Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

Intention and Literary Interpretation

Szu-Yen Lin

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in Philosophy

The University of Auckland

2018

Abstract

This thesis looks into the debate over intention and interpretation in the artistic context, and focuses on literature in particular. The main issue here is whether or not the author's intention is relevant to the interpretation of her work. I critically discuss the major positions in the analytic tradition and defend my view in the final chapter.

The position that rejects appeal to external evidence of authorial intent is called anti-intentionalism, which claims that linguistic convention determines work-meaning and that such convention is all the interpreter needs. This position, alternatively called conventionalism, suffers from the contextualist criticism that contextual factors, aside from linguistic convention, play a crucial part in determining work-meaning.

A variation on anti-intentionalism is the value-maximizing theory, which holds that the interpreter should look for interpretations that maximize the work's value within the limits set by relevant linguistic (and perhaps contextual) factors. However, some philosophers disagree that aesthetic satisfaction is the primary aim of interpretation.

Contra anti-intentionalism, actual intentionalism claims that authorial intention is an indispensable element in interpretation. The extreme version identifies what a work means with what the author intends it to mean, and is widely seen as implausible. The moderate version acknowledges failed intentions, and claims that when an intention fails meaning is determined by convention plus context. I argue that the moderate version has a difficulty in giving convention and context the same credit for meaning-determination as authorial intent.

A middle course between actual intentionalism and anti-intentionalism is hypothetical intentionalism, which maintains that work-meaning is determined by the intention which the appropriate audience is best justified in attributing to the author. A variation on this position holds that it is the postulated or implied author's intention that determines work-meaning. I argue this latter position is more convincing because it does a better job in balancing interpretative freedom and constraints.

Acknowledgements

I would like to thank my supervisors for making this thesis possible. First and foremost, I want to express my immense gratitude to my main supervisor Stephen Davies, who always reads and corrects my manuscripts patiently and meticulously. I have learned and benefited a lot from him.

I extend my grateful thanks to my co-supervisors Fred Kroon and the late Jonathan McKeown-Green. Fred has provided many useful comments to improve my thesis; Jonathan suggested some useful readings at the early stage of my research.

Secondly, I want to thank Stephen's "aesthetics legion." This includes my dear colleagues Justin Horn, Matteo Ravasio, Chung-Chen Soong, and Daniel Wilson (arranged by alphabetical order). They have helped me in different ways during my studies, and I will miss the great times spent with all of them. There are also special thanks to aesthetics visiting students Myriam Albor, Andreas Fukerider, and Mehdi Shams (arranged by alphabetical order).

Thirdly, I want to thank the following visiting scholars for their insightful comments on or suggestions about my research: Theodore Gracyk, Paisley Livingston, Daniel Nathan, and Richard Shusterman (arranged by alphabetical order).

Finally, I treasure the mental support from my family (my parents, my sister, and my dog). And I could not have survived the past four years living overseas without the support and encouragement from my wife.

Table of Contents

1. Introduction

1.1 The problem

When we interpret a literary work, what we mainly do is try to grasp the meaning of the work. For instance, when we read a poem, we may want to know what some particular word or line of that poem means, or what themes that poem should convey. The question hence arises: how do we decide on the meaning of a work? How do we settle this? A quick and intuitive answer would be: we settle this by appealing to what the author intended her work to mean. If I write a novel and you are confused about some ambiguous content in it, it is very natural for you to think that consulting me would clear up your confusion. My intention is authoritative in settling interpretative questions because I am the author. But is this really the case?

Consider the following example. The 1898 novella, *The Turn of the Screw*, written by Henry James, features a young governess who experiences supernatural activities. However, there is ambiguity regarding the nature of the evil in the story. One version of interpretation indicates that the narrator is actually deranged and under an illusion, which adds psychological depth to the work. Another line of thinking sees the work as a gothic ghost story with great atmosphere and hence recognizes the ghosts' reality. The fact is that James himself rejected the psychological interpretation and claimed that he intended the work to be exactly a ghost story. This authorial revelation has raised controversy because some people claim (and some object) that, given textual evidence, the psychological interpretation makes more sense. Should we take James' intention seriously? The answer comes very hard.

The moral of this example is that, though it appears intuitively appealing that authorial intentions settle interpretative questions, in many cases we somehow resist settling which interpretation should prevail by resort to them. In fact, much of contemporary literary criticism tends to disregard and downplay the author's actual intentions.

It is clear that the relation between intention and meaning is the point at issue. In other words, the debate here is mainly about the relation between the following two statements (Wreen, 2014):

(1) The artist intended x to mean p in work w;

(2) x means p in w.

Early intentionalists such as E. D. Hirsch claim that (1) necessarily entails (2). Anti-intentionalists such Monroe C. Beardsley claim that (1) not only does not entail (2) but also cannot offer even adequate evidential support for (2). As the debate goes on, a tangle of related issues have been gradually revealed and the initial problem turns out to be more complicated than people thought it to be.

To give some examples of these surrounding issues: is it a normative or ontological question when we ask how work-meaning is determined? Is there such a thing as work-meaning? Is James' intention to classify *The Turn of the Screw* as a ghost story different in kind from his intention to convey themes or theses (if there are any) in that story? Is there a single right interpretation of a work or can there be many? These, among others, are all difficult questions that probably no philosophers can answer at one go, and very often different theorists have different answers to them.

Though most of the surrounding issues will be revealed and clarified in due course, it will be helpful to elucidate some basic terms such as "intention," "meaning" and "interpretation" before I offer a brief summary of the present debate. There is no consensus on how to define these terms, and my purpose at this stage is not to settle on one view, but to give the reader a rough idea of how these cardinal terms are construed and used so as to dispel doubts about whether the debate is based on problematic grounds.

1.2 What is intention?

Let us start with "intention." William K. Wimsatt and Beardsley define intention as "design or plan in the author's mind" (Wimsatt & Beardsley, 1946, p. 469), which is similar to Paul Grice's suggestion that an intention is an explicit or conscious plan (Grice, 1957, p. 386). Beardsley later elaborates his conception of intention as follows: "The artist's intention is a series of psychological states or events in his mind: what he wanted to do, how he imagined or projected the work before he began to make it and while he was in the process of making it." (Beardsley, 1981a, p. 17) Of course this is still a very rough characterization, but this seems to be a matter of general agreement among most of the early participants in the debate. It is worth mentioning that Beardsley's revised and final view of intention echoes the reductive analysis one will find in a contemporaneous textbook of philosophy of mind: intention is constituted by belief and desire (Beardsley, 1978a).

Some actual intentionalists explain the nature of intention from a neo-Wittgensteinian perspective: authorial intention is viewed as the purposive structure of the work that can be

discerned by close inspection (Lyas, 1992; see also Carroll, 1992, pp. 100-101). This view challenges the supposition that intentions are always private and logically independent of the work they cause, which Noël Carroll interprets as a position held by anti-intentionalists.[1]

The most thorough discussion of intention is given by Paisley Livingston (2005a, pp. 1-30; 2005b, pp. 275-280; 2010, pp. 401-404). According to Livingston, intentions are executive attitudes toward plans; these attitudes are firm but defeasible commitments to acting on them. Contra the reductive analysis of intention, this view holds that intentions are distinct and real mental states that serve a range of functions, irreducible to other mental states. However, Jerrold Levinson points out that there are artistically relevant intentions that should be understood along different lines. For example, there are intentions that concern "the intended reception and treatment of artworks on the part of audiences" such as the intention that something be taken as a sculpture. (Levinson, 2016, pp. 139-140) To this, Livingston replies that the said intention can be re-described by identifying the action-plan involved. Therefore, "intending something be taken as a sculpture" should be restated as "intending to make a sculptural work and so intending to get one's hands into the clay." (Livingston, 2009, pp. 85-86)

One thing worth mentioning is that discussions on intention sometimes involve the issue of authorship. This happens when the recognition of intentions is used as a criterion for making sense of collaborative art-making and collective authorship, which has a lot to do with meshing plans and mutual beliefs about them (Livingston, 2005a, pp. 62-90). Both issues deserve a book-length analysis which is beyond the scope of the present thesis. However, the said proposal of joint authorship seems convincing and may potentially handle controversial cases.

For most of the debate, intention is accounted for by one of the three above proposals; and it is single authorship that is almost always assumed for discussion.

1.3 What is meaning?

It is very important to note that the term "meaning" is typically used in a different way in the philosophy of literature from the way it is used in other contexts, such as the philosophy of language. Meaning in the case of literature is not meaning *simpliciter*; it does not refer to

---

[1] There is controversy over whether Wimsatt and Beardsley hold this view. See Lyas 1992. However, it is reasonable to envisage a version of anti-intentionalism that is based on such a construal of intention.

literal meaning. Rather, it is concerned with secondary or *implicit* meaning. Following this line of thinking, Beardsley has developed a very sophisticated theory of meaning, the discussion of which I will postpone until the next chapter. Here suffice it to say that when interpreting a literary work, the sort of meaning we are after is not what the work literally says, but what it implicitly says. One example of implicit meaning is the themes and theses of a work. As Noël Carroll defines it, "The meaning or content of an artwork can take various shapes. There are themes and theses […] artworks may also possess meaning in terms of exhibiting expressive properties like sadness, joy or gloom." (Carroll, 2013, p. 8) Livingston has a broader construal of meaning, which includes "the events, characterizations, and situations figuring in the story conveyed […] as well as the emotions, attitudes, and themes expressed by any work of art." (Livingston, 2005a, p. 138) This construal is more or less the assumption made by most participants in the present debate.[2]

One point that often confuses readers not familiar with the debate needs to be clarified. *The meaning of a work is certainly an aggregate in the present context.* There are local meanings to be interpreted, such as those concerning implicit word or sentence meanings; there are also regional meanings to be considered, such as themes and theses. So when philosophers of literature use terms like "work-meaning" they typically refer to a composite of meanings. But this does not rule out the cases in which it is the meaning of part of a work that is under scrutiny. Simply put, philosophers of literature discuss examples involving both kinds of meaning but would probably acknowledge that there is an ultimate form of macro-meaning that encompasses both.

An analogy philosophers of literature favour is that between literary works and utterances. This analogy was drawn as early as 1958 in Beardsley's monumental book *Aesthetics: Problems in the Philosophy of Criticism* (more on this in Chapter 2). In terms of this analogy, the meaning of a work is utterance meaning, and literary works are a special kind of utterance—what Levinson calls "literances" (Levinson, 1992, p. 229). However, whether the determination of utterance meaning involves contextual factors such as authorial intent is an issue among utterance theorists, and this will be discussed in more detail in § 5.2.4.

As can be anticipated, there have been objections to the view that a work can possess a meaning, given the unique mode in which it is created and the complex structure it has; and there have been objections to the view that literary works are utterances at all. These objections will be reviewed and countered in § 4.3.2.

---

[2] To save space, similar views will be discussed in due course.

1.4 What is interpretation?

According to Gregory Currie, interpretation is the attribution of meaning (Currie, 2005, pp. 291-292). Here "attribution" can be construed in different ways. It can mean "construction," which means that interpretation ontologically creates a meaning; it can more moderately mean "imposition," which means ascribing a meaning to a work without ontologically creating anything; it can mean "recovery," which means retrieving something already existing in a work. Many of the major positions in the debate to be discussed in this thesis endorse either the impositional view or the retrieval view.

Interpretation is also more broadly characterized as the attempt to explain the features of a work. For example, Carroll holds that aside from digging out meaning, interpretation aims to show how a work's design supports that meaning (Carroll, 2013, p. 8); Daniel O. Nathan holds that interpretation is a matter of asking "why," of seeking an explanation (Nathan, 1992, p. 196). Some theorists characterize explanation as aiming to reveal the work's value. For instance, according to Alan H. Goldman, interpretation seeks to achieve appreciation fully by showing how the elements of a work contribute to values (Goldman, 2013, p. 24); Peter Lamarque sees interpretation as finding a thematic perspective that unites relevant features in a work to form a coherent reading so as to reveal the value of the work (Lamarque, 2002, pp. 297-303).

Though the use of the term varies, interpretation can be seen roughly as involving meaning-attribution as Currie construes. This is because thematic contents and other features of a work, as shown, can be covered by the broad sense of "meaning."

1.5 Scope and outline

As said, the literature on the present debate is voluminous and it is practically impossible to cover all the positions and views in a thesis of limited space. Therefore, the discussion will be selective. My focus will be on the most representative positions and views in the Anglophone analytic philosophy of literature, though there will be occasional visits from literary scholars. The focus is mainly on literature, but examples from the other arts will also be used; the reader should keep in mind that the theories to be discussed can be easily extended across the

arts and that some philosophers, such as Carroll and Robert Stecker, do not maintain that their theory applies exclusively to literature.[3]

As I see it, there are two major camps in the debate: intentionalism and anti-intentionalism. Intentionalism can be divided into actual intentionalism and hypothetical intentionalism. The former divides further into three versions: absolute, extreme, and modest. Hypothetical intentionalism divides into the version based on the actual author and that on the implied author.

The other camp—anti-intentionalism—also divides into two versions: conventionalism and the value-maximizing theory. All the positions mentioned so far are mapped out in the following diagram:[4]
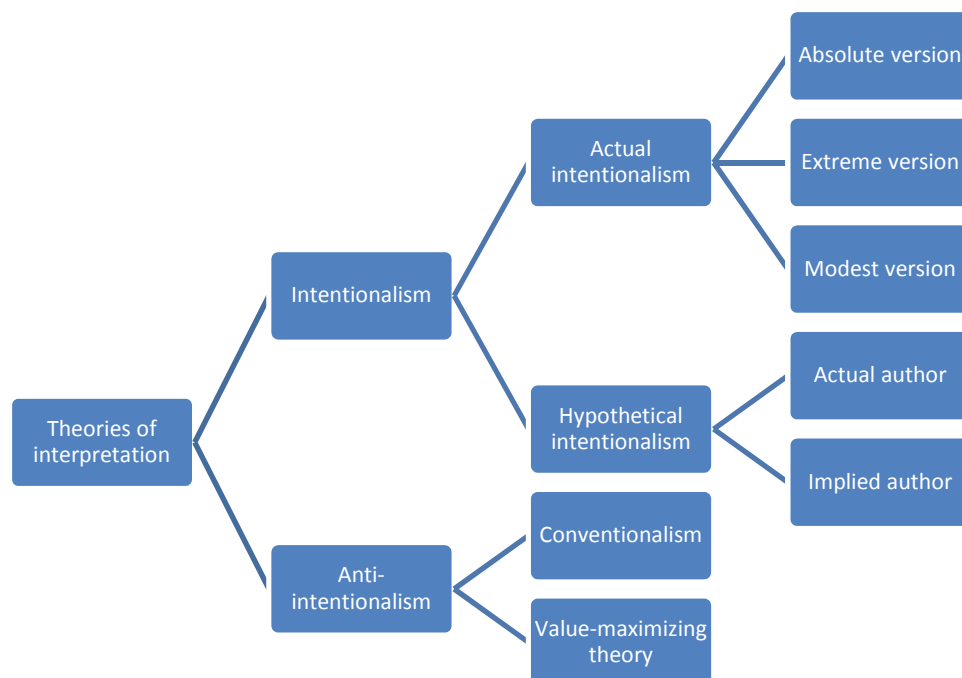
In the chapters to follow I will critically discussed all the positions shown in the above diagram, beginning with anti-intentionalism. The remainder of this chapter is a brief outline of how I arrange the coming chapters and their contents.

Chapter 2 focuses on conventionalism. When used narrowly, the label "anti-intentionalism" refers exclusively to Beardsley's conventionalism; when used broadly, it can cover any positions that degrades authorial intent. For the most part I will adopt the narrow use of the

---

[3] For this reason, from now on the reader does not need to care too much about my shifting use of term between "philosophy/philosopher of art" and "philosophy/philosopher of literature."

[4] For a slightly different taxonomy see Spoerhase, 2007, p. 103.

term and use "anti-intentionalism" and "conventionalism" interchangeably. Simply put, conventionalism maintains that a work is autonomous with respect to its meaning, and that appeal to authorial intent in this regard is irrelevant. The most devastating objection to this view is that it ignores the role context plays in shaping the identity and aesthetically relevant properties of a work, since many philosophers have shown that factors external to the work are not ontologically idle. Nonetheless, there is room for debate over whether Beardsley does hold an absolute anti-externalist position.

Chapter 3 discusses different versions of the value-maximizing theory. This position claims that interpretation should aim to maximize the value of the work within the limits set by linguistic conventions. The maximizing view endorses pluralism given that there are different ways to produce interpretations that enhance a work's value. Nevertheless, it seems that the maximizing view, though convincing in acknowledging critical pluralism, would have a difficulty in totally getting rid of authorial intent, because it implicitly assumes a hypothetical author, as will be explained in Chapter 7.

Chapter 4 turns to the strongest contender in intentionalism: actual intentionalism. The three major versions are reviewed and criticized. The absolute version claims that work-meaning is identical to what the author intended the work to mean. This position is not widely supported because it leads to a Humpty-Dumpty theory of meaning. The extreme version amends this drawback by conceding that authorial intention is necessary, but not sufficient, for meaning. But this view has the odd result that the text becomes meaningless when the author's intention is not realized. The most convincing version—the modest version—adopts a disjunctive scheme: meaning is determined by successful intention or by convention and/or context when the intention fails. I argue that the disjunctive project is unpromising because it faces a dilemma stemming exactly from its disjunctive nature.

Chapter 5 discusses hypothetical intentionalism and centers on the version based on the actual author. This position claims that the meaning of a work is determined by the best hypothesis as to the author's intentions made by the intended, ideal, or competent audience. I critically review these three variations in said order. The biggest challenge to the actual author version of hypothetical intentionalism is that it easily collapses into the implied author version.

Chapter 6 introduces the second version of hypothetical intentionalism, which is based on the implied author. I dub this position hypothetical authorism, to be distinguished from the previous position. This view claims that it is the intention of the implied author that the interpreter should pay attention to. An implied author is the author implied by the text (and

perhaps also by considerations of context), that is, it is how we make out the author to be on the basis of textual features. I review five variations of this position and fend off major objections to the general form of it.

In Chapter 7, the final chapter, I develop and defend a version of hypothetical authorism. My direction is totally different from all the positions discussed so far in that I advance my view neither from the ontological nor from the normative, but from the descriptive point of view. I show that all the current theories based on the utterance model of interpretation end up targeting an implied author. That is, hypothetical authorism best describes how critical practice following the utterance model actually takes place. I then proceed to offer normative arguments in support of hypothetical authorism.

My thesis concludes with some observations about and possible directions for contemporary analytic philosophy of interpretation.[5]

---

[5] For useful survey articles or book chapters on intention and interpretation, see Livingston, 2005b; Currie, 2005; Davies, 2006, pp. 109-134; Lyas & Stecker, 2009; Stecker, 2013; Carroll, 2016.

2. Conventionalism

"Critical inquiries are not settled by consulting the oracle."

—Wimsatt & Beardsley, "The Intentional Fallacy"

This chapter focuses on Beardsley's conventionalism, or *anti-intentionalism* when the term is used narrowly. The core idea of anti-intentionalism in literary interpretation is roughly that appeal to external sources of authorial intent is not authoritative for determining the meaning of a literary work. This position is rigorously defended by Beardsley, who, along with Wimsatt, initiates the intention debate and follows with a series of lively philosophical discussions on literary interpretation and its related puzzles.

This chapter is divided into two parts. In §2.1, I first give a review of Beardsley's early position, presenting its most important assumptions, claims and arguments. Then I move on to discuss some common objections, the most daunting challenge for Beardsley being posed by ontological contextualism: the neglect of context makes his formalist theory less plausible. However, I side with some philosophers in holding that Beardsley actually leans toward contextualism, which means that it may not be adequate to interpret him as an absolute anti-externalist with respect to meaning and other aesthetically relevant properties.

In §2.2, I examine Beardsley's later position, which is based on a version of the pretence theory of art. Beardsley holds that, since fictional works are imitations of illocutionary acts, their meanings are determined by convention rather than by intention. Here Beardsley continues to support anti-intentionalism with the help of a particular view of literary ontology,

2.1 Beardsley's early position

2.1.1 Interpretation and meaning

It would be useful to give a brief review of Beardsley's definition of interpretation and theory of meaning before I consider the major claims of anti-intentionalism.[6]

---

[6] Beardsley's earliest definition of interpretation can be found in his 1953 dictionary entry. What I present below is his revised view.

Interpretation in the broad sense includes explication, elucidation and interpretation in the narrow sense. It aims to discover and report the *implicit meanings* in the work (Beardsley, 1981a, p. xlviii).[7] Put another way, interpretation is the task of digging elusive meanings out of the text.

Explication is the task of pointing out *suggestions* and *connotations* in a work (1981a, pp. 122-126, 129-134). Beardsley builds these concepts on distinctions between primary/secondary sentence meaning and primary/secondary word meaning respectively. *Primary sentence meaning* is the explicit meaning of a sentence, stated by grammatical form. *Secondary sentence meaning* is the implicit, suggested, or implied meaning, what we can infer about the speaker's probable beliefs, and part of the full meaning of the sentence. Beardsley's example is the following statement: "Napoleon, who recognized the danger to his right flank, himself led his guards against the enemy position." From the primary sentence meaning of this statement we can know it at least suggests that Napoleon took this manoeuvre because of the recognition of danger.

As for *Primary word meaning*, this is what a word *denotes*, its *designation*. *Secondary word meaning* is what a word *connotes*, also part of the full meaning of the word. For example:

The word "sea" *designates* certain characteristics, such as being a large body of water; this is its primary word-meaning. It also *connotes* certain other characteristics, such as being sometimes dangerous, being changeable in mood but endless in motion, being a thoroughfare, being a barrier, and so on. These are its secondary word-meanings. (1981a, p.125, emphases original)

Elucidation is the producing of a statement that is an (or part of an) hypothesis about some state of affairs in the world of the work (1981a, pp. 242-247). This kind of reasoning is in principle the same as the kind of reasoning we employ in daily life to determine what actually happens behind the scenes, except that in the latter case we can further test our hypothesis. Elucidation is done to determine the unknown parts of the world of the work based on the parts known plus relevant laws. Some examples of elucidative questions are as follows. Are there really ghosts in *The Turn of the Screw*? Is Hamlet mad? What is the exact location of the wound Dr. Watson got during the war? Of course, the list of this kind of inquiry for a

---

[7] For Chapter 2, all references are to Beardsley's works unless otherwise specified.

story will be endless because it is impossible to specify all the details of a fictional world. But there is no need to answer most of them, for many of the elucidative questions will be irrelevant to our appreciation of the work. For example, it is pointless and trivial to elucidate how many hairs Macbeth has.

Interpretation in the narrow sense is the task of determining the theme(s) and thesis (theses) of a work based on explications and elucidations (1981a, pp. 401-411). *Theme* is an abstract noun or phrase such as "inhumanity" or "the futility of war." *Thesis* is the doctrine or ideological content of the work. An example will be "the work embodies a mystical view of life."

Interpretation in the broad sense (including explication, elucidation and interpretation in the narrow sense) concerns both *local meanings* and *regional meanings* (1970, p. 44). Local meanings apply to words, phrases, and sentences, while regional meanings apply to parts of a work or the work as a whole. The latter depends on the former.

A final point to note is that Beardsley is a realist with respect to meaning and interpretation; that is, he believes there exist objective constraints on interpretation. His position is strong in that he maintains these constraints are sufficient to determine a single right answer to any interpretative question. However, he denies that the author's intention is the right kind of constraint. This issue lies at the heart of the intention debate to which I now turn.

2.1.2 The intentional fallacy

The article "The Intentional Fallacy" by Wimsatt and Beardsley (1946) is the starting point of the intention debate, and still remains one of the most influential papers in the philosophy of literature. At the time the article came out, the most popular form of literary criticism was called biographical criticism, the main idea of which is that, to grasp the meaning of a work, the interpreter needs to study the life of the author, because the work is seen as reflecting the author's psyche. This approach resulted in leading people to consider the author's biographical data rather than her work. Literary criticism became criticism of biography, not criticism of literary works. Wimsatt and Beardsley were strongly against this way of doing criticism, so they collaborated on writing the paper in question as an attack on biographical criticism.

Their main claim in that article is that appeal to *external* evidence of the author's intention is *irrelevant* to interpretation, so such appeal involves a fallacy, namely, the *intentional fallacy*. They famously maintain that the author's intention is "neither available nor desirable

as a standard for judging the success of a work of literary art." (Wimsatt & Beardsley, 1946, p. 468) Here "success" probably means "communicating meaning" rather than "literary value." Their argument for this claim runs as follows. Either the author's intention is successfully realised in the work or it fails; if the intention is successfully realized in the work, appeal to external evidence of authorial intent is not necessary (we can know the intention from the work); if it fails, such appeal becomes insufficient (the intention turns out to be extraneous to the work); the conclusion is that appeal to authorial intention is either unnecessary or insufficient.

One major concern behind this argument is that we should not be led away from the work, that is, we should concentrate on the work itself, because the worry of anti-intentionalists is that we are likely to stray from the work and delve too deeply into the life of the author as biographic critics did. So what the anti-intentionalist needs to do is to establish the authority of the text. Beardsley does so by presuming two pivotal principles, on which a series of interrelated theses are based, and they constitute the focal part of his anti-intentionalism.

The two ontological principles of literary works Beardsley presumes are the Principle of Independence and the Principle of Autonomy. The former states that "literary works exist as individuals and can be distinguished from other things," and the latter states that "literary works are self-sufficient entities, whose properties are decisive in checking interpretations and judgments." (1970, p. 16). The first principle establishes the existence of the object of interpretation. If there is no object to be judged or criticized, literary criticism would be impossible.

The second principle is alternatively called *semantic autonomy* by E. D. Hirsch (1967, p. 4), which affirms that a literary work is a self-contained linguistic entity, the meaning (and other properties) of which is generated by the text, not by some force outside it that controls what it should mean. Beardsley is committed to the view that a work is identical to its text and work-meaning is hence textual meaning (1970, p. 17). The question of how to determine textual meaning leads to the first anti-intentionalist thesis.

2.1.3 The anti-intentionalist theses

(AT1) *Textual meaning is determined by linguistic conventions.* The conventions in question include the semantics and syntax of the words used to form the text, and the cultural and historical resources behind the language concerned (Wimsatt & Beardsley, 1946, pp. 477-478). Roughly, it is our habitual knowledge of language, of how people use it most of the

time. In general, conventions fix meaning without creating indeterminacy. That is, we can usually know the undisputed meaning of a given linguistic unit based on our knowledge of the linguistic conventions. In the case of explication, for example, there is really not much room for what a word will connote or what a sentence will suggest, given its primary meaning, which is stipulated. The limited possibilities can further be narrowed down to a single right answer when put in broader textual contexts.

The same method goes for elucidation. The process is very much like detecting truth in a mystery novel. To figure out who the killer is, the reader will need to put the known facts together and rule out rival possibilities to arrive at the definite conclusion. The relevant facts or clues constitute constraints and the conclusion is in no way random.

Thematic contents are elicited in the same way based on the results of explication and elucidation. For example, it would be absurd to say that the thesis of Macbeth is "love will find a way." Based on the results inferred from linguistic conventions, the play simply does not support that reading. Therefore, it is fair to say that Beardsley's conventionalism is *robust* in the sense that conventions typically indicate a salient meaning.

Since "typically" does not mean "absolutely," Beardsley does accept that there are cases of ambiguity, but he believes this kind of case is rare. Most of the time, the indecisiveness of meaning can be resolved if we study the text closely. All we should rely on to resolve ambiguity is the work itself, instead of appeal to the author's proclamations of intentions. This implies that the most important evidence we need for interpretation is *internal* to the work, which is the main point of the second anti-intentionalist thesis.

(AT2) *External evidence for work-meaning is irrelevant to interpretation.* This point is demonstrated by the dilemma argument presented earlier. When interpreting a work, as AT1 indicates, the best evidence one should draw on is what is discovered in the work through applying the appropriate linguistic conventions. It is in this sense that the evidence is internal. This kind of evidence is also public, since our habitual knowledge of linguistic conventions is of a public nature.

What sits at the other end of the spectrum is external evidence, the kind of evidence involving authorial revelations outside the work. External evidence, such as the author's interviews, diaries, notes or letters, is said to be private, in the sense that it is "not a part of the work as a linguistic fact." (Wimsatt & Beardsley, 1946, pp. 477-478).

Given the Principle of Autonomy and the conventionalist theory of meaning, it goes without saying that interpretation should be grounded on internal, rather than external evidence. Nonetheless, there might be cases in which the author used certain words in a

private or semiprivate way. That is, the words concerned were not used in a conventional way. In such a case, total reliance on conventions for interpretation seems to miss the target.

To solve this problem, Wimsatt and Beardsley make a controversial move: they allow a third kind of evidence to be used in interpretation, one that sits between internal and external evidence: *intermediate evidence*. The sole goal of using intermediate evidence is to reveal the private or semiprivate meanings or associations attached to the words concerned, and nothing more. This is because a word garners its meaning from its history of use (how people have been using it, namely, its conventional use), so we need to take into consideration how the author has previously used it.

One reason for thinking that adopting intermediate evidence does not involve intentionalism is that even in the account of intermediate evidence, meaning is said to be determined by the history of the word's use, that is, how the author has conventionally used it. This could be explained by Beardsley's account of detachment from reference: once the representing practice is established reference is detached from the referrer to the symbol. (See § 2.2.1 for more on detachment from reference.) Of course it might be argued that convention can only be established by a community instead of by one person. In that case invoking intermediate evidence would suspiciously lead to intentionalism.

At the same time, even though Wimsatt and Beardsley suggest intermediate evidence may be used moderately along with internal evidence, they are happy, though reluctant, to concede that the distinction between the three kinds of evidence is not sharp. What they are firmly committed to is the primary status of internal evidence and the dispensability of external evidence. However, it should be noted that the anti-intentionalist is not against using external evidence as a source for generating hypotheses of work-meaning; it is only that these hypotheses need to be tested against the internal evidence.

(AT3) *The author is not to be identified with the dramatic speaker.* We should distinguish between the author and the dramatic speaker (the narrator) in a work. Thoughts and attitudes of a work should be attributed to the dramatic speaker instead of to the author, unless justified by biographical evidence. [8] Broadly speaking, inferences from works to authors are reasonable only in the biographical sense. This means that the biographer can legitimately use works as clues to make inferences about the author, but only in cases where the inferences

---

[8] I argue in an unpublished paper that the narrator as construed in Beardsley's writings is actually an implied author. If this makes sense then Beardsley is implicitly committed to the version of hypothetical intentionalism based on the implied author (hypothetical authorism) that will be discussed in Chapter 6.

can be supported by biographical data. In spite of this proviso, the general attitude of anti-intentionalists is still that this kind of reasoning is not to be made during work interpretation.

The spirit of AT3 is quite clear: anti-intentionalism aims to give literary works full autonomy. It holds a general, if not completely, negative attitude toward seeing the author as materialized in her work. Consequently, inferences going in both directions between works and authors are in theory illegitimate. Particularly, those from authors to works are deemed as intentional fallacies and should be entirely avoided.

(AT4) *That the author's intention is the cause of her work does not entail that it is the standard for judging its meaning.* In the picture of anti-intentionalism given above, what seems to be left for the author is the status of creator. The author is no doubt the cause of her work; more specifically, her intentions are the cause of her literary efforts. Even so, this fact does not grant intention as a standard for judging the meaning of a work. As the Principle of Autonomy suggests, the properties possessed by a work suffice for standards by which criticism can be done. The author's intention at most caused the work, but it can never be used to determine the meaning of the work. To do that, one needs to count on the internal evidence. Anti-intentionalists treat the intended meaning non-preferentially, as one among other potential readings all of which need to be checked against textual evidence.

(AT5) *There is a distinction between speaker's meaning and sentence meaning.* Not many people would dispute that the typical goal of interpreting everyday language is to grasp the speaker's (utterer's) meaning. When someone utters a sentence, one's goal as an interlocutor is to understand what the speaker means by that sentence. A conversation can be carried on only if the interlocutor keeps track of the speaker's meanings.

Apparently, speaker's meaning is determined by the intention the speaker has in mind when she utters the sentence in question. But when a speaker fails to realize her intention, the utterance tends to end up with something else which does not coincide with the failed intention. This shows that when faced with an utterance, we can always ask two questions: (1) What does the speaker (utterer) mean? (2) What does the sentence mean? (1981a, p. 25)

For example, consider a certain sculpture (1981a, p. 21). Suppose the artist who makes the sculpture intends it to symbolize human destiny. Nevertheless, no such symbolism can be discerned at all in the artwork, no matter how hard the audience tries. Such being the case, should we insist that the sculpture indeed symbolizes human destiny, or should we ignore the artist's intention and try to see what the work can really imply? Endorsing the former view is clearly ridiculous, because it means the author or artist can intend her work to mean anything,

irrespective of what the work is. Actual intentionalism is in danger of falling into this implausible Humpty-Dumpty-ism.[9]

Not all intentionalists accept the speaker's/sentence meaning distinction. The absolute version of actual intentionalism maintains that the relation between (1) and (2) is one of identity. That is, the sentence always means what the speaker intended it to mean. Moderate actual intentionalists tend to opt for a third option—utterance meaning—so as to avoid the extremes of sentence meaning and utterer's meaning. (There will be a detailed discussion of actual intentionalism in Chapter 4).

Beardsley rejects the intentionalist analysis. Acknowledging the distinction between utterer's and utterance meaning grants autonomy to language, which is exactly what the anti-intentionalist argues for. Speaker's meaning is one thing, sentence meaning another. We should be careful not to mix them up. On Beardsley's view, linguistic conventions are sufficient on their own to secure textual meaning; the author's intention does not play the key role. Though Beardsley makes this distinction with respect to language in general, he does recognize that the primary goal of understanding daily language is to seek speaker's meaning (1970, pp. 31-32).

In the case of literature, in interpreting a work it makes more sense to focus on what the work means (sentence meaning) than on what the author intended to mean (speaker's meaning). The work, but not the author, should be in the spotlight, because for literary interpretation we have a different aim.

Below I critically discuss a series of arguments Beardsley presents to defend anti-intentionalism.

2.1.4 The anti-intentionalist arguments

Let us start with the dilemma argument already mentioned. Epistemically, this argument shows that we do not need to know the author's intention via external evidence of it to ascertain work-meaning; that is, knowledge of authorial intent based on external evidence is dispensable with respect to literary interpretation. Note that the dilemma argument seems to imply the neo-Wittgensteinian view of intention that we can know what the author's intention is when it is successfully realized in the work.

---

[9] In *Through the Looking-Glass* by Lewis Carroll, Humpty Dumpty said to Alice, "When I use a word, it means just what I choose it to mean—neither more nor less."

Nevertheless, the argument ontologically acknowledges that the author's successful intention can determine meaning. If so, it seems fair to say that what the anti-intentionalist really rejects is the relevance of *failed* intention. Though what the dilemma argument really shows is the irrelevance of appeals to *external* evidence of intent, the total irrelevance of authorial intent to interpretation is indeed presumed by the Principle of Autonomy. It follows that the alleged complete irrelevance of authorial intent is hence undermined by the said presumption in the dilemma argument (Davies, 2014, p. 14).

Moreover, the dilemma argument also suggests that successful intention never alone determines meaning; put otherwise, it at best underdetermines meaning (Davies, 2006, p. 118). Given the Principle of Autonomy, here the role the author's intention plays is embarrassing. It could be that Beardsley had changed his mind when proposing the Principle (Carroll, 2014, p. 39). Otherwise, this admission to intention will significantly undermine another strong anti-intentionalist argument—the two-objects argument, which is only implied, rather than explicitly formulated, in Beardsley's writings. The argument goes like this:

(P1) The work is one thing and the artist (including her intention) is another.

(P2) The interpreter's task is to concentrate on the work.

(C) Any reference to the artist and her intention is irrelevant. (Lyas & Stecker, 2009, p. 366)

The two-objects idea suitably applies to the distinction between sentence meaning and speaker's meaning. This distinction makes it clear that the speaker's intention and textual meaning are two distinct things even if they tend to coincide. Since the critic's task is to concentrate on the work, textual meaning, rather than the author's intended meaning, is what deserves our attention.

Nevertheless, as shown, the dilemma argument implies that the author's intention can be *displayed* or manifested (discoverable by the audience) in the work, which significantly undercuts the first premise of the two-objects argument (Lyas, 1992). Though one wonders how Beardsley would explain the success condition of intention in the dilemma argument, which is a major difficulty for modest actual intentionalism (see Chapter 4), he may not need to do this to maintain his anti-intentionalist view.

As suggested by Carroll, Beardsley later on may have abandoned the idea of successful intention. If so, the issue here will boil down to whether we are willing to accept his Principle of Autonomy, which most theorists reject for reasons to be discussed later.

In addition to the above arguments Beardsley also lays out two normative arguments for favoring textual meaning over authorial meaning (1970, pp. 32-35).

The first argument is drawn from logical considerations, focusing on the availability of authorial intent. When we interpret a literary work, independent evidence of authorial intent tends to be absent, either because the author is long dead or because such evidence is not easily accessible. We can test interpretations only against the work itself. This is why the general and essential task of criticism cannot be the discovery of authorial intent.

Beardsley concedes that the unavailability argument is inconclusive, for the actual intentionalist could maintain that we focus on the text in order to discover the author's intention; or, in Beardsley's own words, the interpreter can be concerned with the authorial meaning the text "discloses." That is to say, a textually attentive reading can still be intention-driven. Now, it is unclear what kind of view of intention is presupposed by the word "discloses." Given that the said response is offered in the paper in which the Principle of Autonomy is proposed, it might be reasonably assumed that here, as Carroll conjectures, Beardsley holds the two-objects view that intention is logically independent of the work, instead of the neo-Wittgensteinian view that intention can be publicly manifested in the work. But in that case Beardsley's reply to his own argument would make less sense, since the text would not be a good guide to authorial intention if the two do not have any logical connection.

Perhaps one could defend Beardsley by holding an intermediate view of intention stating that the text is *indirect* evidence for authorial intent, thereby preserving the independence of intention from the text while maintaining some sort of remote connection. For example, any evidence about the height of the father will be indirect evidence of the height of the son. In truth, this intermediate view is held by Beardsley in *Aesthetics* (1981a, pp. 19-20).

Due to the inconclusiveness of the unavailability argument, Beardsley moves on to propose the aesthetic argument, which is drawn from aesthetic considerations, centering on aesthetic satisfaction. Beardsley claims that the primary goal of literary interpretation is to offer us aesthetic satisfaction by eliciting artistic value from the work. The value of the work presumably resides in the work instead of in the author. When interpreting a work, if we focus on the psychological states of the author we will be approaching the work from the historical point of view rather than from the aesthetic point of view, which, according to Beardsley, does not suit the proper aim of literary interpretation.

Here Beardsley contentiously assumes that the primary goal of literary interpretation is to offer aesthetic satisfaction. Not everyone agrees with this assumption, and one can disagree with many good reasons (Stecker, 2010, p. 147). For one, it is obvious that literary criticism

has been pursued with a variety of purposes, and aesthetic satisfaction is only one among the many (seemingly) legitimate goals. Until the anti-intentionalist proposes an argument in support of the aesthetic goal of interpretation, the fundamental premise of the aesthetic argument is not to be taken for granted.

Alternatively, one can maintain that, though authorial reading might not always be the one that best enhances the value of the work, it can still enhance the work's appreciation in the sense that it adds variety to interpretations. Yet another reason—this one bears more weight—is that normally our initial aim of interpretation is to understand a work's implicit content; by doing so we get aesthetic satisfaction. And reference to authorial intent might be needed to achieve this. Beardsley could reply that, when the interpretation indicated by such authorial revelation is not as coherent and complete as the one best suggested by the text, we should go for the latter, as it arguably gives us better aesthetic satisfaction. However, the case under scrutiny already rules out the possibility that convention points to one salient meaning; presumably that is why the author's intention is called for to disambiguate meaning. To be faithful to his position Beardsley in this case should insist that the ambiguity of the text cannot be "willed away" by the author's intention.

2.1.5 Objections to anti-intentionalism

An early attack on Beardsley's conventionalism involves what is known as the *indeterminacy thesis*, which claims that a word sequence is indeterminate with respect to meaning; only authorial intent can determine its specific meaning; the author's intention is the only way to resolve the indeterminacy (Hirsch, 1967; more on this in Chapter 4). Conventions at best narrow down the range of possible meanings but can never pin down any determinate meaning.

In reply, Beardsley thinks that Hirsch exaggerates the indeterminacy of word-sequence meanings. Texts acquire determinate meaning through conventions. And a word or word sequence is indeterminate only in the sense that "it is capable of acquiring different determinate meanings when placed in varied contexts" (1970, p. 30). We can put it this way: Beardsley's conventionalism is a robust version, in contrast to Hirsch's *thin* version. Robust conventionalism claims that (a) textual contexts tend to rule out alternatives and/or (b) linguistic conventions are often enough to indicate a salient meaning (Irvin, 2006, p. 119). However, as said, this is not to deny that we may have rare cases in which textual meaning is ambiguous.

The debate between robust and thin conventionalism is hard to resolve, since their basic premises contradict each other. And it is hard to say which position explains the relevant linguistic phenomenon better. Some might think the robust version is less plausible because it seems too strong to say that most texts have a definite and uncontroversial meaning guaranteed by convention. This depends on how broadly the robust conventionalist construes the linguistic context. To give an example, the statement "We sell alligator shoes" can either mean "we sell shoes to be worn by alligators" or "we sell shoes made from the skins of alligators." (Nathan, 1992, p. 193) However, if the notion of linguistic convention is rich enough to include background knowledge and assumptions shared by members in the same linguistic community, the second reading of the statement is no doubt the undisputed meaning of this mini-text. How Beardsley construes the linguistic context is a subtle issue, the discussion of which I postpone until later.

Another objection to Beardsley's position is the problem of secret or *idiosyncratic* meanings. In this case, conventions do not help and we can resort only to authorial intent. One possible reply is that the resulting misunderstanding caused by not knowing this meaning will neither be widespread nor serious. It will not be widespread because, when creating a text, the author generally follows linguistic conventions; it will not be serious because if, after close inspection of the work (notably, of intratextual relations), the meaning is still not detected, it must be a marginal meaning (Irvin, 2006, p. 117).

Beardsley could have invoked the use of intermediate evidence to counter this objection. This kind of evidence allows us to interpret the author's idiosyncratic use of words. But since the use of intermediate evidence is controversial, he answers the challenge in another way.

The use of secret meanings is a Humpty-Dumpty way of using words. In philosophy of mind, a popular construal of intention states that intending involves believing and desiring, which is also Beardsley's view. A Humpty-Dumpty way of saying is impossible, for it is impossible for a speaker to believe that she will secure uptake from the audience. Therefore, the challenge of secret meanings fails (1982, pp. 202-203).

What about allusion and irony? In both cases the author's intention seems to play an important role in deciding whether a work alludes to something else or whether it is ironic. With respect to allusion, Beardsley's reply is that, again, the issue is resolved by convention. If convention indicates an allusion in the work, then the work does make that allusion, even if the author did not intend the allusion (1982, p. 200). Whether a work alludes to something or not is eventually determined by internal evidence, the contents of which are in turn determined by public conventions.

As for the case of irony, Beardsley responds in the same way as he does to the objection of secret meanings. For the author to intend to say something, she must not only desire but also believe that she will do it in producing the text. The latter is impossible if there is no evidence in the work that indicates irony, so that she should not believe that uptake from the audience can be secured. In this case, the text, if produced, cannot be said to be ironic, or successfully intended to be ironic. Irony must offer the possibility of understanding. If we and the author cannot imagine anyone taking it ironically, there would be no reason to believe that it is ironic (1982, pp. 205-206; see also 1981a, p. 26). Again, if the notion of the linguistic convention is construed in a broader way then internal evidence will need to be understood as going beyond mere dictionary meanings of the words used. For Beardsley, this move would potentially handle many cases of irony. I now discuss the controversial issue of how broad linguistic context should be construed in Beardsley's writings.

The way Beardsley answers the aforesaid objections reflects his insistence on the work's autonomy: let the work speak for itself. To test an interpretative hypothesis we should always check it against textual or internal evidence, even in the case of irony or allusion. Nevertheless, some philosophers maintain, this dogged insistence neglects the role played by a work's context in shaping the relevant aesthetic qualities (meaning, in particular). One can question whether meaning is hermetically sealed within the work as Beardsley's Principle of Autonomy claims. The tension between a work's autonomy and external influences on meaning is most clearly highlighted by Beardsley's allowance of intermediate evidence.[10] It seems that contextual factors do matter in interpretation. This worry is first addressed by Tolhurst (1979).

Tolhurst plausibly shows that, just as the same sentence on different occasions can mean different things, the same text in different contexts can have different meanings. This demonstrates that textual meaning is determined at least in part by context, which shows we must reject Beardsley's semantic autonomy.[11] Such being the case, a text is better seen as a word sequence embedded in a context, that is, an utterance, in contrast to a bare word-sequence which is context-free. Textual meaning is thus suitably understood as utterance meaning rather than as word-sequence meaning. Once we agree on this, it will be difficult to exclude consideration of authorial intent from interpretation, since utterance meaning is best

[10] Intermediate evidence is never mentioned in Beardsley's writings after "The Intentional Fallacy" probably because of the worry mentioned on p. 14.

[11] Note that Tolhurst is not careful enough to make the terminological difference between work and text. More will be said in this regard in Chapter 5.

defined as the intention best justified in being attributed to the author by the intended audience based on their attitude to and knowledge of the context in which the work was created.[12] The implications here are: (1) the author's intention is still relevant in some way to interpretation; (2) the context of production is taken into account in determining work-meaning (this is the core claim of *contextualism*). Beardsley objects to (1), and seems to ignore (2), *if* we read him as not considering public linguistic conventions at the time of production as part of the context.

Beardsley responds to Tolhurst's pro-intentionalism only briefly (1982, pp. 199-200). He gives a counterexample as a reply and admits that Tolhurst "defends a very persuasive theory." He also makes it clear that he does not contest the thesis that contextual factors present in the creation of a work may play a role in determining work-meaning. But Beardsley died before he had the chance to say more.

However, it has been pointed out by some philosophers that Beardsley actually leans toward contextualism (Currie, 1990, pp. 109-111; Nathan, 1992; Davies, 2005; Livingston, 2005a, p. 141; Davies, D., 2007, pp. 72-73). This is because his construal of linguistic convention can be, as noted, taken as very broad. In "The Intentional Fallacy" it is said that the meaning of a poem can be "discovered through the semantics and syntax […] through our habitual knowledge of the language, through grammars, dictionaries, and all the literature which is the source of dictionaries, in general through all that makes a language and culture." (Wimsatt & Beardsley, p. 477). This rich notion of linguistic convention arguably covers many relevant contextual factors required by contextualism.

Moreover, Beardsley also allows that "apperceptive mass" (apperceptual information such as the style of the work, relations with other works, etc.) can affect what is internally and objectively present in a work (1981a, pp.52-53; see also Davies, 2005, pp. 180-181; Nathan, 2006, pp. 283-286). The right kind of apperceptive information serves as the aesthetically relevant contextual factor that shapes the internal objectivity of the work. The question of whether Beardsley is a contextualist is beyond the scope of this thesis, and I argue for the positive answer in detail elsewhere.[13] If Beardsley is immune to the contextualist criticism, the only possible effective objection to his anti-intentionalism I can think of would be the inevitability of accommodating some sort of intention in interpretation. But I argue in an unpublished paper that Beardsley is committed to a version of hypothetical authorism and

---

[12] Tolhurst's theory will be examined in more detail when we discuss hypothetical intentionalism in Chapter 5.

[13] See my accepted paper (Lin, forthcoming).

hence does not disregard intention entirely. The brief conclusion here is that the reinterpreted Beardsley would be a rival to my version of hypothetical authorism presented in the final chapter.

2.2 Beardsley's later position

In his later life Beardsley develops an account of literary ontology and uses it to support his theory of interpretation. Nonetheless, this move works only with fictional works and to some extent undermines his anti-intentionalist position on nonfictional works. I explain this below, beginning with an introduction to his speech act theory of literature.[14]

2.2.1 Illocution, intention, and representation

Speech act theory was originally proposed and developed by J. L. Austin (1962) to explain deeds performed by utterances in particular contexts. His focus is mainly on illocutionary acts, i.e., speech acts performed in using language. The operational schema can be described at the most general level: an illocutionary act is generated by intentional text-production under appropriate conditions, according to certain language conventions.

The rule-governed nature of illocutionary acts is crucial for Beardsley in that he notably denies the relevance of speaker's intention in determining the correct illocutionary act being performed. An illocutionary act can be performed unintentionally in the sense that it is not necessary for the speaker to have the corresponding intention to perform that illocutionary act (1982, pp. 196-197). Notice that this intention is not the default intention in "intentional text production."

To demonstrate the point, suppose the following conditions hold: (1) I know that you are looking for an Agatha Christie novel to read; (2) I believe that you have read *At Bertram's Hotel*; (3) I do not know that actually you haven't read it; (4) I think that it is not worth reading; (5) I gather that you plan to read *At Bertram's Hotel*. Under these conditions, my uttering the sentence to you "*A Murder is Announced* is ten times better than *At Bertram's Hotel*" constitutes the illocutionary act of advising you to seek *A Murder is Announced* to read. This act is performed intentionally, i.e., with the intention to give you advice. But at the

---

[14] This section is part of my paper (Lin, 2016). The material presented here has been slightly modified.

same time there is another illocutionary act being performed unintentionally: that of discouraging you from reading *At Bertram's Hotel*.

It follows that performing an illocutionary act has nothing to do with having as its condition an intention to perform that act. We do not need to know the speaker's intention to determine which illocutionary act is being performed, which is solely a function of relevant conditions and conventions.

Beardsley claims that the nature of fiction is *representation*. In its broad sense, a representation is a symbolic vehicle standing for something else. By this definition, words, texts, paintings or sculptures are all representations. Plainly, Beardsley's use of the term is narrower.

Representation involves what Beardsley calls *selective similarity* (1981b, pp. 294-295; 1982, p. 191). For X to be able to represent Y, X must bear some crucial aspects that can be perceived to be characteristic of Y, but falls short of being Y. In the pictorial context, we know that a painting is a depiction of, say, a smiling face, because its content exhibits selective visual respects distinctive of a real smiling face. The painting is hence a representation of a smiling face. In the case of dramatic representation, we know that a mime is representing the act of, say, climbing a ladder, because her body language shows crucial aspects that are characteristic of ladder climbing. Her act is one of imitation.

Representation, according to Beardsley, is best understood as some kind of reference: X refers to Y if X represents Y. Such reference may be intentional at the very beginning of the practice of representation, in the sense that what is referred to is determined by the referer's intention. However, once the conventions of a particular kind of representation are established, unintentional reference becomes possible. This is called the *detachment of reference* (1981b, pp. 296-297; 1982, pp. 192-193).

Such being the case, we can have different conventions for representing things. Needless to say, the content of representation in most cases corresponds to what one intends to represent, but what one does represent in the end depends on the relevant conventions. It might also happen that one intends to perform some action but ends up representing it instead, if the action is incomplete. This is to say that representation can be nonintentional in the sense that the intention to represent is not necessary for producing a representation. Of course, when the

performance fails one might end up representing nothing at all, if the similarity produced is insufficient (1981b, p. 297).[15]

So far we can see that a representation is in place when an attempted action is incompletely performed but is still discernible in what has been done. Then, what is a representation of an illocutionary act *per se*? Verbal representation works in the same way as visual or dramatic representation: an illocutionary-act representation obtains when the conditions present are not sufficient for the act in question actually to occur. I will define an obtaining condition of an illocutionary-act representation as an *infelicity condition* of an illocutionary-act performance. And this will presuppose that the attempted action is always of a recognizable kind, since in the current context talk about unintelligible actions is out of place. An example of the obtaining condition for an illocutionary-act representation will be the absence of *uptake*. That is, the utterance is not heard and understood by the audience. The illocution hence loses its force and becomes a representation.

The case of intentional verbal representation is a special case of refraining from illocutionary commitment in order to produce a fiction (1982, p. 191). For instance, when you catch someone stealing your wallet and say "You stole my wallet!", you are performing an illocutionary act of accusation. But, when an actress during a performance says "You stole my wallet!", she is not performing an illocutionary act of accusation but only representing one; she is refraining from an illocutionary commitment in order to produce a fiction or imitation.

This is not to exclude the case in which the actress intends to represent some other illocutionary act but ends up representing the one of accusation due to a slip of tongue. The point is that, since representation is a matter of convention, what is represented is solely determined by convention. Furthermore, just as a failure to perform an act may end up with a representation, so a failure to perform an illocutionary act might end up with an illocutionary-act representation because of the absence of uptake from the audience, for example.

It is now clear that a representation or imitation is *essentially* a fiction. The drawing of a smiling face is a fictional smiling face; the represented action of climbing a ladder by a mime is a fictional action of ladder climbing; the illocution-act representation of accusing someone is a fictional illocutionary act of accusation, and so on.

---

[15] Note that unintelligibility is not equal to ambiguity. The content of a representation can be ambiguous in the sense that it resembles more than one type of act. But when the type of action performed is completely unrecognizable, it is unintelligible.

2.2.2 Fiction, nonfiction, and interpretation

Before we see how the said notions find their places in literature, some important terms should be specified. First, Beardsley uses the word *author* to refer to anyone who intentionally produces a text and to no one else. Thus a parrot or a speaking dreamer is not an author. To *mean* something, someone intentionally produces a text and in doing so intends to *say* something. The word "say" covers two different kinds of speech act: illocutionary acts and their representations (1982, pp. 189-191).

Second, a *compound* illocutionary act is defined as a set of illocutionary acts performed in one single text. A literary text, either fictional or nonfictional, can be seen as reflecting a compound illocutionary act (1970, p. 58).

Third, in every literary work we can envisage an *implicit speaker*[16] who performs the compound illocutionary act, and "whose words the work purports to be" (1981a, p. 238).

Now, let us consider the case of literary fiction. Typically, these are fictional narratives such as novels, short stories and plays. The crucial question is: why does Beardsley think of them as representations or imitations of illocutionary acts? This is because we can always discern in them the infelicity conditions of an illocutionary-act performance, i.e., the obtaining conditions of an illocutionary-act representation. These conditions can be seen as the *marks of fictionality*, which make fictional the illocutionary acts performed in the work.

A first point to consider is whether we should identify the implicit speaker with the author. If the answer is no, the author must be representing the illocutionary act performed by the implicit speaker. This speaker typically appears in two ways: (1) she tells the story behind the scenes and never reveals herself; (2) she narrates from the first person perspective as one of the characters in the story world.[17]

The identification is certainly wrong in the second case: no one would identify Dr. Watson with Conan Doyle, who in most Holmes adventures is only imitating the compound illocutionary act performed by a fictional speaker.

---

[16] A similar concept, that of "the implied author", is later fully developed by Wayne C. Booth in 1961. I elaborate this concept in Chapter 6 and in my unpublished paper mentioned earlier on.

[17] There can be variations. For example, a story is told from the second person point of view, or from multiple perspectives. But the point is that in all such cases there is still someone who is telling the story at any given moment and who qualifies as the implicit speaker.

It is perhaps less clear why such identification in the first case is also illegitimate. A quick example may help. Consider the following passage from Ngaio Marsh's mystery novel *Artists in Crime*, in which an artist's model is murdered:

She examined the body. She states that the eyelids fluttered and the limbs jerked slightly. Miss Bostock attempted to raise Gluck. She placed her hand behind the shoulders and pulled. There was a certain amount of resistance, but after a few seconds the body came up suddenly. Miss Seacliff cried out loudly that there was blood on the blue silk drape.[18]

If we take this passage to be an illocutionary act performed by the implicit speaker, it would be one of describing a crime scene. However, given that the event is not real (no model called Gluck dies in the way described by Marsh), and that the names mentioned here do not refer to actual people (there might be someone called Gluck, Miss Bostock, or Miss Seacliff around us, but we cannot find a perfect match in the real world), the illocutionary act of describing this particular event is not likely to occur. The act is thus pretend. This explains why a fictional narrative is a representation of a compound[19] illocutionary act in nature: there are always nonreferring names and descriptions in such a work that make fictional the narrator and the illocutionary acts performed by her.[20]

Beardsley is unusual in holding that, apart from narrative fiction, there is a second literary category that should also be taken as fiction, and thus as a form of representation: lyric poetry.[21] Sometimes the infelicity conditions in such poems are easy to discern. Beardsley's example is the sonnet by John Keats that begins with the line "Bright Star! Would I were steadfast as thou art—." The speaker here is addressing the star but apparently uptake from the addressee is impossible. The purported illocutionary act is not likely to happen, and Keats is hence only imitating the act of praising performed by the implicit speaker.

One compelling reason for treating lyrics as illocutionary acts is that they are typically spoken from the first person point of view to express personal feelings and sentiments. It

---

[18] Ngaio Marsh, *Artists in Crime* (London: Harper Collins, 2000), Chapter 4.

[19] I will omit "compound" hereafter.

[20] Representation in this context means *depiction*, not portrayal, because the act imitated need not be one that has ever occurred (1973, pp. 31-32).

[21] Beardsley initially speaks of the whole category of poetry (1970), but adopts a more reserved position later (1973, 1981b). Also notice that Beardsley does not claim that fictional narratives and lyrics exhaust the realm of fictional works. I suspect that nursery rhymes, for example, would be other possible candidates.

seems natural and right to identify the speaker in the poem with the author, and thus to attribute those feelings and sentiments to the author, when no obvious infelicity conditions are present. Take for instance the poem "Thoughts on a Silent Night" by the ancient Chinese poet Li Bai:

Before my bed a pool of light—
O can it be frost on the ground?
Eyes raised, I see the moon so bright;
Head bent; in homesickness I'm drowned.[22]

Though researchers do not seem to have any evidence that Li Bai based this poem on a true experience, it is very intuitive and tempting for us to assume he did, and hence to identify the speaker in the poem with him. But one would find this kind of identification problematic if one came to know who wrote the following poem only after reading it:

With Northern grass like green silk thread,
Western mulberries bend their head.
When you think of your home on your part,
Already broken is my heart.[23]

This lyric poem, titled "A Faithful Wife Longing for Her Husband in Spring," is a poem about a woman grieving over her husband's absence due to military service. Simply put, the illocutionary act performed by the lonely wife seems to be one of complaining about the absence of her beloved. At first glance, one might suspect that it was written by a female poet. But surprisingly, this poem is also Li Bai's work. It will be absolutely wrong to identify the speaker in the poem with the author, because Li Bai is a male and does not have any husband. It is impossible that he ever had the emotional experience described in the poem. Li Bai in this case is representing the illocutionary act performed by a lonely wife. One would be yet more surprised to find that Li Bai actually wrote more poems from a female perspective, imitating various types of illocutionary acts that express women's emotions and feelings.

---

[22] Yuanchong Xu, trans., *Selected Poems of Li Bai* (Changsha: Hunan People's Publishing House, 2007), 35.

[23] Ibid., p.67. I omit the last two lines to serve the current purpose better.

So far it is clear why Beardsley rejects the identification of the speaker in a lyric with its author. If this identification is universal, what is true of the speaker must be true of the author, and vice versa. But the truth is that there is no necessary link between what is expressed in a lyric and its author. A lyric is hence best treated as an illocutionary-act representation, rather than as an illocutionary act.

Another reason for rejecting lyrics as genuine illocutionary acts is the so-called "address without access," which constitutes the primary mark of fictionality (1981b, pp. 304-305). A published poem is not addressed to a reader the way a love letter is to him or her. The latter is done in a *pragmatic context*, which the former lacks (1981a, pp. 238-240). The absence of a pragmatic context gives us no reason to identify the implicit speaker with the author, who could just be imitating someone else's speech acts. This is because the act of publishing a poem gives it an impersonal public character. This second locutionary act, to be distinguished from the first of producing the text, detaches the poem from the occasion of utterance and gives us no guarantee that the author is performing an illocutionary act. The poem is hence "freed of illocutionary dependence on its occasion" (1978b, p. 175).

It can be seen that the above reason also applies to narrative fiction. I see no problem in thinking of it also as the reason for treating literary fiction as representational in general.

What about nonfictional works? They tend to have strong realistic elements and lack marks of fictionality, or infelicity conditions.[24] As Beardsley indicates, many pieces of discursive prose are duly classified as nonfiction, including history, philosophy, religious meditations, and personal essays (1978b, p. 174). And it seems wrong to say that the illocutionary acts made in these works are pretend. For example, if a poet publishes an antiwar poem during wartime, it seems inappropriate to see the utterance as a mere representation devoid of the illocutionary force of opposing war (1973, p. 35).

One might object that here Beardsley is self-defeating, given his view that "address without access" is a mark of fictionality. Beardsley has a response. When considering the antiwar poem it is hard to deny that publishing such a poem in the context of its production is

---

[24] There are two worries here. First, many literary critics may be hesitant to refer to nonfiction as literature, because of other non-literary and non-aesthetic functions that can be attributed to the text. Second, the distinction between fiction and nonfiction is not always clear. Beardsley has a reply to both worries (1977, pp. 427-428). The answer to the first really depends upon one's definition of literature. According to Beardsley, the literariness of a text will be defined by whether an artistic intention is at play. For the second worry, Beardsley admits that the line is bound to be fuzzy. However, this does not mean that the distinction is useless, as there are more clear-cut examples.

very close to *using* or *presenting* that text, given that there are no other marks of fictionality present. It is better to treat the context in question as a pragmatic one in which it is legitimate to identify the author with the implicit speaker.

There are indeed reasons for making this move. As Beardsley (1981b, p. 305) points out, nonfictional works tend to be published in a more formal style and targeted at a particular audience. For example, many of Confucius's sayings are addressed to political rulers; a large part of Kierkegaard's literary writings are addressed to the Church. There is a sense of intimacy with the reader in this kind of work, in which the author is keen to secure proper uptake from the intended audience; i.e., she is not addressing without access. But this strong sense of communication is absent from fictional works.

The conclusion we can draw at this stage is that, by rejecting the said mark of fictionality, published nonfictional works are justified as genuine illocutionary acts performed by their authors.

Beardsley draws on what has been argued so far to shed new light on his anti-intentionalism on literary interpretation. As I see it, his complete argument based on speech act theory comes in two parts, in which he tries to show that intention is twice removed from an illocutionary-act representation.

First, the author's intention is not constitutive of illocutionary meaning (of which illocutionary act is performed), because performing an illocutionary act does not require the author's intention. Since nonfictional works are genuine illocutions, it follows that the author's intention is neither necessary nor sufficient in determining their meanings.

But even if authorial intent did play a role in fixing the meaning of an illocution, the same would not hold with its representation. These are cases in which authors appear to withhold illocutionary force in order to produce representations. Since the content of a representation is fully determined by convention, authorial intent is logically independent of meaning-shaping. And since fictional works are illocutionary-act representations by their very nature, the author's intention is then irrelevant to what such works mean.[25]

Beardsley concludes that the intentional fallacy comes from the following confusions (1982, p. 207): (1) Of illocutionary-act performance with its representation; (2) of authorial meaning with textual meaning (what was intended with what was actually said); (3) of the real author with the implicit speaker; (4) of the subject of the work with the occasion of its

---

[25] My presentation of Beardsley's speech-act argument is similar to Wreen's (2014). However, I think that he is mistaken in stating that Beardsley's second sub-argument applies to *all* literary works.

utterance. These confusions are implicated in the intentionalist argument that consists of two sub-arguments: from biographical data[26] to probable intentions, and from probable intentions to work-meaning. The first argument is legitimate, while the second is invalid and is a fallacy. It is this latter type of reasoning that Beardsley is strongly opposed to.

2.2.3 Objections to Beardsley's speech act theory of literature

We can clarify Beardsley's view further by answering some representative objections before we continue with the issue of interpretation.

George Dickie, though sharing Beardsley's anti-intentionalism, is not satisfied with the speech-act argument and points out two problems (Dickie & Wilson, 1995, p. 235; Dickie, 2005, p. 509). The minor problem is that the argument applies only to fictional works. However, Beardsley's *complete* argument works with *all* literary works with respect to the relevance of authorial intent, as I just showed.

Dickie's major criticism is that Beardsley is not addressing the real debate, which is on locutionary meaning rather than illocutionary meaning. This criticism is doubtful for the following reasons. First, it seems that illocutionary meaning does figure in the real debate. For example, Beardsley's major rival E. D. Hirsch implicitly raises the question about illocutionary meaning when asking whether the statement "Pass the salt" is an order, a command, or an entreaty (Hirsch, 1978, p. 71). Moreover, if we accept that irony, allusion and metaphor take place at the illocutionary level, a matter that participants in the debate argue over, they constitute strong counterexamples to anti-intentionalism (New, 1999, pp. 104-105).[27]

Second, what kind of "meaning" is the real debate about? Christopher New distinguishes four levels of meaning figuring in the debate: (1) the vocabulary level (the correct words used); (2) the locutionary level; (3) the illocutionary level; (4) the suprasentential level (themes and theses). Dickie's favorite example of misspeaking (Dickie & Wilson 1995, p. 236, 244) is at the vocabulary level; Mary Sirridge (1978) engages anti-intentionalism at the same level. Noël Carroll (2013) defines the meaning of a work as its "themes and theses," which corresponds to the suprasentential level. Peter Lamarque (2002) also thinks that the

---

[26] The term is used here in its broadest sense, including biographical information of various kinds and reports of authorial intention.

[27] Irony is an especially popular topic in interpretation. See Nathan 1982, 1992, Sirridge 1978 and Tolhurst 1979. As discussed, Beardsley makes great efforts to show that irony is not a counterexample to anti-intentionalism.

debate should be engaged at this level. All this shows how divided the opinions are. The most we can say is that locutionary meaning is undoubtedly part of the real debate, but it is not *the* debate.

Third, an important point: Beardsley does not abandon the discussion of verbal meaning, but addresses it indirectly. Since the meaning of a word or phrase will affect which illocutionary act the sentence in which it occurs reflects, such determination of meaning is done by considering which choice of meaning contributes most coherently to the illocutionary act reflected in the text (1982, pp. 193-194).

Carroll (1992, pp. 103-109) presents four attacks on the soundness of Beardsley's argument. The first is that Beardsley is wrong in claiming that all literature is fictional, which is apparently a misunderstanding of Beardsley on Carrol's part.[28]

Second, not all assertions in fictions are representations of illocutionary acts. For example, in many fictional works there are passages about scientific knowledge, history, or philosophical doctrines. It is inappropriate to say that in these cases the author is always pretending to assert.

Carroll is mistaken in considering only intentional representation. Recall that Beardsley's reason for treating a fictional work as an illocutionary-act representation is that there are always infelicity conditions of illocutionary-act performance present in such a work. In that case, the literary utterance tends to end up with a representation. As said, this includes both cases of intentional and nonintentional representation. It may be true that sometimes the author indeed aimed to assert instead of merely pretending to assert; however, the act eventually ends up with a representation, because there are always marks of fictionality that remove its illocutionary force.

The third objection has it that the thesis projection in a serious literary work, seen as a performance of an illocutionary act, is not always pretended. It would be absurd to say that when Sartre wrote *Nausea* he was just pretending to suggest philosophical ideas.

Again, the real reason for treating Sartre's work as representation is that his utterance satisfies the conditions of representation, not that he *must* have the intention to represent. To reject Beardsley's claim that every fictional work is a representation, what one should do is

---

[28] Carroll in Footnote 25 notices that Beardsley 1977 recognizes that there are nonfictional works. The recognition is not considered because it undercuts the argument in Beardsley 1982. The problem is that the 1982 article never makes the claim Carroll attributes. From my presentation of Beardsley's theory it can be seen that Beardsley has been rather consistent in making a distinction between fiction and nonfiction.

reject the obtaining conditions he identifies, not reject the assumption that all authors must have the intention to represent in every case, which Beardsley also denies.

The fourth objection is that convention cannot settle all interpretative inquiries. Questions about character construction and its function in the overall design of the work can be answered only by appeal to the real author, rather than to the implicit speaker who can be understood only via conventions.

This is a big point of debate. Beardsley could maintain that appeal to authorial intent is still dispensable because textual evidence is sufficient for us to consider character construction and function. If we really need a notion of the author, one based on textual evidence will still be an option for the anti-intentionalist. Defending Beardsley, Daniel O. Nathan (1982, pp. 254-255; 1992) basically holds this view when he draws on the concept of the *ideal author*, whose nature is completely based on textual evidence.

Dickie's and Carroll's main concerns are with interpretation, while Joseph Margolis's target is the speech act theory itself. Margolis criticizes a series of speech act theorists, including Beardsley. He specifies four confusions in the pretense theory of literature (Margolis, 1979, pp. 45-48).[29]

Firstly, verbal imitation is not itself a speech act but an adverbial qualification of it, better described as "mimickingly." Therefore, Beardsley is wrong to say that the verbal imitation is a sort of speech act. Nevertheless, just as we do not say that a mime climbs a ladder mimickingly, so we do not say that a stage actress speaks mimickingly. We say that she *imitates* the speaking of the fictional character. To imitate or to mimic is verbal rather than adverbial in either case. If the stage actress is just speaking mimickingly, what kind of speech act is she performing? Margolis does not say. And verbal imitation still seems to be the answer.

Secondly, not all categories of poetry are fictional discourses, and we cannot even assume that all literary writers have the intention to pretend when composing. These criticisms are again misrepresentations of Beardsley's view that I have already clarified.[30]

Thirdly, based on the first and second objections, it would be better to understand literary distinctions in terms of styles or genres rather than as speech acts.

---

[29] Though Margolis mentions Beardsley and Richard Ohmann when presenting his criticisms, he addresses the latter most of the time. However, since he classifies the two in the same camp, I see no problem in treating all of his attacks as directed at Beardsley as well.

[30] A point to note is that Beardsley sees didactic poems as imitations of arguments: they are philosophy fiction and a subclass of lyric poems (1970: 59-60; 1981b: 303).

Surprisingly, Beardsley (1979) had tried to show that the concept of style can still be accommodated in the speech act model. He distinguishes between dominant (primary) and subordinate (secondary) illocutionary acts (either real or fictional). For example, Caesar's famous triple locutionary act "I came, I saw, I conquered" not only makes three dominant illocutionary acts of assertion, but also makes a subordinate one of asserting that, for him, to arrive was to act. This latter act is subordinate to the other three. Subordinate illocutionary acts can be multiple, as it is possible that one single illocutionary act is accompanied by several implicit ones.

Drawing on this distinction, a stylistic feature is thus defined as a feature of any linguistic form that enables a subordinate illocutionary action. Difference in style means difference in meaning; sameness in meaning means sameness in style. The view that authors can express the same thing in different styles is faulted. To say something in a different way is basically to say that thing plus something else, because a stylistic feature is tied to subordinate illocutionary acts (1981b, p. 306).[31]

As for the concept of genre, Beardsley hints, but does not clearly advocate, that a genre could be defined in virtue of the style commonly found in it (1979, p. 150). That is, a genre G can be defined by a set S of which certain linguistic features are members.

Margolis's final objection is that the mimetic theory confuses imitating a speech act with speaking poetically, and with imagining someone speaking. Let us consider the first kind of confusion. Margolis claims that there is no reason to suppose that it is not the author herself speaking poetically. But the point is, as long as any marks of fictionality are discerned in one's utterance, those marks qualify the utterance as a representation. That is how Beardsley draws a line between fictional works and nonfictional ones, and also between someone who is imitating a speech act and someone who is speaking poetically.

Margolis claims further that there is also no reason to suppose that the author is reporting anything about any speaker: "he is simply speaking, though not necessarily speaking simply." This "simply speaking" oversimplifies speaking. If the speaker is not speaking as herself, she is then representing someone else speaking, that is, the implicit speaker in the story.

The second sort of confusion is a key point. According to Margolis, it is wrong to conflate "reporting what an imaginary speaker 'has said' (granting the fiction) and creating a poem in which an imaginary speaker speaks." Instead of representing an illocutionary act, what an

---

[31] Beardsley (1979, p. 166), however, acknowledges two exceptions in which different expressions result in sameness of meaning: (1) stipulation or definition; (2) elegant variation.

author really does in telling a story is to imagine a fictional world in which characters speak to each other. To imagine this world is a distinct act, but it is not a speech act. Margolis concludes that "the fictional use of language concerns the *presuppositions* that inform a set of speech acts, not those speech acts themselves."

Now, writing a work (be it fictional or nonfictional) necessarily involves intentional text-production, and in performing this very act the author intends to say something. In that case, it is hard to see why this act is not itself a speech act. The author of a fictional work creates or imagines a world by representing what a fictional narrator has said. If she just imagines a story in her mind, that does not involve any fictional use of language; once she outputs her imaginings into words, this very act is suitably perceived as a speech act. Put another way, in writing a fictional work, the author performs a speech act that informs us of an imagined world.

Margolis's objection reflects a fundamental disagreement in the debate on literary interpretation: whether writing a literary work counts as an utterance or not. If the answer is positive, producing a literary work will no doubt be a speech act. I will not pursue the issue here.

2.2.4 Interpreting fiction and nonfiction again

Once we treat a fictional work as an illocutionary-act representation, our interpretative task becomes that of *constructing* the act likely to have been performed, based on textual clues (1981b, pp. 307-308). *Re*construction in this case does not make sense, for the illocutionary act in question is only pretended, one that falls short of being; there is no actual act to be reconstructed at all. Rather, our goal is to construct the act by relying on textual evidence to supply its missing conditions.

By contrast, the *reconstructionist* position can be described as follows:

An ordinary text […] belongs to a set of circumstances, or an encompassing situational context. If we are in doubt about what it says—that is, about exactly what illocutionary actions were being performed in the act of composing it—our task is to reconstruct the situational context and examine the act-generating conditions that were present. Though we can always in principle distinguish between the question: Which illocutionary actions did the author perform? And the question: Which illocutionary actions did the author *intend* to perform, we are often mainly interested in the latter […] if we take a poem as a text that is

to be interpreted in this reconstructionist fashion, we will study the letters of the author and the memoirs of his friends. (1981b, p. 307)

Once we take this stance, the interpretative interest typically shifts from textual meaning to authorial meaning; i.e., we would be more interested in the illocutionary act the author intended to perform than in the one she did perform. This is because an illocutionary act is an act of communication, and typically the audience's goal is to grasp what the utterer actually intends to say. In that case, research into biographical data and other types of external evidence becomes necessary.

Moreover, in explaining how an illocution acquires meaning, Beardsley is already committed to the view that contextual factors, excluding authorial intention, play a role in this regard. It is convincing to conclude that with respect to nonfictional works Beardsley leans toward contextulaism. If all this makes sense then any contextualist attack on conventionalism in this respect misses the target.

2.3 Chapter conclusion

"Let the work speak for itself." This slogan aptly represents the spirit of anti-intentionalism, the greatest merit of which is its shift of focus from the author to the work. The work is its own judge of meaning; the authority of the text is the work itself, not its author. The extreme versions of intentionalism deprive the work of its own life and make it a puppet manipulated by the author.

However, the contextualist claims that Beardsley has gone too far, as his opponents do also. The work does have meaning-determining power to some degree, but it is not fully self-sufficient as he claims. Context also plays a role in the shaping of meaning. Just as contextualism plausibly demonstrates, many relational properties obtaining at a work's creation are crucial to its identity and content. Failure to consider these in interpretation may result in misidentifying the work or missing out on its important features.

Above is the orthodox characterization. However, as briefly suggested, Beardsley's position is compatible with a contextualist framework. In particular, his later position on nonfictional works is plainly contextutalist. How to interpret Beardsley is an issue I do not have space to pursue. It is enough for the reader to know the orthodox characterization because that more or less paves the path for later developments in the debate. Following Beardsley's main ideas about aesthetic satisfaction and the irrelevance of authorial intent, the

value-maximizing theory makes its debut as an improved species of anti-intentionalism. I examine this position in detail in the chapter to follow.

3. Value-Maximizing Theory

"An aesthetic interest concerns itself primarily with the most aesthetically rewarding reading that may be put upon the words used, and this reading may well be different from the one that the author intended his work to receive."

—Stephen Davies, "The Aesthetic Relevance of Authors' and Painters' Intentions"

The main claim of the value-maximizing theory is that the primary aim of literary interpretation is to offer interpretations that maximize the value of a work. The present chapter examines two oft-discussed versions of the maximizing view and related objections to them. The version proposed by Alan H. Goldman will be discussed first because it bears more resemblance in form to anti-intentionalism standardly interpreted. After that I move on to Stephen Davies' version which pays more attention to the role played by context. Finally I critically discuss the most important counterexample to the maximizing position—the movie *Plan 9 from Outer Space*.

3.1 Goldman's value-maximizing theory

3.1.1 A brief overview

With respect to literary interpretation, Goldman holds that the interpretations to be favoured are those that not only are suitably constrained, but also make the work seem better than worse so as to give the reader rewarding experience (Goldman, 1990; 1995, pp. 93-130; 2013). Goldman writes:

The task of an art or literary critic is to be an interpreter, not simply to evaluate on the one hand or to describe or pick out random properties of works on the other, but to select those properties that are value relevant and to guide the audience to appreciation and proper evaluation by showing how those properties are relevant, how they contribute to the values of the works. (Goldman, 2013, p. 23)

Goldman characterizes interpretation as a certain kind of *inference to the best explanation.* What the interpreter does is to explain how the features of the work contribute to literary value. The explanation is best when it successfully enhances, satisfies, or maximizes, the value of the work.

Like most analytic philosophers engaging in the present debate, Goldman is not an anti-realist in regard to literary interpretation. That is, he agrees that there exist objective constraints on interpretation. A significant part of Goldman's account is devoted to elaborating the nature of these constraints.

An inspiring fact noted by Goldman is that interpretations of visual art and music do rest on objective data. To be precise, interpretations in those contexts are based on descriptions that can be universally agreed upon, and this agreement comes from our directly perceiving the work's parts. Normally we do not dispute about coloured shapes and lines in paintings; nor do we dispute about notes and their formal relations in music. This is because we directly perceive them: they are perceivable elements of the work, and can be described without being interpreted.

The matter is somewhat complicated by the fact that lower-level interpretations can also invite universal agreement and thus result in true descriptions. In this case, it seems appropriate to see these descriptions also as part of the objective basis for interpretation. In the case of visual art, many representations can be universally recognized. For example, although it will need some amount of interpretation to know that a painting is a rural landscape, this conclusion is not usually a point of debate. That is, directly perceiving the pigments and shapes in the painting is not quite enough; a minimum degree of reasoning is required for one to reach the conclusion that the combination of perceivable elements results in a meaningful formation, that is, a rural landscape. Goldman calls the first case of description *noninterpretive descriptions*, and the second case *interpretive descriptions*. Both are subclasses of description and constitute the objective data for interpretation.

In like manner, literature also provides an objective basis from which we can form true descriptions. According to Goldman, this basis would be the *text*, which is defined by standard lexical meanings at its time of creation. When we read a literary work, we directly grasp the literal meanings, or the *standard semantic contents*, of the words that make it up. Descriptions based on this basis will then be true and undisputed. For instance, the statement that Hercule Poirot's head is egg-like is a noninterpretive description based on the text's standard meanings, that is, its literal assertions. And the statement that the murderer in *Murder on the Orient Express* finally goes scot-free is an interpretive description all readers

will agree upon, inferred from the denouement of the story. Any interpretation that denies both will hence be unacceptable. Simply put, interpretations should not violate true descriptions of what is in the text, which is the objective constraint on interpretation.

A second moral from nonliterary art is that it is common and natural for the interpreter to go beyond the author's intention. A musical performance, when seen as an interpretation of the composer's work, is normally not constrained to the composer's intention, and is given more space for incompatible performances. This includes anachronistic performances; for example, a classical piece played in a modern way. If we agree that interpretation of nonliterary art can be a paradigm for literary interpretation, then we should allow multiple, even incompatible interpretations of a single work.

A problem now is how Goldman explains whether incompatible interpretations can both be true. According to him, we can speak of truth only relativistically. An interpretative proposition is true if and only if it is within an acceptable interpretative scheme. Otherwise it is false. Fictional truth is hence relativized to the fictional worlds created by different interpretations.

The important assumption behind Goldman's account is this: the aim of interpretation is to achieve full appreciation of a work by grasping its values. This is done by licensing multiple-interpretability. Because there exist different sets of values in a work to be brought out by different interpretations, these readings are all said to be value-enhancing. But this does not mean that the account sacrifices critical objectivity for total interpretative freedom. The former is saved by recognizing the universally agreed elements of a work which function as constraint on interpretation.

3.1.2 Goldman and Beardsley

At first glance Goldman's theory is modelled after Beardsley's conventionalism: both of them put great emphasis on aesthetic satisfaction and linguistic constraints on interpretation. However, their accounts differ in some subtle aspects, which are reflected in Goldman's criticisms on Beardsley.

Goldman criticizes Beardsley on two points: not only is Beardsley's view of meaning-disclosure too narrow, but it does not apply to all categories of literature. As Goldman sees it, interpretation comprises a variety of activities and disclosing meaning is just one among many. For the most part we do not need to discover standard meanings of the text because we just grasp them directly, as argued earlier. How much meaning-disclosure is involved also

depends on which category of literature we are interpreting. In the case of novels, we normally do not interpret the meanings of sentences and phrases. In the case of poetry, meaning-disclosure does feature more prominently, but only as part of a complete interpretation. In either case, the discovery of meaning is not equated with interpretation.

Goldman is aware that Beardsley once defined interpretation as revealing the themes and theses of a work. But this cannot avoid the difficulty just raised because (1) words and sentences still need interpretation in the case of poetry; (2) a simple paraphrase of the whole work will be a paradigm interpretation; (3) the view does not apply to nonliterary arts.

Apparently, the first objection is based on an incomplete understanding of Beardsley's view. Interpretation as revealing themes and theses is only Beardsley's narrow definition of interpretation. His broad definition includes not only interpretation in the narrow sense but also elucidation and explication. Elucidation is about determining the unknown parts of the world of the work based on the parts known, the latter about determining the contextual meaning of words. Put otherwise, interpretation in the broad sense does include revealing local meanings such as sentence or word meanings.

The second objection is misleading because there is an equivocation on the word "interpretation." It is not clear whether it means an interpretive statement or a collection of them. Beardsley does not say that a proposition stating the themes or theses of a work counts as a *complete* interpretation. To know the theme, one has to analyse certain features of the work, such as symbolic contents. A symbolic meaning is subsumed under the theme of the work, and the theme is the broader implication of a symbolism. Moreover, the determination of the themes and theses of a work has to be based on the results of explication and elucidation. In that case, a complete interpretation will never contain a simple paraphrase. In short, the simple paraphrase of themes and theses may be an interpretive statement, but it is just a partial interpretation, and will be supported and elaborated by other interpretive statements that together form a complete interpretation, which is a collection of interpretive statements.

The third objection is extremely contentious. It is not clear why a theory of interpretation must apply across the arts, given the diversity of art media and artistic expressions. Nevertheless, as said, it is also true that a theory of literary interpretation can easily be extended across the arts. The best we can say seems to be that, though Beardsley's theory of interpretation tends to focus mainly on literature, his conventionalism or anti-intentionalism appears to be global, at least in *Aesthetics*.

Goldman's intention behind his critique of Beardsley is to reject the view that interpretation aims to disclose meaning. Interpretation, according to Goldman, aims to enhance value by explaining a work's elements or features. In the main, this view is not incompatible with Beardsley's. Beardsley's early discussion on interpretation indicates that the task of explication "involves analysis" (Beardsley, 1953). To assign meaning to a certain part of a text, the interpreter needs to find the range of possible meanings of the words, construe the syntax and identify the structural relationships among elements. It seems that this task already has a certain amount of explanatory work built into it. Earlier I discussed Beardsley's aesthetic argument. Here are Beardsley's exact words about aesthetic satisfaction as the primary aim of interpretation:

> What is the primary purpose of literary interpretation? It is, I would say, to help readers approach literary works from the aesthetic point of view, i.e., with an interest in actualizing their (artistic) goodness. The work is an object, capable (presumably) of affording aesthetic satisfaction. The problem is to know what is there to be responded to; and the literary interpreter helps us to discern what is there so that we can enjoy it more fully. Now the goodness in which we take an interest (when our interest is aesthetic) is something that arises out of the ingredients of the poem itself: the way its verbal parts—its structure and texture—combine and cooperate to make something fresh and novel emerge. The words have to work on us. They work by manipulating our understanding of parts to make us experience a whole that contains something not in the parts. (Beardsley, 1970, p. 34)

This sounds like a value-maximizing account: the aim of interpretation is to enhance value and give the reader rewarding experience by explaining the features of a work.[32] No doubt Beardsley's target is still meaning-disclosure, and this is what distinguishes his theory from Goldman's, but to reach that goal it is not enough to state only the meanings of a text, because it often involves explanation of the work's elements with an interest in satisfying value. The moral is that Beardsley's view on meaning-disclosure should not be construed in a narrow way.

A major disagreement between Beardsley and Goldman is that the former holds *critical monism* and the latter *critical pluralism* with respect to the number of legitimate interpretations. Goldman criticizes Beardsley's monism for being too narrow, because

---

[32] For this reason, some philosophers label Beardsley a maximizer. See Mikkonen, 2010, p. 76.

seeking one definitive interpretation prevents us from discovering other rewarding interpretations.

The point of contention lies in how strong conventionalism should be. For Goldman, convention is not strong enough to indicate a single correct interpretation (for example, not strong enough to determine irony in an absolute sense). But Beardsley holds the opposite. This is a debate between robust and thin conventionalism, and to adjudicate between them is not a matter that can be easily resolved.

However, it is important to highlight different concerns behind these positions. For Beardsley, the aim of literary interpretation is to recover textual meaning, so the assumption that there exists a definite meaning to be recovered suits his purpose. By contrast, Goldman's concern is with the multiple interpretability of art. For him, acknowledging this point explains the fact that the ultimate differences of tastes and preferences lead to different aesthetic values, as demonstrated in the example of interpretive performances. In that case, thin conventionalism would be a better theoretical basis.

3.1.3 Some objections

Some notable objections to Goldman's view and his replies are worth considering before I present my objection (Stecker, 1991; Goldman, 1991).

According to Stecker, "maximize" is ambiguous, it could mean: makes a work as valuable as possible, in general or with respect to a particular value. If it means the former, it would be incompatible with the claim that there can be multiple acceptable readings of a work; moreover, even if there exists a best interpretation, there is still a point in giving other interpretations since they add value to the work.

However, "maximize" does not imply monism, because there might be several ways to maximize at a given time. Even if one interpretation is best for a time it may grow stale and be superseded by another. A maximizer can hold a relativist position on this point. This is why Goldman also speaks of "enhancing" instead of "maximizing."

The second objection from Stecker is this: it is doubtful that interpretations are explanations. It is strained to say an interpretive performance is an explanation. A related problem is that we do not always explain qualities: we point them out. Goldman's reply is that, to make clear why something is the case is to explain, at least implicitly. So an interpretive performance is at least implicitly an explanation.

These are only minor issues. Stecker's main issue with Goldman is this. He thinks Goldman's attempt to subsume historical interpretations (which are causal explanations) under value-maximizing ones is unsuccessful because they are not the same kind of interpretation. However, Goldman does not claim that value-maximizing interpretations cannot be causal explanations. Interpretations based on appeal to the author's realized intention are both casual and value-maximizing because the historical perspective associated with them presumably enhances the value of the work, as explained early on. The disagreement here between Goldman and Stecker is whether such appeal must be value-enhancing. This is the issue I raise with Goldman in the next section.

3.1.4 A paradox of value

The merit of Goldman's account, some may think, lies in his reconciling multi-interpretability with objective constraints on interpretation. Nonetheless, the major problem of his theory comes exactly from this reconciliation.

To incorporate author intention into the maximizing account, Goldman acknowledges that interpretations based on successful author intentions are also value-enhancing, for guiding the reader to experience the work through the author's eyes is a historical value to be realized. This acknowledgement has an odd implication: any aesthetically inferior interpretation, when suitably constrained, would become value-enhancing as long as it is intended by the author. In other words, if the author's own explanation of the work reveals no or very little value in it, on this view we still have to claim that the author's explanation is value-enhancing and hence acceptable.

The problem here is that the historical value Goldman has in mind seems to be of a different character from the value to which the elements of a work can contribute. Put another way, it is problematic to equate historical value with literary or aesthetic value. We can well imagine a case in which the author's intention provides us with a shallow reading of the work when another reading is available that makes the work aesthetically better. Even when the intention in question is compatible with textual evidence, it would still seem odd to say the shallow interpretation enhances the value of the work. The fact that the effort is the result of an impoverished artistic scheme does not add much value to the work itself. It is true that recognizing the intention in question will change our reading experience and engagement with the work, but it still seems strange to say that this reading enhances the work's value. Undoubtedly, knowing authorial intent gives us a basis for evaluating the author's

achievement, but it does not make the work more valuable, or at least not so much that can be called value-maximizing.

There is no denying that recognizing author intention helps the reader appreciate the work more fully in the sense that she can understand why the author made the text the way it is. Even so, it seems wrong to say that appreciation is always based on grasping aesthetic value, which appears to be Goldman's core claim. Explaining that the author in the above example intended to produce schlock helps the reader appreciate the film more; but this explanation itself fails to make the work aesthetically better in any way. The point is that there is a difference between making the work more valuable and making the reader's experience more valuable. Recognizing authorial intention achieves the latter but not necessarily the former. And this distinction is not clearly drawn in Goldman's account.

The motivation behind the consideration of historical value is to accommodate authorial intention as one of the possible readings that enhance value. However, if what I have argued is right, this move goes so far as to include aesthetically inferior interpretations. A safer strategy would be to say only that the author's intention can very often help us identify one of the rewarding interpretations the work can sustain and yet this is not always the case. Nevertheless, this tack would apparently undermine Goldman's account of aesthetic value.

The inclusion of authorial intention in Goldman's theory of interpretation does seem unfit, because this means that interpretation as inference to the best explanation should be understood as two kinds of explanation: causal explanations and value-maximizing explanations. The criteria for what counts as best differ in these cases. It would be more adequate to speak of just one kind of explanation if we want to use the term in a more coherent sense.

In short, either causal explanations are a subclass of maximizing explanations, or they are contrasted with them. It turns out that the former results in accommodating interpretations that do not really make the work more valuable. This shows that Goldman's attempt to subsume intentionalism under the broader framework of anti-intentionalism is not successful. The upshot is that intentionalism would have to stand alone.

3.1.5 Conclusion

In concluding, there are two things to say about Goldman's view. First, the attempt to subsume causal explanations under value-enhancing ones shows that, even for an anti-intentionalist in the broad sense, like Goldman, authorial intention still plays a role not to be

ignored. This casts doubt on any theoretical attempt to dispose of authorial intent entirely. Second, Goldman will need to confront contextualist criticisms since he takes the text to be the sole standard for interpretation.

Goldman's writings indicate that he approaches the problem of interpretation in the normative direction rather than in the metaphysical direction. Such being the case, he may not be pressed to answer questions about work-meaning and work-identity.

Nevertheless, it remains to be seen whether the value-maximizing theory fares better when based on an account of literary ontology. This brings us to the version developed by Stephen Davies.

3.2 Davies' value-maximizing theory

3.2.1 A brief overview

Like Goldman, Davies claims that interpretation should aim to maximize the work's value, on condition that the interpretation is suitably constrained (Davies, 1982; 2006, pp,122-125; 2007). An initial question concerns what such constraint is.

The answer lies in contextualism, which maintains that a work takes its identity and meaning partly from the circumstances of its creation. Interpretations should be constrained to the extent that the interpreter locates the work in the context in which it was produced. Such being the case, anachronistic interpretations and interpretations that do not respect the work's identity are out of place. The fundamental premise of interpretation is hence to interpret the work *as authored*.[33]

The contextual factors that affect work-identity include, but are not limited to, a work's title, genre, style, and facts about the author's oeuvre, social elements to which the work makes reference, and the artistic and linguistic conventions in place when the work was produced (Davies, 2006, pp. 68-71). Given this contextualist model, interpretations should not violate the above identity-conferring features; otherwise, the interpreter would just end up interpreting a different work.

---

[33] Normally this means "as produced in a specific place and time." However, contextualists disagree on whether the author's identity is part of her work's identity. The debate on contextualism is another topic in the ontology of art, and there is no space here to discuss it in detail. See Gracyk, 2009.

What is crucial here is that different kinds of authorial intent figure in the relevant contextual factors and will be taken into account by contextualists. For example, an author's intention to select the title or category of her work matters because it indirectly affects what the work will mean.[34] Since Davies accepts these kinds of intention, the author's intention is relevant at the level of literary ontology but not at the level of interpretative policy (Davies, 2007, p. 15).

Contextualism in literary interpretation is usually linked to the *utterance model*, according to which a work is seen as an utterance, that is, a time-indexed or contextualized word-sequence.[35] Different theories of interpretation seek to identify utterance meaning in different ways. The maximizing view focuses on what an utterance *could* mean.[36]

Considering utterance meaning, the role played by linguistic conventions is a key one. Unlike Beardsley, who thinks that linguistic conventions in most cases lead to one single interpretation of a work, Davies's view of convention is more moderate. Thin conventionalism is adopted here: based on linguistic conventions and other contextual constraints, a work can sustain more than one interpretation. By this standard, any interpretation that can be put on the work counts as legitimate. To adjudicate between competing interpretations, those that present the work in the best light win out.

A common misconception can be duly corrected here. By the word "those" it can be seen that the maximizing view is not committed to critical monism, which maintains that there can be only one single correct interpretation for a work. Given that there are different ways to confer different values on a work, it is sometimes difficult to say that one way rather than another results in an artistically better reading. Namely, we may have more than one interpretation that can be said to maximize the value of a work equally. The maximizing view hence supports critical pluralism. As shown, all these points are shared by Goldman's version.

The motivation for the maximizing view is aesthetic satisfaction. Davies agrees with Beardsley that literary interpretation seeks to offer such satisfaction. Although the actual intentionalist analogizes literary interpretation to the interpretation of ordinary discourse (see

---

[34] For the power of titles, see Levinson, 1985; for categorial intention, see Walton, 2008, pp. 195-219.

[35] As early as Beardsley's *Aesthetics*, literary works are seen as utterances. But the term "utterance model" is not coined until Meiland, 1981.

[36] "Could" here might mean "relative to our evidence or what we currently know, this is a possibility." (Stecker, 2003, p. 66) This epistemic sense of "could" makes sense in the context of literary interpretation since critics are often working with incomplete evidence. However, it is not clear whether this is the sense adopted by Davies. See the next footnote.

Chapter 4), Davies thinks that this analogy is mistaken. While the latter aims for communication, the former aims for aesthetic satisfaction. Such satisfaction is achieved by exploring rewarding interpretations that can be put upon the work. By reconciling the maximizing criterion with critical pluralism, the maximizing theory does a better job than Beardsley in bringing the aesthetic consideration into full play.

At this point, Davies needs to say something about the truth conditions of incompatible interpretations. His suggestion is that an interpretation is true if it is true to the work, that is, if it deals with one of the meanings the work can sustain (Davies, 2007, p. 202). Since Davies' view is based on contextualism, the truth makers for interpretation will then be the text, contextual factors plus linguistic and art conventions.

Needless to say, usually what the author intends to say in the work coincides with one of the most rewarding readings identified by the maximizing view. In this case, author's intentions are *recommendations* that the work can be read one way rather than another. Considerations of them give us clues of what the most rewarding interpretation may be. But since this move is in principle dispensable, it is wrong to claim that the author's (semantic) intention is essentially relevant to interpretation.

I conclude this overview by noting the major differences between Davies' and Goldman's theory. First, the former has a contextualist basis, which means, when considering legitimate interpretations, the contextual factors present at the time of the work's creation will be taken as a standard. However, in Goldman's theory the object of interpretation is a context-free text. Second, Goldman accommodates intentionalism by acknowledging the value realized by causal explanations, while in Davies' account the author's intention is relevant in the sense that it determines some of the identity-relevant properties such as title and category. While both theories are typically classified as anti-intentionalism, they are pro-intentionalism in different ways.

3.2.2 Some objections

I will consider existing objections and replies before discussing my reservations about Davies' theory in the next section. Most objections here apply to Goldman's theory, too. I begin with two minor objections (Davies, 2007, pp. 187-188).

The first objection holds that the constraints imposed by the maximizing view leave little room for value-maximization. Once we eliminate the interpretations that do not respect the work's identity, the meaning allowed by conventions is usually clear and unique. In reply, the

maximizer says that what is clear and unique is the surface meaning of the work (Beardsley's primary meaning). Given linguistic conventions in place, it is true that in many cases there is not much dispute about what the sentences that make up the work apparently assert. However, interpretation does not merely aim for surface meanings. More often than not it is the implicit meanings that are in need of interpretation. Work-meaning becomes more difficult to pin down as interpretation descends from sentential level to suprasentential level—the themes and theses of a work. At this level there is abundant room for dispute, and hence for different readings that reveal different values.

The second objection has it that literary interpretation calls for understanding rather than just maximizing a work's value. It is not that we are interested only in enhancing the merits in a work; we are also interested in interpreting it with critical eyes. As already mentioned, the contextualist model requires that interpretation be consistent with preserving the work's identity. This requirement does justice to the work. That is, we should want to see the work in its best light, which is what its author should want for it. But even viewed maximally in this way, it might turn out not to be very good.

Now, the next two related objections bring out a worry I will soon reveal. The first objection claims that all meanings that can be put on the work must have been intended consciously or unconsciously, because there must be a logical connection between what one intends and what one does. This objection neglects the possibility of failed intentions. But since the maximizing view is interested in what an utterance could mean, it is the *appearance of intentionality* of an utterance that merits our attention, not the intention realized in it (Davies, 2006, p. 124). This concept is related to the next objection.

It is argued that some features of a literary work must be analysed intentionalistically, such as irony, allusion, metaphor, symbolism, allegory, etc. The maximizer can counter this objection by dealing with intentions more sophisticatedly (Davies, 2007, pp. 117-118). If the relevant features are identity-shaping, they will be respected and accepted in interpretation. In this case, any interpretation that ignores the intended feature ends up interpreting a different work. But if the relevant features are not identify-conferring, more room would be left for the interpreter to consider them. The intended feature can be ignored if it does not add to the value of the work. Or, as said, the interpreter can interest herself in the appearance of intentionality: where such a feature is not intended but can be put on the work, she can still build it into the interpretation if it is value-satisficing.

Now, the basic premise of the maximizing view is that a literary work is an utterance. If there is an utterance there must be an utterer. Since the replies to the last two objections

appeal to the appearance of intentionality, it seems that an *apparent* utterer has to be presupposed. Otherwise, the agent of intentionality would be lost. In other words, one cannot talk about an apparent intention without detaching it from the agent of it.[37] In that case, the value-maximizing theory is in danger of collapsing into hypothetical authorism to be discussed in Chapter 6. Though this worry does not threaten Goldman's version, he will need to handle the contextualist criticisms which Davies' version is free from.

More will be said about this point in Chapter 7 when I discuss how theories of interpretation based on the utterance model are related to an apparent utterer. Now I turn to my reservations about Davies' maximizing view.

3.2.3 Recovery and imposition

As said, the value-maximizing theory seems to have derived mainly from Beardsley's conventionalism. Not only do both put emphasis on aesthetic satisfaction, but also they see conventions as objective constraints on interpretation. There is a notable divergence between these two versions of conventionalism that does not seem to receive much attention in the debate. The core idea of Beardsley's conventionalism is that convention determines work-meaning. The word "determine" here is used in an ontological sense: convention constitutes meaning. Carroll calls this issue the *constitutive* question: what constitutes the meaning of a work (Carroll, 2013, p. 10)? Stecker has another term for it: the work-meaning issue: what is the meaning of a work? (More on this in the next chapter.)

Basing a theory of interpretation on an account of meaning is a common way to support the alleged interpretative aim the theory advocates. For example, actual intentionalism claims that the author's successful intention constitutes meaning. In that case, the problem about how we should interpret a work is automatically resolved: since meaning is determined by intention, interpretation should aim to discover or recover intention. This may sound like an is/ought fallacy, but the strategy has been widely adopted in the present debate and not everyone regards this move as fallacious in the context of literary interpretation.

Anti-intentionalism also takes the same tack by affirming the constitutive status of convention instead of intention. Once we accept that convention determines meaning, it is

---

[37] I am inclined to think that the epistemic sense of "could" does not commit one to talk of an apparent intention. The referent in "what an author could mean" in this sense is still the actual author. However, an apparent intention implies an apparent speaker, for it concerns not a possible intention but a make-believe intention.

natural for us to accept that interpretation aims to recover what is implied by convention. Interpretation is said to be a practice of retrieval when guided by the constitutive perspective, because it retrieves for the reader something that is already present but implicit in a literary work.

According to Carroll (ibid., p. 12), it is unclear how the maximizing view answers the constitutive question. It does not seem to make much sense to say that the meaning of a work is determined by the interpretation that grants the reader maximal aesthetic satisfaction. If retrieval is not the ideal interpretative practice the maximizer envisages, then what kind of activity will it be? If the maximizing interpretation is not constitutive of meaning, what is its relation to the work?

Where interpretations do not uncover something already there, they may be better treated as things we ascribe to, or impose on, a work. In other words, interpretations in this case are *impositions*, and the maximizing theory can hence be described as holding an impositional view of interpretation.[38] This construal seems apt since the interpretations acknowledged by the maximizing view are those that can be "put on" a work.

In Beardsley's eyes, impositions—or superimpositions in his term—are not interpretations proper (Beardsley, 1970, pp. 42-44). Though what he has in mind when referring to impositions are the so-called "theory-driven" interpretations such as Marxist, psychoanalytic, feminist readings, he does explicitly mention that an imposition is what a work can be "taken as" and that it does not "bring out of the work something that lies momentarily hidden in it" (ibid., p.44). Interpretation aims to recover what is objectively there, rather than impose what we think can be there. Once we agree that something constitutes meaning and is recoverable, it is hard to see why interpretation should aim for imposition rather than recovery.

Although Davies does not seem to offer a thorough theory of utterance meaning, what he thinks about how authorial intent contributes to meaning appears to be at least partly similar to Beardsley, who allows that intentions can be successfully realized in a work, and that intentions alone are insufficient to constitute meaning in the absence of conventions. Davies also acknowledges that the author's intention can be successfully realized in a work and in that case authorial meaning will be one of the legitimate readings generated by convention (Davies, 2007, p. 200, 202).

At this point, two conjectures about meaning constitution in the maximizing view can be attempted. First, conventions alone are sufficient to determine meaning, which is the core

---

[38] For a discussion of recovery and imposition, see Lamarque, 2007, pp. 152-157.

idea of Beardsley's robust conventionalism. Since robust conventionalism in theory supports the claim that there can be only one single interpretation generated by conventions, it is hard to see that this position will not be at odds with the maximizing view's moderate conventionalism, which allows multiple interpretations to be imposed on a work. It is puzzling to say that an utterance has one single meaning constituted by convention and multiple meanings allowed again by convention. If by convention the interpreter can reach one unique meaning, it would be impossible to come up with more meanings to impose by the same standard.

The second conjecture seems more plausible: utterance meaning is a function of authorial intent and linguistic convention. This is exactly the position of moderate intentionalism (see Chapter 4), the main claim of which is, roughly put, that the intentions to be taken as constitutive of meaning are only those successfully realized in the work. This position, unlike robust conventionalism, leaves some room for textual ambiguity, and it is hence feasible to impose multiple interpretations.

Now the maximizer may respond that a theory of interpretation is not under an obligation to answer the constitutive question. That is, the question of meaning constitution can be completely jettisoned, or at least ignored. That being the case, the strategy of basing a theory of interpretation on a definition of utterance meaning is abandoned. This tack is different from that taken by the maximizing view's two major rivals, actual (moderate) intentionalism and hypothetical intentionalism, both of which approach the issue with an answer to the constitutive question. Precisely speaking, moderate intentionalism claims that the proper aim of interpretation is to recover successful intention, because utterance meaning is determined by such intention; hypothetical intentionalism claims that the proper aim of interpretation is to reach the best hypothesis of the author's intention, because utterance meaning is determined by such hypotheses. Nonetheless, the maximizing view does not claim that the proper aim of interpretation is to enhance a work's value for the reason that utterance meaning is determined by the value-satisficing reading.

A fundamental disagreement here is whether recovery trumps imposition in interpretation. Since what interests the maximizer is what an utterance could mean instead of what it means, an impositional view of interpretation would be more appropriate. But people supporting Carroll might object that a theory of interpretation appears incomplete if it fails to offer an answer to the constitutive question. Something seems missing when we can explain what an utterance could mean and yet are silent on what it means.

The maximizer might maintain that my second conjecture does hold and in that case the impositional view does not exclude meaning recovery. If the maximizing interpretation gets to what the author intended, it would be reasonable to describe what is going on as recovery. But once an account of utterance meaning is offered, the general difficulty posed by Beardsley emerges: if we have an interpretation that brings out what lies hidden in a work, why can impositions that cannot do so achieve equal force? Put otherwise, once one has an answer to the constitutive question it is hard to see why imposition weighs as much as constitution.

But this issue cannot be easily resolved, since one can always separate normative issues from semantic ones. A theory of interpretation can be either *hortatory* or *semantic* (to borrow Andrew Huddleston's phrases, see his 2012, p. 244); the former concerns an interpretative policy while the latter concerns a theory of meaning.

Assuming that we have a good reason to accept the impositional view of interpretation, there is still another problem to be addressed. The maximizing view is unclear about what counts as the kind of interpretation an interpreter should really pursue. At some places, Davies maintains that all meanings that can be put on the work count as legitimate or true interpretations (Davies, 1991, pp. 35-41; 2007, pp. 17, 191-197, 198-206). Put another way, as long as an interpretation that respects the work's identity does not violate textual constraints, it counts as a true interpretation, and there can be many. At other places, however, a mandate is added to pick out the most rewarding interpretations from the true ones (Davies, 1982, p. 65; 2006, p. 122; 2007, pp. 15, 151, 166-190; Stecker & Davies, 2010, pp. 307-308; 2014, p. 13). This means that the interpretations that satisfy the maximizing requirement do not always exhaust the true interpretations of a work. That being so, one wonders why the maximizing view needs two standards for selecting what is really needed. What should the interpreter do with those interpretations that can be put on the work and yet fail to be value-satisficing/maximizing? In his earliest discussion on this issue, Davies suggests that we reject the inferior readings, but it is unclear whether this remains so in later discussions (Davies, 1982, pp. 68-69).

If we stick to the interpretations permitted by conventions, the position is better called conventionalism rather than the value-maximizing theory.[39] In that case, the label of the value-maximizing theory seems misleading unless it is maintained that every interpretation

---

[39] This is exactly the position in Davies, 1996, pp. 20-21. This paper is slightly revised and collected in the 2007 anthology, where Davies changes the label "conventionalism" to "value-maximizing theory."

that can be put on the work is value-enhancing. But if we are serious about the maximizing mandate, as the current label suggests, it appears that the constraints of imposition should be reset. For what the interpreter is after in the end is not every interpretation that can be put on the work but those that can not only be put on the work but that also maximize the work's value. In other words, the condition for a legitimate interpretation on this view comes in a conjunction: it is one that can be put on the work and is value-satisficing (this is Goldman's position). The value-maximizing theory, as it is currently presented, thus has vague delimitation on qualified interpretations. This also makes its label not as neat as its contenders should want.

In sum, I do not think these reservations are major problems that cannot be fixed. Davies' version seems quite convincing and my main issue with it is still the inevitable postulation of an implicit speaker, which is to be discussed in the final chapter.

3.3 The *Plan 9* debate

I have postponed the discussion of the most important objection to the value-maximizing theory because it deserves a separate treatment. The objection here is that the maximizing view is in danger of turning a mediocre work into a masterpiece. Ed Wood's film *Plan 9 from Outer Space* is used as an example by Carroll and it has inspired much discussion (Carroll, 1992).

Plan 9 from Outer Space, probably the worst film ever made, can become a postmodern classic if viewed from the maximizing perspective:

> *Plan 9 from Outer Space* is a cheap, slapdash attempt to make a feature film for very little, and in cutting corners to save money it violates—in outlandish ways—many of the decorums of Hollywood filmmaking that later avant-gardists also seek to affront. So insofar as the work of contemporary avant-gardists is aesthetically valued for its transgressiveness, why not appreciate *Plan 9 from Outer Space* under an analogous interpretation? (Carroll, 1992, p. 119)

The point is, in attributing the postmodern interpretation to the film for the sake of aesthetic interest, the maximizer unduly turns a bad work into a classic. Carroll calls this "aesthetic hedonism." The value-maximizing theory is a species of aesthetic hedonism, which

assumes aesthetic pleasure as the primary interest in interpretation. Robert Nozick's experience machine argument is then invoked to reject such hedonism:

> […] we have an investment in really encountering interesting and brilliant authors, not simply counterfeiting such encounters. Knowing that *Plan 9* is a schlock quickie, but responding to it as if it were superbly transgressive, is akin to knowingly taking the heroics performed in Nozick's experience machine as if they were actual adventures. (Carroll, 1992, p. 123)

This counterexample can be rejected in different ways. I begin with less promising replies. A first quick response is that the film might not offer sufficient clues for being treated as a subversive message (Nathan, 2006, p. 289; Davies, 2007, p. 187). The problem of this response is that it is not a universal solution to all counterexamples. Presumably, there are cases in which the value-maximizing interpretation is consistent with the work's observed properties and yet unduly maximizes the work's value.

A similar solution is to claim that deviations from conventions, even if intentional, do not always contribute to aesthetic value (Goldman, 2013, p. 51). The case of *Plan 9* is exactly a case in which the originality displayed is not merited. Again, this reply might be convincing in this particular case but not in all cases.

The anti-intentionalist can claim that the postmodern reading of *Plan 9* is an anachronistic interpretation not allowed by the anti-intentionalist stricture (Nathan, 2006, p. 289). This is because the relevant conventions and background knowledge should be those at the time of the work's creation. But the problem remains the same: not all counterexamples can be removed in this way.

A different strategy is this. We should not base our interpretations on false descriptions of a work (Goldman, 2013, p. 51). True descriptions of a work's properties include those of relational properties. Since an "avant-garde attack" is a relational property involving an intention, we should not misdescribe it. That is, in this particular example the avant-garde interpretation will not be a true description of the work and hence should be rejected.

This reply has two problems. The first is that it is (again) not a universal solution to the real problem raised by the *Plan 9* example. The second is that appeal to relational properties seems to create more difficulty. Since all of art's contextually relevant properties are relational, Goldman's move here would force him to exclude most, if not all, non-

contextualist interpretations. This is clearly not what he wants, since he has been arguing from a non-contextualist basis.

A more convincing solution suggests that we can be aware of Wood's intention and yet suggest a better interpretation at the same time, just as in daily conversation we often polish our interlocutor's thought by suggesting a better possibility (Huddleston, 2012, pp. 251-253). This solution will face problems if its underlying assumption is attacked: literary interpretation is akin to conversation. This controversial claim will be discussed in more detail in the next chapter when we get to Carroll's conversation argument.

As I see it, the most promising reply to the *Plan 9* counterexample is to claim that the postmodern interpretation of the film does not respect its identity (Davies, 2007, p. 187). In other words, the interpretation at hand does not interpret the work as made by Wood, whose *categorial* intention to make a B-grade film rather than a satirical postmodern film is accepted by the maximizer, who denies that the postmodern interpretation is targeted at Wood's film as such. [40] Of course, not all counterexamples have to do with categorial intention; but to respect a work's identity, all interpretations that do not identify the work in its context of production will be ruled out. As Davies, says, "The interpretation needs to be *true to the work*, not merely *consistent* with it." (Davies, 2006, p. 123, emphases original) This imposes strict constraints on the sort of aesthetic hedonism which Carroll is against.

3.4 Chapter conclusion

In this chapter I have discussed two representative versions of the value-maximizing theory; they are mainly distinguished on a contextualist basis. As shown, Goldman's version has two major problems: for one thing, he needs to deal with the contextualist criticisms; for another, he has a general difficulty in accommodating intentionalist interpretations in the anti-intentionalist framework. These challenges make his version less appealing than Davies'.

That said, there are two reasons why I hesitate in accepting Davies' value-maximizing policy for interpretation. First, appeal to value raises the difficult problem of what value is. One merit of Goldman's theory is that it is backed up by a separate and comprehensive account of aesthetic value (Goldman, 1995). Absent such an account, it is unclear whether a

---

[40] Carroll, who is also an expert on film studies, writes: "[…] all the evidence indicates that Edward Wood did not have the same intentions to subvert the Hollywood style of filmmaking that contemporary avant-gardists have […] it seems that the best hypothesis about his intentions is that he was attempting to imitate the Hollywood style of filmmaking in the cheapest way possible." (Carroll, 1992, pp. 119-120)

given interpretation counts as value-enhancing. But since aesthetic or artistic value is another big debate, we may not want to press this point too hard. Therefore, I did not raise this issue in my discussion. Second, there is a need for an apparent utterer which I think is crucial. This worry points to some deep issues about the role of intention in interpretation to be discussed from the next chapter on.

The focus of my discussion will now shift to different versions of intentionalism. I begin with the strongest contender in the debate: actual intentionalism.

4. Actual Intentionalism

"To banish the original author as the determiner of meaning was to reject the only compelling normative principle that could lend validity to an interpretation."

—E. D. Hirsch, *Validity in Interpretation*

Actual intentionalism comes in at least three forms, giving different weights to authorial intention. The *absolute* version claims that textual meaning is fully determined by the author's intention; the *extreme* version claims that the text ends up being meaningless when the author's intention is incompatible with it; the *moderate* version claims that either the author's intention determines meaning or if this fails meaning is determined by convention and/or context instead.

Below I discuss these three versions in order as well as related criticisms. My conclusion is that the odds are on the moderate version because there are major problems with the other two.

4.1 Absolute actual intentionalism

4.1.1 Authorial intent as the infallible standard for interpretation

Absolute actual intentionalism (absolute intentionalism hereafter) is notably championed by Steven Knapp and Walter Benn Michaels, who claim that authorial intention and textual meaning are identical (Knapp & Michaels, 1982, 1987).[41] They justify this relation by inviting us to imagine a case of intentionless meaning. Suppose you walk along a beach and see a squiggle in the sand. It looks like a stanza from Wordsworth's poem:

> You will either be ascribing these marks to some agent capable of intentions (the living sea, the haunting Wordsworth, etc.), or you will count them as nonintentional effects of mechanical processes (erosion, percolation, etc.). But in the second case—where the

---

[41] See Knapp & Michaels 1992 for the abridged version of their papers. Also note that their essays, along with papers discussing their views and their replies to the discussion, are collected and published in Mitchell 1985.

marks now seem to be accidents—will they still seem to be words? Clearly not. They will merely seem to *resemble* words. (Knapp & Michaels, 1982, p. 728, emphasis original)

That is, you would no longer view the squiggle as language (and hence as having meaning) if it turns out that it is caused without an agent of intentionality. The marks become meaningless as soon as you know that they are intentionless. The conclusion is, hence, that authorial intention and textual meaning are one and the same thing.

The argument for identity here is apparently invalid. What it really has proved is that authorial intention is necessary for something being language (and hence meaningful). It seems reasonable to think that if we know the marks in the sand are caused by the waves then they cease to be language at all. It also seems reasonable that, to imagine another scenario, we might doubt whether a computer has a mind and hence intentions if it randomly chooses a word sequence that can be taken as a poem. In the latter scenario, if scientists conclude that the computer in question does not have a mind then we would hesitate to call the randomly produced word sequence a real poem. On this point I think Knapp and Michaels are right, that there cannot be cases of intentionless meaning.[42]

It is less obvious whether they have successfully proved that authorial intention is sufficient for something being meaningful, because there is a confusion of terms with "intention" and "meaning." The former can mean the intention to produce a text or the intention to mean something, while the latter can refer to textual meaning or human language. Since the crux of the debate on literary interpretation is the role of semantic intention, one would expect that it is this very point that Knapp and Michaels are addressing. This seems to be so as evidenced by their mention of Hirsch and Beardsley.

The point at issue, as I see it, should be whether the author's semantic intention can be infallibly realized in the text, namely, whether the author's semantic intention is sufficient for the text to have that intended meaning. Knapp and Michaels might have intended to argue for a positive answer to the said question. However, as shown, what they really prove turns out to be that intentional text-production is necessary for being a language (and for having textual meaning, since the former is necessary for the latter). Their mistake also leads many of their

---

[42] Note that Grice (1957) allows for natural meaning. For example: spots on someone's skin mean measles; the smoke we see in a forest means fire. Here the connection is causal rather than intentional. The "meaning" discussed here excludes natural meaning.

critics to interpret them as proposing a Humpty-Dumpty theory of meaning, which claims that a text infallibly means what its author intended it to mean.[43]

A Humpty-Dumpty theory of meaning is not promising because it does not allow the indubitable fact that authors sometimes fail to say what they intended to say (Shusterman, 1992, pp. 168-169). Such a theory implies that the author can mean anything by any word or text, which is absurd. This consequence follows because the account does not distinguish between speaker's meaning and sentence meaning, or what Currie calls intentional meaning and conventional meaning (Currie, 1990, pp. 111-116). It turns out that on this view no features of the text can generate any meaning at all. But if that were true, we could not even say that we often succeed in daily conversation.

As said, it is right to say that the intention to produce a word sequence is necessary for the latter to be part of a language, which is the real force of the wave-poem argument.[44] It follows that it is impossible for us to interpret without positing any agent of human intentionality (Knapp & Michaels 1982, p. 726). Nonetheless, we can question whether the speaker required in this case has to be real instead of hypothetical (see Chapter 6).

4.1.2 Textual identity

It is of interest to see how Knapp and Michaels further advance their identity thesis. They claim that authorial intention is the criterion for textual identity. That is, not only do they argue that authorial intention is identical with textual meaning, but also that meaning is a necessary and sufficient condition for the identity of texts. If we have a different meaning then we have a different text, and vice versa. It follows that if the interpreter disregards authorial intent then she would end up misidentifying the text.

The proof starts with the rejection of the conventionalist account of textual identity.[45] How does the conventionalist explain textual identity in her picture? One appealing option is to identify the text with the word sequence syntactically construed (Goodman & Elgin, 1986).

---

[43] For example, Currie, 1990, pp. 111-116; Livingston, 2005, pp. 139-146; Lamarque, 2009, pp. 125-126; Carroll, 2016, p. 311.

[44] Beardsley (1982, p. 195) makes a similar point in claiming that intentional text-production is necessary for generating illocutionary acts.

[45] Philosophers have drawn a distinction between work and text, the former being a text considered in its context of production. This distinction is not well articulated until Currie 1991a. For an overview, see Shusterman, 2009. The distinction will be disussed later.

Nevertheless, this criterion is unsatisfactory in two respects. First, given this criterion, it is hard to see how we can still maintain an interest in distinguishing one sort of text from another that has the same word sequence. For example, suppose there are two syntactically identical texts but one is literary and the other scientific. In this case, not only do the texts have different meanings but also they are categorially different. It seems reasonable to take them as different texts, but we would not be able to do so while adopting the syntactic criterion, as it is indifferent to textual meaning (Knapp & Michaels 1987, pp. 58-59).

Second, the syntactic criterion has the consequence of allowing an infinite range of meanings. This is more obvious in cases involving short texts than long ones. For example, the word "go" means what it means in English as well as "what it could be used to mean in any language, living, dead, or yet to be invented" (Knapp & Michaels 1992, 61).[46]

They go on to claim that the conventionalist is left with two options for textual identity: authorial intent and conventional meaning. Since the first option is not available to conventionalism, it has to be the second, which works as follows. By linguistic convention, a text bears a limited and determinate range of meanings in terms of which the text can be individuated. Typically authorial meaning will be among those conventional meanings. But Knapp and Michaels suspect that this criterion is arbitrary. If the advantage of the conventional criterion is to limit the range of possible meanings, many other rules will do as well. Then why favour this one over others? The answer they can think of is that the conventional criterion provides a range of meanings plausibly related to authorial meaning. But this is not necessary because there are counterexamples. Then, they maintain, the relevance of convention can be explained only by the assumption that conventional meanings serve as clues to authorial meaning. However, if this is the case, the view that textual meaning transcends author intention would be false. The conventionalist thus faces a dilemma: either she embraces authorial meaning as the criterion for textual identity, or she opts for the conventional criterion which turns out to be incoherent. Here comes the conclusion: not only does authorial intention fully determine what a text means but also it functions as the criterion for textual identity.

In their reasoning Knapp and Michaels appear to confuse text-identity with work-identity. My understanding is that they use the term "text" to mean "this-word-sequence-as-written-by-this-person" rather than a brute word sequence; otherwise their assimilation of texts with

---

[46] It is usual, though, to index syntax to a language. Even in that case we have the phenomenon of word ambiguity in a language.

utterances would not make sense. Since they want to cover all sorts of texts, not only literary ones, the term "work" is not used probably because it could imply being a literary text. But even if we use "work" to mean "this-word-sequence-as-written-by-this-person" (and not necessarily a *literary* work), the criterion they advocate would be at best incomplete. We can well imagine a case in which two authors produce the same text with an intention with the same content. In this case, Knapp and Michaels would claim that the authors produce the same work, but this result is as unsatisfactory as that in the thought experiment on personal identity in which we call the two offshoots resulting from fission the same person, even if they are both psychologically continuous with the original person (Olson, 2017). This case shows that they neglect the context of production as part and parcel of work-identity; author intention is only part of the context. Since they assimilate texts (to return to their term) with utterances, the occasion on which a text is uttered has to be taken in its entirety in order to individuate texts. For example, it might suit their project better if they argued that the author's identity, rather than the author's intention, is part of the text's identity.

A second mistake Knapp and Michaels make is being arbitrary in drawing the conclusion that the conventional criterion ends up providing clues to authorial meaning. This conclusion is problematic, for, as Iseminger points out: "Why should not this plausible relation consist precisely *in the fact that they are all things that the author might mean in the sense that they are all things that, given the conventions of the language, the text can be used to say* […]" (Iseminger, 1992, p. 90, emphasis original)

Given the above two concerns, I do not think that Knapp and Michaels have succeeded in defending authorial intention as the criterion for textual identity. The conclusion of this section is that an absolute intentionalist position on identity and meaning is not promising.

4.2 Extreme actual intentionalism

4.2.1 Authorial intent as the necessary condition for meaning

Extreme actual intentionalism (extreme intentionalism hereafter) is defended by E. D. Hirsch (1967, 1976) and improved by Gary Iseminger (1992, 1996, 1998). Below I start with Hirsch's account, followed by modifications suggested by Iseminger.

Unlike the absolute intentionalist, Hirsch accepts the *codetermining influence* of linguistic convention (Hirsch, 1967, p. 27). According to him, any given word sequence represents an array of possible meanings constrained by linguistic convention. That is, a word sequence is

always indeterminate with respect to meaning. Only the author's intention can resolve the indeterminacy. In a key passage Hirsch writes:

> A determinate verbal meaning requires a determining will. Meaning is not made determinate simply by virtue of its being represented by a determinate sequence of words. Obviously, any brief word sequence could represent quite different sequences of verbal meaning, and the same is true of long word sequences […] But if a determinate word sequence does not necessarily represent one, particular, self-identical, unchanging complex of meaning, then the determinateness of its verbal meaning must be accounted for by some other discriminating force which causes the meaning to be *this* instead of *that* or *that* or *that*, all of which it could be. This discriminating force must be an act of will, since unless one particular complex of meaning is *willed* […] there would be no distinction between what an author does mean by a word sequence and what he could mean by it. Determinacy of verbal meaning requires an act of will. (Ibid., 1967, pp. 46-47, emphases original)

Even so, in his discussion Hirsch is inconsistent about how convention plays its role. He paradoxically claims that in some cases linguistic convention can fully determine meaning (ibid., p. 28; see also Dickie & Wilson, 1995, p. 236). This concession contradicts what he says in the cited passage. I shall ignore Hirsch's undermining of his argument; otherwise his project against semantic autonomy would just fail. That is, I will take Hirsch as accepting that, absent authorial intent, a word sequence is always indeterminate in the sense that convention can only constrain the range of possible meanings. Hirsch calls this the Boundary Principle (ibid., p. 54).

A related inconsistency in Hirsch's account is the status of authorial intentions. At one point he comes close to absolute intentionalism when claiming that a poem means what the author intended it to mean even if her intention fails in the sense that it does not correspond to one of the meanings permitted by linguistic convention (ibid., p. 12). This is probably what prompts Beardsley to call Hirsch's thesis the Identity Thesis: "what a literary work means is identical to what its author meant in composing it" (Beardsley, 1970, p. 17). But if this were true, the codetermining influence of convention would again be lost. Given that in his book Hirsch more or less bases his discussion on the co-influence thesis, I shall again ignore his self-defeating remarks.

It is now clear that the codetermining influence of convention secures room for a version of conventionalism in Hirsch's account. This *thin* conventionalism contrasts with the sort of *robust* conventionalism held by Beardsley, avoiding the Humpty-Dumpty-ism resulting from absolute intentionalism. The main disagreement between these two versions of conventionalism is whether convention alone can guarantee a single determinate meaning for any given word sequence. This is a difficult issue and it appears that who wins the debate on this point often depends upon whose examples are more convincing, and there are always examples in favour of both sides.

According to Hirsch, if meaning were not determinate it could not be understood and shared, because it would have no self-identity and could not be entertained by any human mind (Hirsch, 1967, p. 44). This seems to imply that an indeterminate word sequence is therefore *meaningless* (Meiland, 1978; Nathan, 1992; Iseminger, 1996). Combine this view with another claim Hirsch makes about meaning and we get a clear picture of how he thinks about language. He claims that meaning is an affair of consciousness, not of words; this entails that a word sequence means nothing in particular until someone means something by it or understands something from it (Hirsch, 1967, p. 4). Therefore, a word sequence is by nature meaningless in the sense that it does not refer to any determinate and hence understandable object, and meaning is actualized only when entertained by the human mind, either the critic's or the author's (notice that to be actualized this meaning has to fall within the confines of convention). Appeal to the critic's meaning is not promising because different critics will come up with different meanings for the same work, and it will be hard to adjudicate between them. This leaves the author as the sole option. Hirsch concludes that appeal to the author's meaning is the only compelling normative principle for the validity of interpretation. In other words, he takes the normative step that, between the actualized meaning of the author and that of the critic, we *should* choose the first instead of the second.

It has been argued that Hirsch's support for his recommendation is not strong enough, because he still acknowledges that the critic's will can make meaning determinate (Meiland 1978, p. 34). For those who embrace critical pluralism such as Goldman, they may claim that interpretative objectivity can still be preserved as long as we have a criterion by which illegitimate interpretations are ruled out. And Hirsch's thin conventionalism already offers this criterion: linguistic convention constrains the set of potential meanings to be actualized by critics. Therefore, a fundamental disagreement here is whether the only option for objectivity in interpretation is critical monism.

A possible way out for the extreme intentionalist here is to build the author's intention into work-identity, as suggested by Iseminger (1992, pp. 92-93): the author's intention, when compatible with the text, determines the identity of the work. This Revised Identity Thesis is less radical than the one adopted by the absolute intentionalist, for it works only for cases where meaning possibilities are actualized by authors. This thesis excludes cases in which different critics construct different meanings for the same text for the reason that they are simply interpreting different works. Nonetheless, it still has the consequence of identifying two textually indistinguishable works as the same work produced by two different authors with the same semantic intention, but one would think there might be crucial artistic differences between them. (Levinson, 1992, p. 239). It is unclear whether one should accept the ontological picture recommended here, since one can alternatively opt for the utterance model that bases work-identity on the broader context of the work's production. Iseminger simply begs the question on this point since he does not justify his account of work-identity.

Moreover, Iseminger explicitly admits that his thesis has difficulty in explaining cases involving irony and other figurative speeches, as meanings of this sort are typically out of the confines set by linguistic convention (Iseminger, 1992, pp. 92-93). This predicament holds for Hirsch too and leads to a dilemma (Nathan, 1992, pp. 187-188). Take irony for example. The first horn of the dilemma is this. Constrained by linguistic convention, the range of possible meanings has to include the negation of the literal meaning in order for the intended irony to be effective. But this ends up with naïve or absolute intentionalism again: every expression would be ironic as long as the author intends it to be. But—this is the second horn—if the range of possible meanings does not include the negation of literal meaning, the expression simply becomes meaningless in that there is no appropriate meaning possibility for the author to actualize. It seems that Hirsch needs a broader notion of linguistic convention to explain figurative language. But if he makes that move his intentionalist position will be undermined, for authorial intention would need to take a back seat after convention in such case. However, this is normally not a problem for modest intentionalists because for them the possible range of work-meaning is determined by convention plus context, and ironic meaning and the like could well be contextually permissible. More about context will be said when we get to modest intentionalism.

Hirsch is well aware that the practice of literary interpretation is very varied and that his account would fare better if it can be more tolerant of other modes of meaning. He then takes two crucial steps. First, he legitimizes the interpreter's interest in the work's *significance*.

Second, he acknowledges *unconscious* meanings. The second move gives rise to greater controversy than the first.

As maintained by Hirsch, significance refers to the relationship between meaning and things that lie beyond the work such as conceptions, attitudes, opinions, values, or feelings (Hirsch, 1967, p. 8). It is how the interpreter relates the work's meaning to those things. Briefly, significance is meaning-as-related-to-something-else (Hirsch, 1976, pp. 79-80). For example, the interpreter may consider how a 19th century work appears to us in our time. What the interpreter considers here is not meaning but significance; she is relating meaning to the situations and values in the 21st century. This is what a lot of critics do. Although they think that they are talking about meaning, in Hirsch's eyes they are not. However, significance is the legitimate object of *criticism*, but not of interpretation, which is exclusively concerned with meaning (Hirsch, 1967, pp. 56-57).

Of course this distinction can be challenged. However, since most philosophers engaging in the present debate seem to be satisfied with this distinction, I will accept it for the moment and focus instead on Hirsch's account of unconscious meaning which I found much more problematic.

A text is bound to have many potentially public meanings the author is not conscious of. These fall within the range of meanings constrained by convention, as the Boundary Principle shows. Aside from the one intended by the author, the rest are all potential meanings to be actualized by the critic. To broaden the notion of authorial meaning and thus accommodate more interpretations that seem legitimate, Hirsch develops the Continuity Principle as the criterion for meanings unconsciously entertained by the author. He calls these unconscious meanings *implications*. But how does the Continuity Principle work? Hirsch draws an analogy between meanings and icebergs. The exposed part of an iceberg is like the meaning consciously entertained by the author; the submerged part is the set of unconscious meanings. Just as the submerged part of an iceberg is continuous with the exposed part, so too are the unconscious meanings to the conscious one (ibid., pp. 53-54). In other words, among the potential meanings, those that are coherent with the one consciously willed by the author count as unconscious meanings and hence are also intended by the author.

Beardsley thinks that this is impossible. One can never mean something one is not conscious of (Beardsley, 1970, pp. 20-21). He maintains that what is unwitting is unwilled, that Hirsch cannot legitimately distinguish unconscious meanings from unintended meanings. Unfortunately, Beardsley rejects Hirsch's claim without saying anything about his Continuity Principle. Below I try to say more.

I do not think that the Continuity Principle is tenable for a very simple reason: the present principle falsely implies that, for a text, every convention-permitted meanings coherent with authorial meaning must be unconsciously meant by the author. Why can coherence justify this? Why can't one unconsciously will something that is not coherent with her conscious meaning and yet permitted by convention? Perhaps when Henry James wrote *The Turn of the Screw* he unconsciously intended the work to be a psychological drama, the meaning of which is apparently not coherent with the ghost-story reading he explicitly proclaimed that he intended. But then Hirsch would say there is no way the psychological reading could have been unconsciously intended by James, which seems implausible.

One can venture to come up with other relations for unconscious meanings, but I suspect that the difficulty remains the same: the definitional relation will always be arbitrary, and Beardsley just seems right to say that a distinction between unconscious meaning and unintended meaning cannot be drawn.

4.2.2 Objections and conclusion

Three much-discussed objections to Hirsch's account deserve a brief look before I close this section. I argue that all of them fail.

Against the Identity Thesis, an objection goes that there are textual meanings without authorial meanings (ibid., pp. 18-19). For example, an intelligible poem generated by a computer apparently can be meaningful. But this is not a counterexample to Hirsch's view, for he does acknowledge that meaning can be actualized by the interpreter. What he really claims is that we *should* favour the authorial meaning over the interpreter's meaning.

Another objection states that the meaning of a text can change after its author has died (ibid., pp. 19-20). This is close to what Hirsch calls "radical historicism" which claims that textual meaning changes from era to era (Hirsch, 1967, p. 6). To this, Hirsch might object that the sort of ontology favoured by the historicist is not appealing, not only because it fails to keep ontology simple but also because it has difficulty in explaining the work's self-identity over time. Alternatively, one can argue that Beardsley might have abandoned this meaning-changing argument because it does not fit other parts of his theory.[47]

A final objection has it that Hirsch's indeterminacy thesis results in an infinite regress: authorial proclamations of intentions will be indeterminate too. In that case we can never be

---

[47] I argue this view in my accepted paper (Lin, forthcoming).

sure of what the author really meant (Dickie & Wilson, 1995, p. 238). Nevertheless, Hirsch makes it clear that although it is true that author intention cannot be known with certainty, we can still aim for probable understanding (Hirsch, 1967, p. 17). This makes it clear that in his account of interpretation Hirsch is not making an epistemic claim about authorial intention but a metaphysical one.

In conclusion, extreme intentionalism has the appeal of neatly fixing the valid interpretation. Aside from the complications discussed in the previous section, an unsatisfactory consequence of extreme intentionalism is that it may strike many as unconvincing to say that an utterance ends up being meaningless when speaker's intention fails to actualize one of the convention-permitted meanings of a text. Some philosophers argue that convention and context fix meaning when that happens. And this is where moderate actual intentionalism comes into play.

4.3 Modest actual intentionalism

Modest actual intentionalism comes in different versions. Below I will discuss the most representative versions: *modest actual mentalism* (Noël Carroll), the *unified view* (Robert Stecker), and *partial intentionalism* (Paisley Livingston). The order of discussion is arranged by the complexity of the theory. I start from the most elegant of them.

4.3.1 Modest actual mentalism

Noël Carroll is one of the leading modest actual intentionalists (the word "actual" will hereafter be omitted in similar contexts). He only recently changed "intentionalism" to "mentalism." For convenience, I will use the former label until I explain what mentalism means at the end of my discussion of his view.

4.3.1.1 Convention-constrained intentionalism

Carroll claims that the author's intentions to be taken seriously are only those compatible with the text based on linguistic and literary conventions (Carroll, 2000, pp. 76-77). The author's intention fails if it does not square with the text, and hence is out of our consideration. In that case, meaning is fixed by convention alone. In other words, this move

avoids not only Humpty-Dumpty-ism but also the unfavourable result of meaningless texts, as in the case of extreme intentionalism.

The standard for realizing an intention is low in Carroll's account: as long as the text can be read as the author intended, the intention in question counts as successful, even if it is not the reading most likely to secure the audience's uptake (Carroll, 2011, p. 119). That is, if a literary text supports interpretation *A*, *B*, and *C*, and the author intended only *B*, then *B* is the correct interpretation of the work, even if *A* and *C* are more easily inferred from the text. Thus, Carroll's account preserves the merit of authorial intention as a disambiguating tool without granting it full authority.

This modest version of intentionalism is not without problems. The first problem is whether convention alone can lead to a single determinate meaning when the intention fails. Presumably, the modest intentionalist has to take textual ambiguity as pervasive; otherwise, the disambiguating function of intention would simply be lost. But this means that convention alone seldom fixes meaning when the intention fails. The predicament here is obvious: Carroll has a difficulty in reconciling textual ambiguity with convention robustly construed (Irvin, 2006, pp. 118-120). To this Carroll gives the unsatisfactory reply that textual ambiguity is not as pervasive as expected (2011, p. 131), which downplays the role authorial intent is supposed to play in an intentionalist theory.

The second problem is about context. According to Carroll, both authorial intention and convention are constitutive of meaning, but context does not have this meaning-shaping power. In other words, utterer's meaning plus word-sequence meaning are more than enough to get the work done. Carroll's example is this. Suppose that someone utters the sentence "The fish are on the bank" and that there is a pile of fish behind her on the bank of the river. Now, this mini-text supports at least two interpretations based solely on linguistic convention: the fish are right behind her or on the steps of a financial institution. If we take context into consideration, the first interpretation makes more sense. However, on Carroll's account, if the speaker actually intends "bank" to mean a financial institution, the second interpretation will be the correct one, because in this case the speaker's intention seems to be a more effective and legitimate meaning-determiner than the physical environment (2011, pp. 130-131).

Carroll here is rejecting the view that meaning is understood as utterance meaning on a contextualist basis, which claims that what a text means is what it ends up conveying in a particular context. This is because Carroll is basically following a Gricean framework of meaning, which focuses primarily on speaker's meaning. This runs against his earlier view (Carroll, 2002, pp. 326-327). I will assume that Carroll renounces his earlier claim. If this is

the case, the current account—Stecker calls it *convention-constrained intentionalism*—will face two difficulties (Stecker, 2010, pp. 149-150).

The first difficulty is that it is too narrow. This is because in many cases we realize our intention using words that go beyond convention. Irony is a case in point. When I say "You're such a genius" and intend it to be ironic you know my intention is successful because you know you are doing something stupid when I say it. Context apparently plays a role here; without it, the irony intended cannot be discerned in this mini-text. If we follow Carroll's account we will be forced to conclude that most, if not all, irony-related intentions are bound to fail, for no speaker's intention will square with the text they produce without consideration of context. Of course, irony is just one example; there are many other cases in which one intends to convey things that go beyond the conventional meaning of the text produced, and the intention in these cases is suitably seen as successful as long as it is permitted by context.

The second difficulty is that the account is too broad, because it ignores meaning-fixing context. It could be that an intention is compatible with convention but is hardly permitted by the context of utterance. In the case of the bank example, some philosophers like Stecker would think that the sentence "The fish are on the bank" exclusively means that the fish are on the riverbank even if the speaker intends "bank" to mean the financial institution, for the former interpretation is supported by context and is what the utterance ends up conveying . In this case context supplies a referent independently of the speaker's intention. It remains to be seen whose account makes more sense, but if the parallel between literary works and utterances makes sense then the odds would be with Stecker.

4.3.1.2 Conversation argument

It seems that no theory of interpretation can succeed on descriptive grounds, because critical practice is very varied. Offering a theory of meaning is of course a way to secure one's stance in the debate, but not all theorists (especially literary critics) are interested in the philosophy of language. The best way would be to defend a theory of interpretation on normative grounds, which concerns the aims of interpretation. Carroll has notably offered an argument to support his normative claim that the interpreter *should* aim to recover the author's intention when interpreting a work. The main idea of his argument is that we have a moral responsibility to understand others correctly and to find truth (Carroll, 1992; see also 2011, p. 133).

Carroll contrasts his position with the maximizer's, which maintains that interpretation should aim to maximize the work's value, subject to linguistic plausibility, for obtaining the fullest aesthetic satisfaction is the primary goal of literary interpretation. The maximizer's aesthetic argument does not preclude authorial interpretation but recommends that we consider it only when it maximizes the aesthetic satisfaction derivable from the work. Literary interpretation should always aim to maximize aesthetic satisfaction and seek the most aesthetically rewarding meanings the work can sustain.

To the aesthetic argument, Carroll objects that our overriding concern in art interpretation is *conversational interest*, not aesthetic satisfaction. The point here is that in ordinary conversation our interest in what is said derives from an interest in what the person meant: we are all actual intentionalists when it comes to ordinary conversation. If the interest in art is analogous, we should be actual intentionalists there, too. We would then want to converse with the author or artist by understanding what she is trying to say in her work. Maintaining this "human encounter" is a major impetus motivating our interest in engaging artworks. The aesthetic argument reduces this encounter to an I/It relationship instead of I/Thou, to borrow Martin Buber's terminologies. Ignoring the I/Thou relationship undermines the mutual respect we have for the author and lowers our self-esteem, because we would be insincere and self-deceptive with respect to the conversation in which we are engaging. If we treat a bad work as a masterpiece just because it affords a more aesthetically rewarding interpretation than the author actually intended, we are undermining the role we play in the artistic conversation with the creating individual.

As discussed in Chapter 3, the aesthetic argument appeals to something like a species of aesthetic hedonism. It assumes that aesthetic pleasure is the primary legitimate interest in art interpretation. [48] Such being the case, Nozick's experience machine argument can be adequately used to reject it, since his argument was originally designed to reject traditional hedonism, which claims that pleasure is the sole value we should pursue in life (Nozick, 1974, pp. 42-45). Nozick's argument shows that maintaining contact with reality is much more important than seeking pleasure; therefore hedonism is wrong. Applied to the aesthetic case, it shows that we have an investment in having a genuine conversation with the author, not just in counterfeiting such conversation to obtain aesthetic pleasure. It turns out that aesthetic

---

[48] One might disagree with Carroll's characterization of aesthetic pleasure here, for it denies/underestimates the cognitive involvement required to generate aesthetic pleasure. It is not a nice feeling; it is the pleasure that goes with understanding. See Levinson 1996.

pleasure is not all that matters to us in art interpretation, and that the conversational interest trumps the aesthetic concern. Even in cases where authorial meanings are aesthetically inferior to counterfeits, we resist the latter, just as we resist the idea of having pleasant experiences by using the experience machine. For we value being in contact with reality, both in the case of life experience and art interpretation.

Carroll's argument can be summarized as follows: (P1) The standard goal of an interlocutor in a conversation is to grasp what the speaker intends to say; (P2) Art interpretation is one form of conversation; (C) Therefore, the primary aim of art interpretation is to discover what the artist intends to convey.

In summation, given the conversational interest as a major concern, authorial intent should always figure in art interpretation, and any clues that help us to understand such intent are welcome (for example, biographical information). Since our primary aim of interpretation is to have a human encounter with the author, other purposes in interpretation should at least be constrained by it.

4.3.1.3 Objections to the conversation argument

Carroll's conversation argument has been widely criticized. In this section I examined six notable objections in detail, from the weakest to the strongest. The first two of the objections can be quickly dismissed, while the next two are harder to meet. The last two objections are the most formidable and satisfactory replies on Carroll's part become harder to find.

The first objection is that conversations are typically carried on in a natural language and are accompanied by facial expressions and gestures, while this is not the case in the "conversation" between artist and public (Wilson, 1997, p. 311; Dickie, 2006, p. 80). Nevertheless, with the rapid growth of communication software and people's increasing use of it, this description is no longer true. Nowadays many people spend more time facebooking friends than talking to them face to face, and yet they are no doubt engaging in conversation.

The second objection goes that conversations can be taken up again, but artworks, once finished, stand on their own (Dickie, 2006, pp. 80-81). Carroll could respond that of course people do say things like "Let's continue our conversation," but the continued conversation may well be a different utterance, provided that it is continued on a different occasion. The issue here is that Carroll would need an account of how to individuate conversations, but I assume that is not a big problem. A conversation that is continued after thirty seconds' break may count as the same conversation as the one before the stop, since the context is more or

less the same; a conversation taken up after a long break may count as a different one, since it is more distantly time-indexed. No doubt there will be borderline cases but such an account is not practically unworkable.

The third objection challenges the first premise of the conversation argument. It has been argued that grasping speaker's meaning is only the *occasional* rather than standard goal in a conversation (Dickie & Wilson, 1995, pp. 245-246; Dickie, 2006, pp. 71-73). Instead, most of the time the standard goal of a hearer is to understand utterance meaning because we generally assume that an utterance means what the utterer intends it to mean. In the normal course of a conversation, we do not think "what does she (or he) really want to say to me" when hearing every sentence from our interlocutor; we do so only when the sentence puzzles us, but this cannot be said to be a common phenomenon. In other words, the goal of understanding utterer's meaning is invoked only when the utterance is puzzling due to utterance failures or other special reasons.

Carroll's initial response is that Dickie and Wilson do acknowledge that we have the conversational goal of ascertaining speaker's intention when the utterance is puzzling. And it is commonly known that works of art are often puzzling. In that case, basing a theory of art interpretation on conversational puzzlement is not beside the point (Carroll, 1997, p. 307). Of course, this response does not meet the objection head on; it shows only that the objection overlooks the nature of artworks and that the application of the conversation argument in art still works. Carroll goes on to argue that recognizing speaker meaning is certainly our standard goal in a conversation and that it happens all the time (ibid., p. 308).

Carroll does not downplay the importance of utterance meaning, for it is the best guide to speaker meaning: in conversation we make presumptions and conjectures based on the utterance of our interlocutor to ascertain her intention. For example, when asked by someone on a street "Where is the Empire State Building?" one would think about whether the person asking this question is, say, a traveller or a pollster, before giving the answer. This means we tend to construct a framework in which to place our interlocutor in response to her utterance. More often than not, the assumptions we make about our interlocutor may be default assumptions not brought before the court of consciousness, but they are still assumptions. The relevant assumptions and conjectures constitute part of what is involved in the process of figuring out the speaker's intention. They are a necessary ingredient in conversation. Applied to the case of art, how we place the artist and her purpose is relevant to artistic interpretation.

Dickie replies that, still, in many situations we do not conjecture about speaker's meaning. In most cases the speaker makes what she says come across so that we do not inquire into her

intention (Dickie, 2006, pp. 71-73). However, Carroll could reply that the background assumption made when we attend to utterance meaning is that what is said is what is meant (Huddleston, 2012, p. 245).

Probably realizing the previous objection is not going to get anywhere, Dickie proposes another objection stating that Carroll holds an inconsistent conception of intention (2006, pp. 74-77). When Carroll initially developed the conversation argument he held a neo-Wittgensteinian view of intention, stating that authorial intention as the purposive structure of a work can be discerned by close inspection (1992, p. 101), but what he says later about utterance meaning as our best guide to speaker meaning ends up treating the speaker's intention as separate from the work, which runs counter to the earlier neo-Wittgensteinian view of intention.

Abandoning the neo-Wittgensteinian view of intention brings back the two-objects criticism that authorial intention is never part of the work. In that case, the attempt to build intention into utterance meaning automatically fails. However, if the conversationalist chooses to secure the neo-Wittgensteinian view of intention, that speaker's intention is manifest in utterance meaning, then the goal of ascertaining the speaker's intention would be superfluous given that attention to utterance meaning would be enough.

I do not think Carroll is being inconsistent here. Rather, I think Dickie has misinterpreted Carroll's latter view as holding a different conception of intention. As already shown, Carroll claims that the successfully realized intention determines work-meaning, and when he talks about the neo-Wittgensteinian view, what he has in mind is, presumably, the intention of this kind. In other words, the neo-Wittgensteinian view of intention finds its application only in the case of successful intention. This does not run against the view that utterance meaning is our best guide to the speaker's intention, because the latter view does not entail that the speaker's intention cannot be a component of utterance meaning. For Carroll, the distinction between speaker meaning and utterance meaning is not absolute. When successfully realized, the speaker's intention constitutes utterance meaning, and this is when the reader can discern authorial intention in the work. However, ascertaining a successful intention by close inspection of the utterance does not mean that the utterance cannot potentially sustain different readings. As Carroll sees it, an utterance could sustain multiple readings constrained by convention, and a successful intention will determine the correct utterance meaning. Dickie makes a second misinterpretation when he thinks that, in the case of successful intention, an utterance presents no ambiguity. He thus draws the wrong conclusion that in this case appeal to speaker meaning is gratuitous. Certainly, it is a problem for people like Carroll

to explain what constitutes a successfully realized intention, but at this stage the reply can work without such an account. (See§ 4.3.3 for more about the success condition.)

A more challenging objection has it that the reader can have equally legitimate interpretative goals other than the conversational one (Wilson, 1997, p. 311; Percival, 2002, pp. 202-203; Stecker, 2003; Huddleston, 2012, pp. 254-255). It seems that there are other legitimate and valuable goals of interpretation an interpreter can pursue, and it would be repressive to dictate one single interpretative policy. For example, when the interpreter finds a more interesting reading than the one the author has in mind, why not let it stand? Saying that this is deeply wrong seems unreasonable.

This objection might not hold if we acknowledge that interpreting a conversation is absolutely the same as interpreting an artwork. But this does not seem to be so. Even if we grant that art interpretation is some sort of conversation, it is still a very different sort of conversation that allows other interpretative aims, as some of Carroll's opponents would claim. Alternatively speaking, it is certainly true that we can have a conversational interest in interpreting art, but this need not trump other interpretative interests that are equally legitimate.

Carroll replies that he claims only that the conversational interest should *constrain* other interests (Carroll, 2002, pp. 321-322). Indeed, one of Carroll's claims is that actual intentionalism is compatible with the hermeneutics of suspicion, which is a skeptical attitude—often heavily politicized—adopted toward the explicit stance of a work. The symptomatic political contents may include, but are not limited to, sexism, ablism, classism, racism, imperialism, etc. To give Carroll's example, the suspicious hermeneut may find sexism in the works by George Bernard Shaw, even if he was a self-proclaimed proponent of women's rights. The suspicious hermeneut wants to attribute meanings the author might not have intended, including meanings contradicting apparent authorial intentions. It is believed that there is value in making these political interpretations not constrained by authorial intention, and that actual intentionalism stands in the way of this kind of criticism. Nevertheless, Carroll thinks that hermeneutical suspicion presupposes actual intentionalism, for the questionable content has to be non-ironic in order for the attribution of the skeptical interpretation to be proper. For example, in attributing racist tendencies to *Mysterious Island*, in which the black slave Neb is portrayed as docile and superstitious, we need to suppose that the tendencies are not ironic; otherwise the suspicious reading becomes inappropriate. To

give another example, in attributing sexism to Ian Fleming, we must assume that he intends Bond to be a male paragon (Carroll, 1993, p. 251).

Nonetheless, the suspicious hermeneut could reply that what she actually means is that the attributed meaning is the actual meaning of the work and that the challenged one is only apparent. If so, the latter is just a mistake we make when *decoding* the real meaning of the work, and thus we do not have issue with intention in this case. If we think that *Mysterious Island* is actually a racist work, then the non-racist reading is simply mistaken. Carroll's distinction between intended and unintended meanings implicitly commits the hermeneut to intentionalism. A potential way out for Carroll is to stick to the claim that a linguistic expression is bound to have both intended and unintended meanings. Such being the case, an interest in one necessarily presupposes the existence of the other, and work-meaning can be either or even a mix of both. Once this is granted there is room for talk of intentions.

The most challenging objection to the conversation argument is the *monologue* objection, which states that the greatest disanalogy between conversation and art is that in the latter case our response to artworks is utterly one-dimensional (Wilson, 1997, p. 311; Nathan, 2006, pp. 289-290; Dickie, 2006, p. 80; Livingston, 2005a, p. 151). That is, the reader's or the critic's remarks about a work are not a reply to the artist; interaction is not a necessary component in art interpretation, as it is in normal conversation. This objection has been further elaborated by Huddleston (2012, pp. 249-251). He claims that there are two conversational conditions Carroll fails to meet: the *mutuality condition* and the *openness condition*. The former refers to the lack of interchange in the literary "conversation," while the latter refers to the lack of minimal openness to the views of the conversational participant (the reader).

According to Huddleston, Carroll could save his account by construing the artistic conversation as a meta-level discussion, in which the reader need not stop with what the author intends to say. In this meta-level discussion, the reader first tries to understand what the author intends to say, but then she can contribute to the discussion by proposing *independent* and *interesting* interpretations she thinks are better than the author's. The mutuality condition is met because the conversation is not one-sided; the openness condition is met because the reader's opinion is valued. This meta-level discussion seems to be a model of a good conversation, for both sides of the conversation have something to contribute and show respect for each other.

However, as Huddleston indicates, if the actual intentionalist adopts the suggestion of meta-level discussion to change the monologue into a conversation, she will need to abandon the view of authorial intent as a constraint, because the meta-level discussion allows her to go

beyond the author's thought. In that case, the conversation metaphor militates against what the actual intentionalist wants to protect. The conclusion then is that the conversation argument is bound to fail, either because the artistic encounter is not a real conversation or because the argument will in the end betray the tenet the actual intentionalist strives to secure.

One way to meet the monologue objection is to specify more clearly the role of the conversational interest (Jannotta, 2014, pp. 373-376). As said, what Carroll actually claims is that the conversational interest should constrain other interests such as the aesthetic interest. In other words, other interests should be *reconciled* or *work with* the conversational interest. This point is epitomized by Carroll's acknowledgment of interpretations based on the hermeneutics of suspicion. In these, interpretations are constrained by the author's non-ironic intention, and hence they are legitimate on Carroll's account. The conversation we have here is mutual, since there is minimal interaction at play. For example, in the case of *Mysterious Island*, "Verne speaks; Carroll listens, understands, and replies. Verne, admittedly, can offer no rejoinder, but we have nevertheless one full exchange. Thus mutuality is satisfied." (Ibid., p. 374)

The conversation in the above example is also open, since the reader's view is valued. It would be a fault on Huddleston's side to claim that the reader's independent interpretation has to be both non-intended and irrelevant to authorial intention, for the irrelevance condition does not make sense in a conversation, which tends to be mutually purposive.

Even if Carroll appeals to the proposal of meta-level conversation, he can still retain the conversation metaphor without selling out actual intentionalism (ibid., pp. 376-380). This is because there are two problems with Huddleston's characterization of meta-level conversation. First, as said, he falsely holds that the reader's view has to be independent in the sense that it is neither intended by the author nor relevant to her intention. This kind of independence is at odds with the mutual purposiveness of a conversation. Second, the interestingness Huddleston has in mind sounds like aesthetic maximization (best). But the conversationalist need not construe "best" as aesthetically best. It can be both epistemically and aesthetically optimal. Given these two points, the meta-level conversation does not deviate from the intentionalist picture.

4.3.1.4 Some general concerns with modest actual mentalism

So far the objections are directed at the conversational aspect of Carroll's position. But there are controversies over the position in general. In this section I discuss two major ones.

One frequently expressed worry is about how intention figures in Carroll's story. Some have raised the objection that intentional actions can happen even when there is no conscious act of intending (Freeland, 2014, pp. 21-24; Auxier, 2014). That is to say, the artist might not know what she means until the work is finished or it is experienced by the audience.

Carroll's reply is that actual intentionalists (not only himself) never claim that the relevant intention has to be a fore-intention (Carroll, 2014, p. 43, 45). What the objector says is completely compatible with actual intentionalism. What is more important, Carroll's position is *mentalism* instead of intentionalism. According to Carroll, this creates more space for tacit and unconscious assumptions to come into play. Carroll writes:

Modest actual mentalism contends that the meaning of a poem is determined by the actual intentions and underlying, though not necessarily concious, assumptions of the poet. That is, modest actual mentalism holds that the cognitive or, more broadly, mental stock of the artist fixes the meaning of the work, so long as said intentions, assumptions, etc. are consistent with what is available in the text." (Carroll, 2013, p. 14; see also Carroll, 2011, pp. 120-121).

This shift from intentionalism to mentalism has the benefit of avoiding the narrowness of "intention" and of accommodating the broadness of "meaning" in the current debate. Apparently, philosophers of art do not use the term "meaning" in the same way philosophers of language use it. The former typically concern themselves with a broader sense of semantic meaning. Specifically, the term tends to be loosely construed as constituted by various mental features a critic might refer to such as intentions, attitudes, feelings, values, beliefs, thoughts, etc. Interpretation, typically understood as the search for meaning, is accordingly construed as targeting "meaning" broadly defined. Obviously the term "intention" is too narrow to capture all of these features. Carroll's move is suitable in the context of artistic expressions, since the said features are typically what would interest a critic.

Discussion of intention introduces a big challenge to Carroll in the case of *authorial endorsements* (Davies, 2014, pp. 15-16). The most important claim in Carroll's account is that the author's successful intention determines the meaning of a work. Then, what would Carroll say about interpretations not only compatible with the text, but also endorsed by and yet not intended by the author? Do authorial endorsements weigh at all in this kind of case if the author's mental stock, conscious and unconscious, determines what a work means?

This is not a problem in the case where authorial intention is ambiguous and the proposed interpretation can be covered, but would be a problem if one claims that the intention in question is unconscious. Appeal to unconscious intention makes Carroll's account circular and unfalsifiable. It makes sense only when we have strong evidence that the relevant intention lurks in the author's subconsciousness; in that case, the author is probably trying to suppress her awareness of it. It is also not satisfactory to meet the objection by saying that the author is mistaken in agreeing with the proposed interpretation, because this amounts to arbitrarily disregarding the author's words about intention.

To this, Carroll replies that authorial endorsements must be assessed gingerly against textual and contextual evidence and are not to be taken at face value; we can thereby exclude cases of the author lying, misspeaking, being in delirium, or simply ingratiating herself with the audience (2014, pp. 40-41). After all, if we accept authorial endorsements without critical eyes the result will be no different from Humpty-Dumpty-ism. However, one would think that quite a lot of the time there is no compeling evidence for us to make the judgement. If so, this reply might seem to be a question-begging dodge.

As for the cases in which the author sincerely makes the endorsement, these might be where the critic articulates an idea that was vague in the author's mind when she created the work, so says Carroll. This solution is not particularly appealing. The author might endorse an interpretation simply because it praises her work, not because she vaguely entertained it in the process of creation. Then in practice how do we distinguish one case from another? Carroll's reply is in danger of conflating the two.

4.3.1.5 Section conclusion

From the aforesaid discussion I summarize three of the most serious problems for modest actual mentalism. First, the modest actual mentalist needs to say more about context. Does Carroll reject a full-fledged contextualism? If this point remains unexplained then it will remain unclear what kind of evidence the interpreter can legitimately use. If context does not matter, the interpreter would just focus on evidence of authorial intent, ignoring the relevant historical and social conditions of the time in which the work was created. This result does not square with what Carroll says when he compares his account with hypothetical intentionalism: "[…] absent the so-called private avowals, the two views will arrive at comparable hypotheses." (Carroll, 2014, p. 38)

The second problem concerns the conversation argument. Assimilating art interpretation to conversation is not promising. Carroll claims that art interpretation is on a continuum with ordinary conversation. However, the empirical facts do not seems to lend succor to this claim: many people do not interpret works of art in the way they interpret daily utterances. If a significant number of people do not treat art interpretation as conversation, then the claim that art interpretation is continuous with conversation just does not hold. There must be crucial differences that prompt people to treat the former otherwise. It is probably more plausible to say that art interpretation is a different game from daily conversation (Levinson, 1992, pp. 240-241; see also Nathan, 2006, pp. 291-292).

The last problem is the controversy over unconscious mental states. Pinning down the acceptable interpretation in this case becomes a troublesome task. To know whether or not a reading is intended unconsciously by the author, we need to look closely into the relevant evidence. Nevertheless, we can imagine that in many cases the evidence underdetermines such an hypothesis, for the critic can always come up with seemingly plausible interpretations that await testing. Modest actual mentalism ends up producing too many interpretations that are only "provisional." Carroll could have introduced the useful concept of mentalism without trading in unconscious mental states.

The unified view, a second version of modest actual intentionalism developed by Robert Stecker, is free of all the above concerns. I now turn to the discussion of his view.

4.3.2 The Unified View

Stecker's version of modest actual intentionalism is dubbed the *unified view* (alternatively called the *disjunctive view*). He develops this view and discusses associated issues at several places (Stecker, 1993, 1997, 2003, 2006, 2008, 2010, 2013). In this section I will present his up-to-date account.

The way Stecker approaches the subject matter at hand is systematic and comprehensive. As Stecker sees it, there are three tangled issues in the debate on literary interpretation: the work-meaning issue, the proper-aim issue, and the critical monism-pluralism issue. Strictly speaking, the unified view is the solution only to the work-meaning issue, which is arguably the nucleus of the debate.

Briefly put, the work-meaning issue is the debate about the nature of work-meaning; the proper-aim issue concerns the legitimate aim(s) of interpretation; the critical monism-

pluralism issue discusses whether there is a single right interpretation for a work. In what follows I will examine these issue in the above order.

4.3.2.1 Work-meaning issue (1): the utterance model

Most participants in the debate on interpretation agree that there is such a thing as work-meaning; as shown in the previous sections and chapters, they tend to assimilate it to utterance meaning.

The distinction between sentence and utterance is that the latter is a *use* of the former. An utterance is a sentence (or better put, word-sequence or text) used on a particular occasion. Such being the case, sentence meaning and utterance meaning will be determined in different ways. Sentence meaning is the meaning "attachable to a sequence of words taken in the abstract in virtue of the operative syntactic and semantic (including connotative) rules of the specific, time-indexed, language in which those words are taken to occur." (Levinson, 1992, p. 222). As already discussed in Chapter 2, the anti-intentionalist is generally interpreted as appealing to sentence meaning in determining utterance meaning. However, utterance meaning as it is used in the philosophy of literature has been more sophisticatedly defined by philosophers after Beardsley. Roughly speaking, utterance meaning is what one ends up conveying by the words uttered on a particular occasion, which requires us to take into account both the words produced and the context in which they are produced.[49]

But how does the author's intention figure in this picture? This is where the unified view comes into play and turns the above rough characterization into more detail by unifying intentionalist and anti-intentionalist requirements. It claims that when the author successfully realizes her intention, the utterance means what she intended it to mean; when the intention fails, meaning is fixed instead by convention and context in place. In other words, utterance meaning is determined by successful intention *or* convention plus context (hence the *disjunctive* view).[50]

This formulation has the merit of acknowledging the role played by context, the lack of which appears to be a downside in Carroll's account. But the addition of context to the theory

---

[49] Stecker claims that the utterance model can be extended across the arts as long as we substitute the meaning-conveyed medium for words. Nonliterary artworks do not literally have utterance meaning but they are indeed meaningful.

[50] Strictly speaking, modest actual intentionalism of whatever brand will need to be disjunctive; otherwise it cannot specify how meaning is to be fixed when the author's intention fails.

also means that, to define what counts as "success," Stecker would need a more complex account than Carroll's. Precisely, an utterer successfully intends a meaning, X, if (1) the utterer intends X; (2) the utterer intends that the audience will grasp this in virtue of the conventional meaning of the words uttered or a contextually supported extension of this meaning permitted by conventions; (3) the first intention is graspable in terms of those conventions or their permissible extensions. Call this formulation the *intermediate view* (Stecker, 2010, p. 151; see also his 2006, p. 429).

One thing to add before I discuss the intermediate view is that, for the third condition, "graspable" is used in a stronger sense than it is in Carroll's account: the content in this case has to be the one most likely to secure the audience's uptake, while Carroll requires only the possibility of uptake (Stecker, 2003, p. 14; see also Davies, 2007, p. 170).

Note that the above Gricean formulation, unlike Grice's original account, is not intended to be a complete account of meaning, for Stecker still uses the concept of conventional meaning and leaves it unanalysed. What the above formulation says is that if the three conditions are met, the utterance means X. However, it does not go vice versa: if the utterance means X, it does not follow that the three conditions have been met. Utterance meaning could result from unrealized (improperly realized) intention as well as unintended meanings. Where there is no properly realized intention to fix utterance meaning, yet the utterance appears to have meaning, then this is fixed by convention and context directly.

Stecker believes that the intermediate view avoids the two drawbacks of Carroll's convention-constrained intentionalism discussed in the previous section: not only does it allow an intention to be realized by contextual extensions, but it also excludes intentions that are compatible with the words uttered (text) and yet hardly permitted by context.

However, the formulation of the intermediate view seems flawed to me. It does not successfully remove the second drawback in Carroll's account. To demonstrate, let us use again the bank example given by Carroll. Suppose that, with a pile of fish on the riverbank behind me, I utter "The fish are on the bank," intending this to mean that the fish are on the steps of a certain financial institution. Moreover, I also intend that my interlocutor will grasp my intention in terms of the conventional meaning of the words uttered. Then the first and second conditions of the intermediate view are both met. Contrary to Stecker's expectation, the third condition, arguably, is also met, for my first intention is indeed graspable by virtue of the relevant linguistic conventions: one of the conventional meanings of the word "bank" refers to a financial institution. Since the third condition is a disjunction, for it to be true only one disjunct needs to be true. It follows that my utterance in this case means what I intended

it to mean, that is, the fish are on the steps of a financial institution, not on the riverbank, which is what Stecker thinks is the correct conclusion to draw.

This is just a technical error that does not pose a serious problem. A modification will do. The intermediate view can be improved by rephrasing the second half of the third condition, which states that the utterer's intention to mean X is graspable in terms of relevant conventions or their permissible extensions. The sentence after "in terms of" can be changed to "the conventional meaning of the words uttered *understood contextually*." Therefore, the third condition in its revised form is: "the first intention is graspable in terms of the conventional meaning of the words uttered *understood contextually*." This avoids the aforesaid mistake caused by the disjunctive formulation and preserves the function of context. For the remainder of this section I will assume that this modified formulation is the correct version of the intermediate view and use it in my discussion.

4.3.2.2 Work-meaning issue (2): rejection of the utterance model

The most fundamental and controversial issue here is whether a work does have such a thing as meaning (Olsen, 1982). Olsen thinks that talk about literary works having meanings makes a category mistake. Saying that a work has a meaning is like saying that someone's voice is green. It would be natural to talk about the meaning of a word or utterance, but it would be very odd, and not useful, to ask what the meaning of a work is. For example, what is the meaning of *Macbeth* or *Harry Potter*? This way of asking seems unnatural. Moreover, he does not think that literary works have the same meaning-producing features as sentences or utterances. The meaning of a sentence is a function of two factors: the meanings of the words and how they are combined. It is very unclear how a literary work has the analogous parts and mode to produce meaning.

Peter Lamarque (2002) uses Beardsley's interpretative model to demonstrate the point. Recall Beardsley's distinction between explication, elucidation and interpretation in the narrow sense. It seems that the utterance model (or other linguistic models) is most convincing at the level of explication, especially in the case of poetry. Nonetheless, explicative questions seem out of the question when it comes to literary works of greater length. For novels, our interpretative debates are typically not about verbal meaning.

As for elucidation, there are many cases where a determinate right answer is available, but there are also cases where indeterminacy arises. However, most of the missing details are

irrelevant to the appreciation of the work. For example, no one would care about how many hairs Macbeth has, and there are infinite questions of this kind for a story.

Interpretation, the level of characterizing themes, seems to be the right level with which the debaters should be concerned. The elicitation of themes does not rest on thematic summaries (which is Beardsley's position) but on joining elements of the work together under themes of general interest. Finding a theme is finding a unifying vision that reveals the value of the work. It turns out that literary interpretation calls for *appreciation* rather than understanding, for interpretation aims to reveal the work's interest or value, not meaning.

To address the worries expressed by Olsen and Lamarque, we need to rethink how we choose to construe "meaning." If the term is understood in a broad way, then talk of work-meaning does not seem to be beside the point. We have already seen how Carroll resolves this problem. Stecker takes a similar approach: when the interpreter tries to identify work-meaning, what she does is to identify what the author *does* in the work. The author's doings may include "saying, representing, expressing, presenting, alluding to, allegorizing, and so on." (Stecker, 2003, p. 55) Paisley Livingston has a similar claim: "I take it that this [meaning] includes the events, characterizations, and situations figuring in the story conveyed by a novel, play, film, or other narrative work, as well as the emotions, attitudes, and themes expressed by any work of art." (Livingston, 2005a, p. 138) Nathan puts the point neatly: "Meaning […] is the product of coming to understand a work, where understanding encompasses a full range of kinds of awareness of the artistically pertinent properties of the work in question." (Nathan, 2006, p. 282)

As early as 1978 when the utterance model was first introduced by Tolhurst, he makes the following note which seems to be largely overlooked by most objectors to the idea of work-meaning:

There may be some philosophers of language who would object to my use of the term 'utterance meaning'. They might argue that the concept of meaning should be restricted to word sequence meaning, and that what I have called utterance meaning should be described in some other way […] I believe that what I have called utterance meaning captures much of what critics are interested in when they try to elucidate the 'meaning' of a work of literature and that most of the problems in the interpretation of literature do not hang on questions of word and sentence meaning. The difficulty which I have had in understanding a poem […] has nothing to do with not understanding the meaning of the sentence which the poem instantiates and quite a lot to do with the point or force of the

sentence which is given in a specification of its meaning as an utterance. (Tolhurst, 1978, p.14)

There are two points to note here. First, if the above way of construing literary meaning makes sense, then it is not clear why a work does not have meaning. Needless to say, this macro-meaning is a complex of micro-meanings (as was discussed by Beardsley in Chapter 2). But this complexity is not a reason against viewing it as a whole. Second, if ascertaining the *point* or *force* is an important aspect of interpreting an utterance, one wonders why the same does not hold with literary works. We do ask questions about the point of designing a particular character, plot, or narrative structure; we also ask questions about the force of a work. For example, is the work meant to castigate capitalism, report the poor conditions in undeveloped countries, or recommend an alternative political regime? In this respect the differences between literary works and utterances seem to dissolve.

With this qualification, there may not be much of a distinction to draw between appreciation and understanding. Finding a unifying way to reveal the value of a work can well be a process of recovering meaning. Of course, the latter does not necessarily invoke value; conflating the two may not be an adequate move. But appreciation, as defined here, is indeed in danger of collapsing into the broader class of understanding. The crux of the issue then becomes whether the meaning in question is *recovered* or *imputed*. However, if the utterance model applies to literature, it is hard to see why a work, construed as an utterance, does not have a meaning itself and, instead, can only be assigned one.

4.3.2.3 Work-meaning issue (3): two objections

One of the most serious challenges Stecker has to face after meeting the above objection is the circularity problem (Levinson, 1996, p. 180). Defining meaning by the notion of successful intention is circular. One cannot define work-meaning as a successful intention without an independent notion of meaning, because without that one will not be able to explain what that success amounts to. To demonstrate: a work W has a meaning property, M, and M is the product of a successful intention because the author intended M and W has M. This argument is circular.

As said, the intermediate view already specifies an uptake condition that deflects this worry. The third condition of the intermediate view makes it clear that the work, with convention and context in place, is such that the artist's intention is capable of uptake from the audience

(Stecker, 2003, pp. 46-47; see also Trivedi, 2001, p. 197). So, if this uptake is secured then the intention counts as successful.

But what does the audience actually recognize in this case? Presumably, the audience recognizes that the author intended W to mean M, which still presupposes the notion of W meaning M (Levinson, 2010, pp. 145-146). The formulation of meaning remains circular.

Perhaps one can modify the uptake condition as: the audience would recognize that the author intended to mean M by W. This way, the notion of W meaning M is not obviously presupposed. But this move, though avoiding circularity, brings about another question: W's meaning M is then based on the audience's recognition that the author intended to mean M by W. This sounds wrong because it makes impossible the recognition that the author fails to mean M by W. The moral so far, according to Levinson, is that there is an unbridgeable gap between utterer's meaning and utterance meaning. The latter is just not analysable in virtue of the former plus audience uptake. In his 2010 reply to Levinson's criticisms of modest actual intentionalism, Stecker, for reasons unknown, addresses other criticisms except this one (Stecker & Davies, 2010).

Aside from the circularity problem, a second troublesome objection states that references to context and to convention are enough to secure meaning. If this is so then the author's intention becomes superfluous. Stecker replies that this may be true "when all that is needed to understand what is said is to take the words, *already disambiguated*, at the most literal level." (Stecker, 2003, pp. 48-49, emphasis original). Nonetheless, this is not true when we ascend to higher levels of interpretation.

Here Stecker seems to claim that at higher levels of interpretation, even with the help of context and convention, we still need the author's intention to disambiguate meaning. But this claim brings us back to the dilemma of ambiguity presented to Carroll's convention-constrained intentionalism: if context plus convention are not enough to remove ambiguity in a number of cases, work-meaning will unfavourably remain ambiguous in these cases when the author's intention is unrealized; however, if context plus convention can always disambiguate meaning, then the objector wins—we do not need the author's intention for disambiguation. This dilemma, again, justifies Levinson's concern that building utterer's intention into utterance meaning is not promising.

4.3.2.4 Proper-aim issue

The work-meaning issue is about the ontological question of what work-meaning consists of, while the proper-aim issue is about the normative question of how we *ought to* interpret a work. These two issues constitute the core of the debate on literary interpretation.

Stecker recognizes that there are a number of legitimate aims of interpretation: " […] if people engage in certain interpretive practices that make no straightforwardly false assumptions and that aim at valuable goals, then it is not clear how we can object to such practices." (Stecker, 2013, pp.312-313) All the theories of interpretation discussed in this thesis have such legitimate aims: the anti-intentionalist aims to retrieve work-meaning through linguistic convention; the value-maximizer aims to offer interpretation(s) that enhances aesthetic satisfaction; the actual intentionalist aims to discover what the author means; the hypothetical intentionalist aims to hypothesize what the author means (Chapter 5); the hypothetical authorist aims to explore what the author could mean (Chapter 6).[51]

These are all legitimate aims of interpretation and presumably there are more. One example is the search for significance, as discussed early on in this chapter: how the audience relates the work to something outside it. A second example is the attempt to explore how the work can be taken by different perspectives such as the Marxist, feminist, or psychoanalytic frameworks. There is no reason to push back the search for significance and so-called theory-driven interpretations since there is no point to denying that they are both guided by valuable and plausible goals. It follows that Stecker advocates a plurality of interpretative aims. Call this *proper-aim pluralism*. The fact that the work has a meaning does not compel us to limit our aim to discovering it, which would be the is/ought mistake Hume reminds us not to make.

At this point the following objection is anticipated: if one already has an account of what a work actually means then it is hard to see why recovering meaning should not be the sole aim of interpretation. Call this *proper-aim monism*. Even if one thinks that proper-aim pluralism makes more sense, a natural thought would be that recovering meaning should at least be the *central* aim of interpretation. Call this *hierarchical proper-aim pluralism*, in contrast to *indiscriminate proper-aim pluralism*, which is Stecker's position, stating that all legitimate aims of interpretation are of the same weight.

It is true that the is/ought consideration makes perfect sense in many different contexts, but in the present debate theorists tend to identify the aim of interpretation with the search for meaning, as can be seen from my discussions on most positions in this thesis. It would be odd

---

[51] Note that some of these aims can be encompassed by the broader aim of recovering work-meaning. This is normally true for interpretative positions packaged with a theory of meaning.

(or even insane) for, say, the hypothetical intentionalist to claim, "Although work-meaning is determined by our best hypothesis of the author's intention, interpretation should seek to find significance for the reader." If so, why spend so much time elaborating how meaning is determined if that is not going to have some connection to the normative dimension of interpretation? It seems to me that one crucial move in the debate on interpretation is to base the normative account on the ontological account. This is at least how Beardsley and Wimsatt started the debate, and this may be Carroll's point when he says all the participants in the debate have to answer the questions of what constitutes meaning and of what evidence is to be used for discovering meaning (Carroll, 2013, p. 10). The way Carroll frames the debate apparently assumes the normative purpose behind the ontological move. In short, Stecker's pluralism would make more sense if he at least made meaning-recovery central among other legitimate aims of interpretation.

However, an objection Stecker would face if he made that move is that hierarchical proper-aim pluralism renders a number of interpretations "second best," which implies that these interpretations bear less weight. I think here Stecker could make either of the following two moves. First, bite the bullet and accept that there are second best interpretations. I do not see harm in acknowledging this, for it does not reject the legitimacy of the interpretations pursued by secondary aims. Things of less importance are still worth doing and we do them frequently in our life. Discovering what the author intended to do in the work may be very important, but this does not mean that considering a work from the Marxist perspective is not worth doing at all. The interpretative freedom of the reader is thus preserved.

Alternatively, Stecker could opt for proper-aim monism. That is, he could just stick to the aim prescribed by the unified view and defend the former by defending the latter. This strikes me as a natural way to defend a theory of interpretation. This is also what Stecker and I disagree upon.

Finally, Stecker still seems to overlook the nature of the debate: it is precisely because people disagree about the aim of interpretation that they start arguing. It seems that people tend to have preferences for some aims over others. Indiscriminate proper-aim pluralism is like a qualified version of interpretative anarchism that does not shed much light on the perennial query: how should we interpret? It just gives up answering this question. As said, hierarchical proper-aim pluralism would be a more plausible version to adopt if one still finds proper-aim pluralism appealing.

Stecker might reply that what he tries to defend is the legitimacy of interpretative goals rather than interpretative norms, because there is no need to argue for the norms for

interpretation once we concede that there are a number of interpretative goals worth pursuing (hence legitimate goals). This reply is not satisfactory, as the question remains that people disagree with each other even over those "legitimate" aims. Once we find a strong preference for a particular goal of interpretation it is hard not to treat it as a norm for interpretation instead of simply viewing it as "legitimate."

4.3.2.5 Critical monism-pluralism issue

Critical monism is the thesis that there is a single right interpretation for each work of art; critical pluralism rejects monism by claiming instead that a work of art is able to bear multiple acceptable interpretations (Stecker, 1994, p. 193). Since Stecker holds proper-aim pluralism, it follows that he accepts critical pluralism.

But one can also defend critical pluralism by defending a particular aim of interpretation that seeks plural interpretations for a work. The value-maximizing theory is a case in point, so is the significance-seeking approach. Combining these multiple and equally acceptable interpretations into a single interpretation is not appealing, as doing so creates a hodgepodge rather than a coherent whole.

Interpretations reached by aims that prescribe critical pluralism are typically not truth evaluable, and that is why the word "acceptable" is used instead. However, Stecker claims that many of these interpretations make at least weak claims and, hence, can still be either true or false, but he does not press on this point (Stecker, 2003, p.55).

It should be less controversial that interpretations reached by the unified view are truth evaluable. These interpretations spell out what the author *does* in the work on a particular occasion, including intended and unintended doings. As required by the utterance model, these interpretations will all be historically accurate statements made true by successful intention or context plus convention.

At this point one might venture to suggest one sort of monism that conjoins or disjoins all the true interpretations (reached by whatever aim) into a mega-interpretation. But this suggestion is unappealing in the same way the suggestion is that we combine all acceptable interpretations obtained by a pluralist aim of interpretation. The result will be a pointless combination. Stecker hence rejects this proposal (ibid., p. 56).

Monism, however, makes sense for the interpretative aims that seek a single right interpretation for a work. Most actual intentionalists, the anti-intentionalist and the hypothetical intentionalist all embrace this aim. If one thinks that recovering meaning is *the*

aim of interpretation, then critical monism will be the proper label for this view. In the case of actual intentionalism, where the author intends to make a multi-layered, ambiguous work, it might be that many readings will be possible. But here the one correct interpretation would be the conjunctive one.

For Stecker, critical monism and pluralism are compatible, because proper-aim pluralism implies that both positions are equally legitimate. A notable objection is worth considering. One could argue that if an interpretation does not respect the identity-relevant contextual factors of a work, then it is not acceptable, because in that case the object of interpretation will not be the same work. To the extent that the unified view leads to an interpretation that respects the relevant contextual factors present at the time of the work's creation, interpretations pursued by other aims have to be ruled out in that they fail to respect the work's identity. It follows that in Stecker's account critical monism is incompatible with critical pluralism (ibid., pp. 64-69).

Nevertheless, for aims that seek to answer "What could the work mean?" or "In what way can the work be taken?" we need to be clear about what these questions are really asking. There is a pragmatic sense of "could" and "can be taken" that is relative to a point or set of constraints. In the case of interpretation, this is when we bracket off something we know in pursuing a particular aim of interpretation. Stecker's example is that "we may know a work has an essentially polemical aspect, but may want to bracket it off to see if we can find a more general significance in it." (Ibid., p. 66)

Performative interpretations make a case for this claim. A modern presentation of *Macbeth* is still a performance of *Macbeth*, even if it fails to respect many of the relevant contextual factors of the original work. The pragmatic "could" or "can be taken" is apparently at work in this kind of performative interpretation.

I have misgivings about Stecker's solution. How does he justify the claim that interpretations that do not respect work-identity are on a par with those that do? At one point when talking about the interpretation of *Richard III*, Stecker writes, "So the play does not represent a totalitarian everyman. Yet I might, and people have, interpreted the play without being ignorant of its contextually fixed meaning." (Ibid., p. 68) On the same page he goes on to claim that "It is usually allowed […] that they [performative interpretations] are productions of the work." Stecker is in danger of saying that the interpretations in question are of the same work because people take them to be so. To be charitable, we can ignore the ill support here and yet still acknowledge the pragmatic sense of "could." However, this pragmatic sense appears too liberal to me, as it could cover almost all interpretative aims

except the one prescribed by the unified view. After all, interpretations that do not respect work-identity are presumably those in which we bracket off some piece of contextual information on purpose.

But then a hierarchical proper-aim pluralist would have less difficulty in coping with this issue, for she could maintain that an interpretation that does not respect work-identity is not as worth pursuing as one that does, for it is not an acceptable interpretation *of the work*. But this does not mean that it is not an acceptable reading of the *text* or some other kind of object of interpretation.

4.3.2.6 Section conclusion

The unified view does not have the three drawbacks of modest actual mentalism discussed at the end of §4.3.1.5: (1) it recognizes what context can do in fixing meaning; (2) it does not need to answer formidable objections to the conversation argument; (3) it does not appeal to unconscious mental states. Do these advantages make the unified view a more convincing account than modest actual mentalism? Not necessarily. Below I summarize some general concerns.

As stated, the unified view is still haunted by the dilemma of ambiguity, even after context comes into play. Building utterer's meaning into utterance meaning poses a general difficulty for actual intentionalists, as further shown by their strenuous effort in coping with the circularity problem.

Second, holding indiscriminate proper-aim pluralism gives the game away. What would count as an implausible aim for advocates of this position? Stecker does not say. Then the position boils down to anarchism: any aim of interpretation is as worth pursuing as any other.

Third, the ambition to justify the compatibility of critical monism and critical pluralism is questionable in Stecker's framework. This problem is apparently rooted in his insistence on indiscriminate proper-aim pluralism. A more plausible kind of compatibility might be achieved by adopting hierarchical proper-aim pluralism that does not limit acceptable interpretations to one type of object of interpretation.

Livingston's partial intentionalism is free of the latter two worries. It remains to be seen whether it has a chance of meeting the first challenge. It is to his account that I now turn.

4.3.3 Partial intentionalism

Like the previous two versions of modest actual intentionalism, Paisley Livingston's partial intentionalism maintains the view's disjunctive nature: work-meaning is determined by successful intention or by convention and/or context when the intention fails (Livingston, 1996, 1998, 2005a, 2009, 2010).[52] Here there are two major points worth considering: (1) Does Livingston, like Carroll, conceive context as metaphysically idle in shaping work-meaning? (2) How does the partial intentionalist cope with the problem of specifying success conditions? Below I will start with (2). Concerns with (1) will be revealed in due course.

4.3.3.1 The meshing condition

The following passage shows how Livingston defines the success condition:

My own proposed solution to the problem under discussion runs as follows: the intention to mean *q* by saying or otherwise representing *p* is successful just in case the intention to imply *q meshes* sufficiently with what is written, spoken, or otherwise put on display. (Livingston, 2009, p. 99, emphasis original)

Livingston concedes that this meshing condition is vague, but can be roughly understood as follows:

The meshing condition applies to the relation between the content of the intention and the text's or structure's conventionally determined meanings as well as the explicit and implied ideational relations that give a work its coherence, such as rhetorical connections between its various parts. The meshing of intention and structure requires a high degree of coherence between the content of the intention and the display's rhetorical patterns. (Ibid., p. 100)

An example of meshing offered by Livingston is this. Suppose that a film tells a story in which a fisherman in a small village accidentally acquires a large fortune and becomes rich. The story then centres on the stark contrast between his life and the life of the other villagers: the former is contaminated by vices and misfortunes while the latter remains peaceful and simple. Suppose further that the director's intention in this case is to convey the thesis that

---

[52] In the first two papers Livingston opts for a Gricean account but abandons it in his 2005 book.

wealth corrupts. Then the intention does mesh sufficiently with the cinematic display because the intended thesis coheres highly with how the film's segments are arranged.

In order to further clarify when meshing obtains, Livingston offers two *symptoms* of it (Livingston, 2010, p. 414). The first symptom is that the audience is apt to infer the authorial intention from the text. This is very much like Stecker's uptake condition, but Livingston does not define success in terms of it. The second symptom is the audience's ability to develop a coherent interpretation based on textual evidence after being informed of the author's intention. These two symptoms thus serve as useful means for checking whether an intention meshes well with the text. Note that a symptom, like a clue, is neither a necessary nor a sufficient condition for successful intention; it is evidentiary but not definitional.

A final note before I assess the meshing condition is that meshing does not require that the intention concerned bear a strong conceptual association with every textual feature. Since it is reasonable to assume that a work will inevitably have many unintended features, requiring a global meshing would be too demanding.

4.3.3.2 Objections to the meshing condition

Livingston's reason for rejecting Stecker's uptake condition is that the interplay between the audience's recognition on the one hand, and the relevant conventions and contextual factors on the other, is difficult to specify. In other words, it is hard to say under what circumstances the audience's recognition is warranted (Livingston, 2010, p. 414). However, this may not be a serious problem, since Stecker is not obliged to articulate the relevant mechanism. All his account demands is uptake from the well-prepared audience, and nothing more.

Does the meshing condition fare better than the uptake condition? Stecker disagrees (2008, pp. 40-41; cf. Davies, D., 2007, p. 83). One problem of the former is that it could end up allowing private meaning. Recall that the audience's uptake in Livingston's account is only a symptom. This means that there can be cases in which an intention meshes well with the text without being recognized by a well-prepared audience. This requirement seems too liberal because it leaves room for idiosyncratic meaning.

One motivation for upholding the meshing condition is to recognize that interpretation typically aims to dig out the *implicit content* of a work. If something is implicit, it may not be easily grasped or aptly recognized; very often we have interpretative queries because the content of the work is implicit. To preserve this observation, a minimal success condition

such as meshing is preferable. This is because, if a stronger condition such as Stecker's is adopted, the implicitness of work-content would be automatically removed, which means that the disambiguating function of authorial intent would be lost. This brings us back to the internal conflict in Stecker's account: the author's intention seems epistemically idle, because, if we take the uptake condition seriously, the author's intention never disambiguates meaning, as meaning is always what the audience is apt to recognize (the one most likely to secure uptake from the audience), even in the case of failed intention (cf. Davies, D., 2007, p. 81). The moral is that Stecker's success condition ends up being too strong for the author's intention to function even as a heuristic. As discussed, Stecker contradicts himself by claiming that the author's intention typically disambiguates meaning at higher levels of interpretation. This claim entails that linguistic ambiguity is so prevalent that it is often not an easy thing for the well-prepared audience to discern salient meaning from the work. In that case the uptake condition would not make much sense, for the well-prepared audience would tend to recognize only ambiguity instead of clear-cut contents.

The tricky part here is that the second horn of the dilemma goes against Livingston as well. That is, accepting that the meaning of a work is typically implicit amounts to acknowledging the pervasiveness of textual ambiguity, which is the phenomenon that the meaning of a linguistic expression is often unclear in the sense that it can sustain many different possible meanings. Such being the case, how does Livingston account for the determination of work-meaning when the author's intention fails? At one point he writes: "Some implications are, of course, conventional, but other implied meanings of an utterance derive from relations between the author's intention, features of what is said, written, etc., and the context." (Livingston, 2005a, p. 150) This leaves some room for convention and context to play their role, but this room, again, appears to be too little to serve the function of convention and context, because the combination of the two can seldom point to one salient meaning in the work.

What puts Livingston in a more difficult plight than Stecker is the exact role of context. At some places, such as the aforesaid quotation, Livingston mentions convention together with context. However, at other places, particularly where he talks about the meshing condition (see the previous section), only convention is referred to. If what Livingston has in mind is the latter, he would face two problems. First, without the help of context, textual ambiguity will be as pervasive as it is on Carroll's account. Second, excluding context runs against his defence of the utterance model (Livingston, 2009, pp. 94-99); ignoring context also means Livingston has to meet the difficulties presented by Stecker for Carroll (see §4.3.1.1).

To conclude, partial intentionalism is ambivalent between modest actual mentalism and the unified view, and more or less bears similar problems. One merit of it is that it articulates the weak standard of success condition in Carroll's theory. Where Carroll is unclear about what counts as success in the minimal sense, Livingston elaborates it. Despite this contribution, partial intentionalism, as discussed, is still trapped in the same predicament as its kin.

4.3.4 Two dilemmas for modest actual intentionalism in general

Modest intentionalism is the most plausible version of actual intentionalism, as it not only acknowledges unintended meanings but also failed intentions. Its great sensitivity and sophistication makes it the strongest contender in the camp of actual intentionalism. However, the said sensitivity and sophistication also creates some general worries that undermine the theory's plausibility.

The biggest worry, as I see it, is the dilemma of ambiguity, which has been discussed several times. I believe this is the most critical blow to modest actual intentionalism. The modest intentionalist has great difficulty in giving equal credit to authorial intent on the one hand, and convention and/or context on the other: either convention/context is effective in determining meaning or not; if it is effective, it achieves at the cost of authorial intent; if it is not effective, the attempt to involve convention/context automatically fails. Either horn renders futile the disjunctive formulation of modest intentionalism. As far as I know, no modest intentionalist has yet offered a solution to this problem.

Another notable dilemma—the *epistemic dilemma*—is proposed by Saam Trivedi (2001). The dilemma starts with the epistemic question: how do we know whether an intention is successfully realized? Presumably, we figure out work-meaning and the artist's intention respectively and independently of each other. And then we compare the two to see if there is a fit. Nevertheless, Trivedi claims, this move is redundant: if we can figure out work-meaning independently of actual intention, why do we need the latter? And if work-meaning cannot be independently obtained, how can we know it is a case where intentions are successfully realized and not a case where intentions failed? It follows that appeal to successful intention results in redundancy or indeterminacy.

Stecker rejects the second horn by claiming that when an intention is successful we do not need to know it to know work-meaning (2010, pp. 154-155).[53] In interpreting an utterance, we form hypotheses about the utterer's intention; with the help of context and convention, we will eventually reach the right utterance meaning as well as the successful intention after eliminating wrong hypotheses.

However, this does not answer the question of whether we can know that the author's intention is successful or not, for we cannot know whether the identified utterance meaning matches the author's intention unless we compare it with the author's intention independently determined. Here Stecker is in danger of conflating utterance meaning with utterer's meaning (Trivedi, 2015, p. 704). In a related manner, Stecker's reply reinforces the first horn: if we can figure out work-meaning independently of knowing the author's intention, why do we still need the latter?

There is a better way to reject the second horn of the epistemic dilemma: the modest intentionalist does not determine the success of an intention by comparing independently obtained work-meaning with the author's intention. As already discussed, modest intentionalists propose different success conditions that do not appeal to the identity between authorial intention and work-meaning. For Carroll, success is defined by the intention's compatibility with the text; for Stecker, it is defined by the audience's ability to grasp the intention; for Livingston, it is defined by the degree of meshing.[54] The second horn of the dilemma therefore fails, for even if work-meaning cannot be independently obtained, we can still know whether the artist's intention is successful or not.

The first horn of the dilemma assumes that work-meaning can be obtained independently of knowledge of successful intention, but this is false for Carroll and Livingston, for they explicitly assume a substantial degree of textual ambiguity. But that result immediately makes them fall prey to the second horn of the dilemma of ambiguity. In conclusion,

---

[53] Note that this is not a reply to the circularity problem, which is a *constitutive* challenge to how a successful intention determines work-meaning. Rather, this is a reply to the *epistemic* challenge to how the audience knows whether or not an intention is successfully realized in a work. For this distinction, see Trivedi, 2015, p. 701. Despite the distinction, Trivedi admits that these two issues are somehow related. But he does not say how.

[54] Defending modest actual intentionalism (especially Carroll's brand), Lintott (2002) makes a similar point, claiming that success is measured by whether a given semantic intention is compatible with one of the work's possible meanings. Trivedi seems to assume that a work can always have a salient meaning, as indicated in the next paragraph of my discussion.

Carroll's and Livingston's accounts are immune to the epistemic dilemma, but are susceptible to the ambiguity dilemma; both dilemmas, unfortunately, work against Stecker's account.

4.4 Objections to actual intentionalism in general

Here I consider three representative but surmountable objections, and two big challenges to actual intentionalism in general. The former are frequently expressed worries in the literature; however, the latter are usually treated as presenting less serious problems. I think otherwise. My opinion is that the latter two are real threats that significantly reduce the plausibility of actual intentionalism. I start with the three surmountable objections.

A recent objection is the *publicity paradox* (Nathan, 2005). The main idea is this: when S means p by U there is a second-order intention that the audience need not go beyond U to reach p; that is, there is no need to consult S's first-order intentions to understand U. Similarly, when an artist creates a work for public consumption, there is a second-order intention that her first-order intentions not be consulted; otherwise it would indicate the failure of the artist. Actual intentionalism hence leads to the paradoxical claim that we should and should not consult the author's intentions. I discuss this objection and Stecker's reply in more detail in Chapter 6, but briefly note here that I think neither is successful.

A second important objection is this: given that the goal of interpretation is to offer us aesthetic satisfaction through understanding the work's literary language, we should focus on the work itself rather than something outside it (Beardsley, 1970, p. 34). To this Stecker convincingly argues that, typically, the initial aim of interpretation is to understand a work's implicit content; by doing so we get aesthetic satisfaction. It is authorial intent that achieves this implicit meaning (Stecker, 2010, p.147).

A third objection concerns secret meanings. It seems implausible to require the interpreter to attend to this kind of meaning. If a novel contains an intended secret code that can be deciphered only by reading the first word on every odd page, the code should not qualify as a valid literary meaning. I think that we should distinguish between two senses of secret meaning: secret code and idiosyncratic meaning. For the former, it may be controversial to view it as literary meaning, and the actual intentionalist can accordingly dismiss it. For the latter, the modest intentionalist can invoke the weak success standard or insist that it is crucial to make room for this kind of meaning, because even the anti-intentionalist tries to accommodate it by licensing intermediate evidence. (Modest intentionalism does not seem to be very clear about which way to go).

Now I consider two objections which I think are real challenges to actual intentionalism. I call the first the *epistemic worry*, which asks: is intention knowable? Actual intentionalists tend to dismiss this worry as insignificant, and maintain that in many contexts (daily conversation or historical investigations) we have no difficulty in discerning another person's intention. In that case, why would things suddenly stand differently when it comes to literary interpretation? (Livingston, 1996, pp. 80-81; Carroll, 2002, p. 326; Stecker, 2013, pp.311-312) This is not to say that we succeed on every occasion of interpretation, but that we do so in an amazingly large number of cases. That being said, we should not reject the appeal to intention solely because of the occasional failure.

Most of the opponents of actual intentionalism seem to feel comfortable with this reply, but I hold an opposite view based on two reasons. First, the fact that we do often fail to discern the intention of our conspecifics is significantly downplayed by actual intentionalists. Miscommunications and misinterpretations do happen a lot in our life (and probably more in historical speculations), not to mention that there are cases in which we think we have successfully grasped the speaker's intention but actually we have not, and cases in which misinterpretations do not affect our interaction with people. The ground on which the actual intentionalist relies is shaky.

Second, even if we grant that in daily life successful cases prevail, it is less clear that the same stands in the case of literary interpretation. Most literary works are very much longer than daily utterances and successfully grasping the utterer's intention becomes much more difficult. Moreover, it is very important to note that, in daily conversation, it is a rule of thumb to get what we say across as clearly as we can; a conversation is considered bad if one keeps saying things that are ambiguous or unclear. Nonetheless, this is not the case in the production of literary works. In the realm of literature, producing works with ambiguous or obscure contents is not only legitimate but also encouraged, especially in the case of poetry and so-called serious literature. For example, a story with an open ending, whether the author in fact intends it to have one particular reading or not, would normally be praised by the critic and the reader for its multiple interpretability, but we would normally expect a conversation to end with clear and understandable remarks. Puzzlement does not merit acclaim in everyday discourse.

The conclusion is that the success rate in daily interpretation cannot be happily maintained in literary interpretation, which renders unconvincing the interpretative policy recommended by actual intentionalism. I hence have strong reservations about the easy access to authorial intention as suggested by the actual intentionalist.

The modest intentionalist might object that we do not attempt to infer utterer's intention from the text, but from the author's proclamations of her intentions, which, presumably, could be as short as daily utterances. But this contradicts the claim that the text is the best guide to authorial intent (as made, for instance, by Carroll, 2000, p.77). Abandoning this reply brings actual intentionalism back to the Beardsleyan criticism that the interpreter easily shifts their attention to external evidence of authorial intent and ignores the text.

To make things worse, this reply only reinforces the epistemic worry that, in the case of literature, the utterances are typically harder to decipher than daily utterances, and this is why we often appeal to information outside the text for clarification about work-content, which is not the case in everyday conversation.

My second challenge to actual intentionalism involves a different take on Beardsley's argument of unavailability (1970, p.33). According to Beardsley, in the usual case the evidence of authorial intent is unavailable and that gives us a good reason to focus only on textual meaning rather than authorial meaning. There are several reasons why this kind of evidence is unobtainable: (1) the author is dead and there is no way to consult her; (2) the author is still alive but never reveals her intention; (3) the author has revealed her intention, but it is difficult for the reader to obtain it (for example, the revelation takes the form of private sources, or the sources, though public, are difficult to find). Indeed, the impression seems to be that, for most literary works, authorial intent tends to be absent due to at least one of the above three reasons.

However, Beardsley thinks this argument is inconclusive because the actual intentionalist can claim that we appeal to the text to discern authorial intent. My inclination is to think that Beardsley was not aware of the real force of the unavailability argument. The real thrust of this argument does not lie in the unavailability of authorial intent, but in the *consequence* of it: the unavailability in question creates a number of *provisional* interpretations whose provisional status might remain forever. The author's intention is normally unavailable unless she is asked to reveal it, which does not happen to most authors. This means that many of our hypotheses about the author's intention are bound to remain provisional, if we follow the actual intentionalist's recommendation.

To worsen the matter, we need to be aware that the meaning of a work is a composite: as Beardsley reminds us, it contains local meanings and regional meanings. Even if the author has revealed her intention, it is almost always only part of this complex of meaning. To get the correct meanings of words, phrases, passages, chapters—even only for a selective set—a

die-hard follower of actual intentionalism will need to prove all of her hypotheses for all of the relevant linguistic units in a work, which seems to me an impossible task.

If theories of interpretation are expected to be practically useful for interpreters—apparently they are—then actual intentionalism would not be a very ideal position, for the above two concerns create obstacles that prevent the interpreter from arriving at correct interpretations.

4.5 Chapter conclusion

So far I have critically discussed three versions of actual intentionalism, that is, absolute intentionalism, extreme intentionalism, and modest intentionalism, from the least to the most convincing.

Absolute intentionalism is the least appealing because it has unacceptable results: it leads to a Humpty-Dumpty theory of meaning, which is apparently untenable. Extreme intentionalism claims only that the author's intention is a necessary condition for meaning-determination, but sacrifices the possibility of unintended meanings. Modest intentionalism is the most plausible version of actual intentionalism, but it too suffers from some major difficulties, notably the dilemma of ambiguity. Also, there are major worries that haunt actual intentionalism in general. It remains to be seen whether the actual intentionalist has convincing replies to the epistemic worry and the updated version of the unavailability argument

Based on the discussions so far, I conclude that actual intentionalism is not an appealing position in regard to literary interpretation, even though it has merits on its own. Notably, the last two challenges just discussed point to hypothetical intentionalism, which is the focus of next chapter.

5. Hypothetical Intentionalism

"A work of art says what, on the basis of the work contextually construed, it would be reasonable to impute to its artist as a view that he or she both significantly held and was concerned to convey."

—Jerrold Levinson, "Messages in Art"

A compromise between actual intentionalism and anti-intentionalism is *hypothetical intentionalism*, the core claim of which is that the correct meaning of a work is determined by the best hypothesis of the author's intention made by a selected audience. The aim of interpretation is then to hypothesize what the author intended when creating the work from the perspective of the qualified audience.

Two points call for attention. First, it is hypothesis, not truth, that matters. This means that a hypothesis of the actual intention will never be trumped by knowledge of that very intention. Second, the membership of the audience is crucial because it determines the kind of evidence legitimate for the interpreter to use.

Hypothetical intentionalism was first proposed by William Tolhurst (1979), later modified by Jerrold Levinson (1992, 1995, 1996, 2006, 2010) and Saam Trivedi (2001, 2004, 2015). The main difference between these variations lies in the qualification of the audience. The audience concerned is described respectively by the three proponents as *intended*, *ideal/appropriate*, and *competent*.

In what follows I will discuss the three versions in said order.

5.1 Tolhurst's theory

5.1.1 The utterance model

Tolhurst aims to settle the early dispute between Beardsley and Hirsch. As already mentioned, Beardsley claims that textual meaning is determined by word-sequence meaning in virtue of linguistic rules; Hirsch claims that textual meaning is determined by the author's intentions. Tolhurst thinks that these two theories go to extremes, so he tries to find a place

somewhere in between. He argues that the author's intention is relevant to, rather than constitutive of, textual meaning. [55] Therefore, Beardsley is wrong in saying that author intention is irrelevant, and Hirsch is wrong in saying that author intention is necessary for meaning.

Tolhurst begins by distinguishing three types of meaning: *word-sequence meaning*, *utterer's meaning* and *utterance meaning*. Though these have been discussed in previous chapters, it would be useful to restate them, since these concepts are crucial in the debate.

Word-sequence meaning, alternatively called *sentence meaning*, is meaning determined solely by the linguistic conventions of the words concerned, regardless of the occasion on which the word sequence is produced; this view, according to the orthodox interpretation of Beardsley, is Beardsley's position. Utterer's meaning is whatever an utterer means by her utterance. This is alternatively called *speaker's meaning*, in contrast to sentence meaning. Utterer's meaning is identified with the intention the speaker had in mind when producing a word sequence. This last construal of meaning is endorsed by Hirsch.

To give a very simple example from ordinary conversation. If I say to you "You're such a great genius" after reading the novel written by you and really appreciating it, what I said is that you are a really talented writer (call this Case 1). But the same words do not mean the same thing when you have done something extremely stupid, e.g., mistaken salt for sugar and put it in a cake (Case 2). In the latter case, I am not saying that you are a really talented cook but that you are unusually dumb.

The point is that sentence meaning cannot capture what words convey nonliterally, and hence is insufficient as a theory of meaning. The account of utterer's meaning does not fare better, not only because people often fail to say what they mean, but they cannot intend any sentence to mean just anything. Neither in Case 1 nor 2 can one make the sentence mean "Pass me the salt" simply by intending it to mean so; it is also unsatisfactory to say that here

---

[55] Notice that the distinction between work and text, as mentioned, is probably not cautiously drawn until Currie 1991a (see also Livingston, 2005, pp. 112-127; Lamarque, 2007, pp. 71-78; Shusterman, 2009). Very often contemporary philosophers of literature use the terms "text" and "word sequence" interchangeably, both of which refer to a contextless string of words. In contrast, the term "work" refers to a contextualized literary text. Tolhurst is basically using "text" to mean this-word-sequence-as-written-by-this-person, literary or nonliterary. However, Currie in the said article interprets Tolhurst as attaching text-identity to its causal origin (see also Livingston, 2005, pp. 116-119), while he soon admits that this is just a matter of terminology, and that Tolhurst is roughly making the same distinction using a different term. For convenience's sake, I will follow Tolhurst's usage of the term when discussing his view.

the utterance becomes meaningless, as Hirsch would claim, for apparently it is not. Neither sentence nor speaker's meaning captures what is said in Case 2. Tolhurst thinks that to correctly interpret texts in general, we need to construe textual meaning as utterance meaning.

An utterance is not to be identified with a word sequence *simpliciter*: it is a word sequence embedded in a particular time and place. Therefore, utterance identity and meaning are determined in part by the word sequence uttered and in part by the context of discourse in which it is uttered. My uttering "You're such a great genius" in Case 1 counts as a different utterance than that in Case 2 because the word sequence concerned is produced in different contexts. Also, the contexts are diverse enough for the words uttered to generate different meanings. Of course, none of this excludes the possibility that two utterances can have the same meaning. We can imagine Case 3 in which the context is much like Case 1 (for example, you have produced another great novel) and my uttering the same word sequence would have the same meaning as that in Case 1.

Tolhurst suggests that we construe work-meaning as utterance meaning. Since the production of a literary work is like the occasion of a word sequence, we will interpret a work the way we interpret an utterance. This avoids the extremes of treating a work as a brute text or, alternatively, as an infallible embodiment of authorial intent. This analogy is later called the *utterance model* (Meiland, 1981). As said, subsequent philosophers engaging in the debate of interpretation are more or less committed to this model.

5.1.2 Textual individuation

Tolhurst points out that there is one serious problem with the utterance model, and he manages to offer a solution to it. However, most proponents of the model do not pay attention in this regard. It seems to me that Tolhurst's solution may need to be assumed to make tenable any theory of interpretation based on the utterance model.

As shown, the relationship between a word sequence and the utterance of it is a type/token relationship. A difficulty stands if we view a text as an utterance: a text still seems to be a type of some sort because it can have many tokens. A published text is bound to have many copies and very often even reprints, but surely we do not want to say that all the reproductions count as different utterances. A theory of textual individuation is needed if we want to hold on to the utterance model (Tolhurst & Wheeler III, 1979).

A text is a type in the sense that its tokens are tokens of it not solely in virtue of the word sequence instantiated, but also in virtue of a particular causal relation the tokens stand in to

the original instance. We may use the term *primary token* to refer to the original production of the word sequence of the text, and the term *replica* for the subsequent copy of such a product. What makes a replica a token of a text are the word sequence concerned plus the right causal relation. Typically this causal relation can be explained by intentional human agency: my copy of *And Then There Were None* is a token of Christie's work of the same title because the publishing house intentionally copied down the text produced by Shakespeare. This is at least a sufficient condition for textual identity. Tolhurst thinks that it might not be necessary to give an account of necessary conditions for textual identity, since the present account explains the normal situation.

Given this theory of textual individuation, a text can be a type of some sort even though textual meaning is utterance meaning.

5.1.3 The intended audience

So far we have been talking about utterance meaning as if we already knew the underlying mechanism of it, but the notion needs unpacking. We have seen that moderate intentionalists view utterance meaning as a function of the actual intention, linguistic conventions and perhaps context in place, while Tolhurst takes a somewhat different route.

Utterance meaning, according to Tolhurst, is understood as the *hypothesis of utterer's intention* which is most justified based on those beliefs and attitudes which one possesses qua intended audience. In the case of literature, textual meaning is construed as the intention best justified in being attributed to the actual author by the intended audience based on the attitudes and knowledge they have in the context in which the text was created. What the interpreter does is make hypotheses of such intentions, relying on the resources available to the intended audience. Two reasons support Tolhurst's formulation: (1) understanding what a person means by her utterance is essentially a matter of discerning what she intended by it; (2) to interpret an utterance the audience is normally not required to have a detailed knowledge of the speaker's psyche.

To apply the theory to literature, consider Jonathan Swift's satirical essay "A Modest Proposal," the narrator of which suggests that the poor in Ireland might ease their economic pressure by selling their children as food to the rich. The text, taken as a word sequence, hence endorses infanticide and cannibalism. However, if it is taken as an utterance, its meaning would be quite different. The text appears to be adequately understood as ironic, and the theory of utterance meaning just proposed helps to explain how we reach this reading.

Based on the relevant attitudes and beliefs of the audience Swift was addressing, the best hypothesis of the author's intention would be that the text is a satire that criticizes the heartless attitude toward the poor and Irish policy in general. That would be the intention best justified in being attributed to the author by the intended audience (whether or not that was the one s/he actually had).

The spirit of Tolhurst's theory can be summed up as follows: authorial intent is relevant in the sense that, by and large, the reader is hypothesizing such when interpreting a literary text; but it is also true that the reader need not possess detailed knowledge of the author's mental life when engaging in interpretation. The actual author's intention hence is not in a privileged position with respect to interpretation. In consequence, the theory of utterance meaning successfully captures the insights of actual intentionalism and anti-intentionalism, and avoids the weaknesses of both sides.

5.1.4 Some reservations about the notion of the intended audience

Tolhurst himself concedes that there are concerns with the concept of an intended audience, but he claims that they are not serious problems. The first concern is that sometimes it is hard to determine who the intended audience is, for they might not be whom the author seems to be addressing. In most cases we can easily identify who the intended audience is. For example, if an article is printed in a magazine, it is reasonable to suppose that the intended audience is the average reader of that magazine. Tolhurst thinks that exceptional cases are rare and this concern is hence not a practical problem.

Second, in some cases the intended audience might be better put as the *qualified audience*; that is, the membership of the audience is determined by a certain kind of qualification the author expects them to have. This happens when the author produces a text that requires the reader to possess a certain level of knowledge for interpretation. For example, the author writes a very difficult novel full of metaphors from a wide range of disciplines; only a polymath can suitably make sense of the story. In this case, the author does not have any particular group of people in mind as her audience; instead she is addressing an idealized audience that might or might not exist. Or, to give another example, the author is visionary and produces a novel only readers in the distant future can fully appreciate. In this case the qualified audience the author has in mind does not belong to any extant group of people but is something that might never come into existence. However, it might also happen that the

author is too inept to make her intention grasped by the qualified reader, since the text in this case tends to be a more challenging one to compose.

Tolhurst expects the objection that there should be constraints on the qualification the author demands from the reader; otherwise the author can always require an unduly esoteric base of knowledge in the reader. But this is not a big problem, for authors normally expect and desire to be understood, so unreasonable demands on the reader would be automatically ruled out by this prudence.

But the notion of the intended audience is more problematic than Tolhurst thinks. With the first concern mentioned above, the problem is that in many cases the author publishes a text in some medium not because she targets as audience the average reader of that medium but because other media reject the publication of her text. I do not have any statistics about this, but I think that typically authors aim for the best medium for publication but the rejection rate of such a medium tends to be the highest. This may be enough to show that it is dangerous to identify the intended audience with the average reader of the medium in which a text is published.

As for the second concern, the most obvious problem with it is that Tolhurst allows that the qualified audience can be an empty class at the time of a text's production. In that case, it is hard to see how the theory of utterance meaning is going to work because there is no uptake from the audience. Moreover, since the audience does not exist at all, we cannot anchor them in any context; absent the context, we are thus unable to have knowledge of the audience's relevant beliefs and attitudes by which we hypothesize the author's intention. If the intended audience is a necessary condition for meaning, we would be forced to claim that in this case the text is meaningless (Dickie & Wilson, 1995, p. 244).

What the idea of a qualified audience reveals is the fact that in many cases the author does not have any specific kind of audience in mind, especially in the case of fiction writing. When, to give an example, a New Zealand crime writer at a certain time publishes a crime novel set in New Zealand, it is reasonable to assume that her intended audience are crime fiction enthusiasts in New Zealand at that time. But the intention here is very general, indefinite, and vague. No doubt the author may intend her novel to be read by the said readers, but she may also intend her novel to be read by readers of other genres, countries, or even future generations. The fact is that normally authors of fictional works intend their stories to be widely read. If so, the notion in question needs to be more broadly construed. Given the generality of the intention, it may be better to define the audience by a broader qualification than intention. I am not denying that there are cases in which the author has a concrete

intention to produce a text for a particular kind of audience, but I doubt that this intention dominates over the more general intention that her text be read by a wider audience, especially in the case of fiction texts. Though this is an empirical claim open to demands for empirical grounding, it is reasonable to think that the point here can be made in terms of the nature of fictional works, or at least categories of fictional works. The dominant intention in question figures prominently in the case of popular fiction, which typically aims for fame and commercial success. Such being the case, popular fiction might be roughly defined as the kind of text made with an intention to make fame and profits by producing aesthetic satisfaction. There are of course counterexamples but this sketchy formulation shows the possibility to link the nature of fiction to the intention of reaching popularity.

5.1.5 Objections to hypothetical intentionalism based on the intended audience (1)

In this section I discuss some objections that can be disposed of quickly (Dickie & Wilson, 1995, pp. 244-245). I keep the more formidable objections for the next section.

The first objection is about the case of misspeaking. Normally in daily conversation we can discern what a speaker intended to say even if she misspoke. But Tolhurst's theory would compel us to claim that the best hypothesis based on the misspoken words yields the same utterance meaning as that whose words are correct. (For a similar objection, see Davies, D., 2007, p. 86). One can dispose of the case of misspeaking by offering a criterion for deciding the definitive word sequence of a particular text (Currie, 1991a, pp. 325-340; New, 1999, pp. 97-99). The text should be established before one sets to interpreting. Typically, the text to be interpreted is the one the author intends to produce. When an author makes a mistake due to a slip of the pen, we correct the text, for the author intends to write correctly. Correction in this case hence yields the definitive text.[56]

A second way to respond to this objection is to deny that the two utterances in question have different meanings. This is not a problem since Tolhurst's theory allows that different word sequences in different contexts can bear the same meaning. Suppose that I see a fly in your soup and warn you of that. Unfortunately, I misspeak "suit" for "soup." It is not unreasonable to say that my uttering "There is a fly in your suit" *in that context* means "There is a fly in your soup."

---

[56] Currie gives a sophisticated account of how to establish the text. This will be discussed in more detail when I come to Currie's account in the next chapter.

The second objection goes that most of the time we cannot infer utterer's intention without understanding utterance meaning: Tolhurst's theory gets us reasoning backwards. This objection is flawed, as the term "utterance meaning" used by Dickie and Wilson here cannot be the same as how Tolhurst defines it. If it were so, the argument would not hold because understanding utterance meaning requires us to infer utterer's meaning first. It follows that the term "utterance meaning" in the objection must refer to a different sort of meaning. The best candidate seems to be word-sequence meaning, since Dickie and Wilson in their paper are defending Beardsley's view. But Tolhurst does not deny that we should understand word-sequence meaning before inferring utterer's meaning. The hypothesis the interpreter makes is based on the relevant knowledge of context *plus textual data*. The interpreter needs to take into account the word-sequence meaning and hypothesizes utterance meaning based on that.[57]

The final objection states that, if one conceives of intentions as linguistic in nature, a serious problem will arise. On this view, having an intention implies having some set of words that bear a certain meaning for that intention so that one is able to represent the intended meaning to oneself. This objection works both against actual and hypothetical intentionalism. An utterer's hypothetical intention will then have to be understood based on another hypothetical intention with a word sequence representing it. This generates an infinite regress. However, as Dickie and Wilson admits, this objection works only for those who think that intentions must be linguistic in nature. For those who maintain that intentions can be nonlinguistic phenomena, this objection will be inconclusive.

5.1.6 Objections to hypothetical intentionalism based on the intended audience (2)

Tolhurst's work marked a new direction in the philosophy of interpretation (Meiland, 1981). However, this first version of hypothetical intentionalism is generally thought to be untenable. The most devastating objections centre on the audience-identifying intention; they indicate the weakness in Tolhurst's account.

One of these objections has it that Tolhurst's theory does not work when the author holds false views about the reading ability or relevant knowledge of the intended audience (Beardsley, 1982, pp. 189-199). For example, the author writes a story for eight-year-olds but overestimates the range of their vocabulary. In this case, the most justified hypothesis made

---

[57] Trivedi's version of hypothetical intentionalism is immune to this objection, as we shall soon see.

by the intended audience does not appear to be the most adequately attributed speaker's meaning.

It is also questionable why the audience-identifying intention can determine which kind of evidence is to be used in interpretation (Livingston, 2010, p. 408). If one group of the intended audience is familiar only with the public persona of the author, while another group with the author's true self (probably from private exchanges), then this sort of intention seems to imply arbitrariness of legitimate evidence for interpretation.

The most developed objection to the audience-identifying intention is raised by Nathan, who claims that Tolhurst's argument either leads to actual intentionalism or to anti-intentionalism (Nathan, 1982). On the one hand, if the intended audience is an extremely small group possessing esoteric knowledge of the author, meaning becomes a private matter in this case, for the text can only be properly understood in terms of private information between author and audience, and the difficulties characteristic of strong intentionalism arise. On the other hand, if the author mistakes the nature of the intended audience, as the first objection notes, why should we bother with the author's unreasonable audience-identifying intention (hence the evidence possessed by the intended audience)? It seems more reasonable to interpret the text based only on public facts. In that case, the interpretative method we follow turns out to be anti-intentionalist.

One way to respond to the first problem is to show that the author's audience-identifying intention is significantly different from and less troubling than her semantic intention (Levinson, 1992, pp. 227-228). But I do not think that Nathan disputes this; what he does dispute is that Tolhurst's view can escape from strong intentionalism in the special case in which the author uses an esoteric language known only to the intended audience. To satisfactorily get rid of the dilemma, one probably needs to abandon the intention qualification of the audience.

It can be seen now that all effective objections to Tolhurst's view attack the notion of an intended audience. At the close of Tolhurst's paper he concedes that there is still work to be done with this imperfect notion. This unfinished work is continued by Levinson, whose view I now turn to.

5.2 Levinson's theory

5.2.1 The ideal audience

Levinson replaces the concept of an intended audience with that of an ideal or appropriate audience. Such an audience is said to be familiar with the context in which the work was created, and seeks to anchor the work in its context of creation. More specifically, in Levinson's own words, the ideal reader is "versed in and cognizant of the tradition out of which the work arises, acquainted with the rest of the author's oeuvre, and perhaps familiar as well with the author's public literary and intellectual identity or persona." (Levinson, 1992, p. 228). So this class of audience is not necessarily targeted by the author's intention.

Levinson holds that the qualifications of the ideal audience are ground rules in critical practice and also constitute the implicit contract between author and reader. Following Tolhurst's line of thinking, it is unreasonable to require that the reader should possess detailed knowledge of the author's psyche and life. However, it is reasonable to require that the audience possess relevant knowledge to a certain extent, and this kind of audience is not necessarily a member of the intended audience. The notion of the ideal audience is a broader and more flexible one than that of the intended audience, effectively circumventing the objections to the audience-identifying intention.

Levinson also refines what counts as the *best* hypothesis. There are two senses of "best," epistemically best and aesthetically best.[58] The former refers to the hypothesis that is most likely to be true, the latter to the one that makes the work aesthetically better. In a key passage Levinson writes:

Principally, a "best" attribution is one that is epistemically best—that has the most likelihood of being correct, given the total evidence available to one in the position of ideal reader. But secondarily, a "best" attribution of intention to an author might involve, in accord with a principle of critical charity, choosing a construal that makes the work artistically better, where there is room for choice, so long as plausibly ascribed to the author given the full context of writing. In other words, if we can, in a given case, make the author out to have created a cleverer or more striking or more imaginative piece, without violating the image of his work as an artist that is underpinned by the total available textual and contextual evidence, we should perhaps do so. (Ibid., pp. 224-225)

---

[58] In this thesis I do not draw the terminological distinction between "aesthetic" and "artistic." For such a distinction, see Davies, 2006, pp. 53-55.

What is new here is that the aesthetic consideration is added to form the best hypothesis. Levinson seems to suggest that the aesthetic consideration comes as a tie breaker. That is, when we reach two or more epistemically best hypotheses, the one that makes the work artistically best wins out. This is a significant modification to Tolhurst's account, as it introduces a second criterion for construing what counts as a best hypothesis.

5.2.2 Categorial intention and multiple interpretations

Another important refinement Levinson makes is the distinction between *semantic intention* and *categorial intention*. This is clearly explicated in another key passage written by Levinson:

> An author's intention to mean something in or by a text T (a semantic intention) is one thing, while an author's intention that T be classified, taken, approached in some specific or general way (a categorial intention) is quite another. Categorial intentions involve the maker's framing and positioning of his product vis-à-vis his projected audience; they involve the maker's conception of what he has produced and what it is for, on a rather basic level. The most general of categorial intentions of concern here would be the intention that something be regarded as literature (or art) at all, which obviously enjoins certain modes of approach as opposed to others […] Semantic intentions […] do *not* determine meaning, but categorial intentions, such as a literature maker's basic conception of what is made, *do* in general determine how a text is to be conceptualized and approached on a fundamental level and thus indirectly affect what it will resultingly say or express. (Ibid., pp. 232-233, emphases original)

This is reminiscent of Kendall Walton's account of art categories (Walton, 1970), and here Levinson is following Walton in accepting the artist's categorial intention. Presumably, there are broad categories and narrow ones. Categories in the broad sense can mean different artforms (literature, music, sculpture, films, paintings, etc.) or sub-artforms (novels, plays, and poems are three sub-artforms of literature). Alternatively, they can refer to the classes of texts (in the case of literature) distinguished by an art-qualifying intention; that is, an

intention that the text be viewed as literature at all.[59] On the other hand, categories in the narrow sense typically refer to genres. Hence an author's intention that her work be viewed as a sci-fi rather than mystery novel is a categorial intention in the narrow sense.

Since categorial intention determines how a work is to be conceptualized on a fundamental level, it affects semantic meaning indirectly. For example, if a work is taken as a sci-fi story rather than a mystery story, some plots might be taken differently, because different sets of genre conventions and standards would be used to interpret the work. The same plot might have different implications in different literary genres. Or, if a text is taken as a grocery list rather than an experimental story, we will interpret it as saying nothing beyond the named grocery items.

Levinson claims that categorial intentions are not as transparently evident in the work as semantic intentions and hence should have a different status. This is so not only because categorial intention is inherently part of art making, but also because the projection of semantic intention to a text must be underpinned by awareness of categorial intention, display of which cannot be found in the text. Contra semantic intention, categorial intention is thus extrinsic and (virtually) infallible.[60]

We cannot miss out on categorial intention because if we do we would misidentify the work. We have to locate the author's work correctly before interpreting it. By contrast, semantic intentions do not matter that much for interpreters because they do not determine meaning, and are not a condition for the work's identity.

A final refinement Levinson makes to Tolhurst's view is how we deal with competing interpretations. It can be envisaged that for some works one can come up with more than one best hypothesis and the decision in this kind of case would be hard to make. Levinson suggests that, in the case where competing interpretations are available, we can always

---

[59] Trivedi (2015, p. 708) separates the last kind of intention from categorial intention; he calls it art-making intention. Either approach raises a worry: accepting art-making intention seems to commit oneself to a cheap definition of art (or literature), that is, something is art if and only if the maker intends it to be.

[60] Now, if one is committed to contextualism and accepts categorial intention as identity-relevant, one should indeed claim that a work's categorial status is externally determined. Such being the case, one should also accept that the author's categorial intention infallibly determines the work's categorial membership because as an identity-relevant factor, categorial intention along with other contextual features contributes to work-identity. Acknowledging the fallibility of such intention undermines its identity-conferring status and hence the contextualist premise. This is a potential problem for Levinson and other contextualists. I think a full treatment of categorial intention and its relevant problmes is in serious need. For reasons of space, addressing this issue awaits another occasion.

combine these first-order interpretations into a *global* or *subsumptive* reading of the following form:

> W's meaning is such that it is partly given by/aptly viewed under interpretation 1, partly given by/aptly viewed under interpretation 2,…and partly given by/aptly viewed under interpretation n, where those embedded interpretations, I1, I2, and so on, are understood as first-order sub-interpretations subject to the higher-order interpretation I*, which subsumes them, though not simply disjunctively or conjunctively. (Levinson, 2006, pp. 304-305)[61]

This way we can still talk about *the* meaning of a work, therefore preserving critical monism.

Levinson's theory is the representative version of hypothetical intentionalism and has been extensively discussed and criticized by many philosophers. I see a need to engage these objections and undertake this task in the next three sections.

5.2.3 Objections to hypothetical intentionalism based on the ideal audience (1)

I start with four objections that are based on misunderstandings of Levinson's view. It is important to clarify these misunderstandings so that similar objections won't arise again.

It has been claimed that allowing categorial intention threatens the autonomy of the work that hypothetical intentionalism aims to preserve (Iseminger, 1996, pp. 322-323). As a middle course between actual intentionalism and anti-intentionalism, hypothetical intentionalism avoids taking authorial intent as constitutive in order to make textual data and context central in fixing work-meaning. Allowing for categorial intention ruins this project. However, the sort of autonomy defended here is restricted (Levinson, 2006, pp. 308-309). The hypothetical intentionalist never purports to be a successor to the anti-intentionalist that seeks full semantic autonomy. As a middle course, defending a restricted autonomy seems to be a reasonable result. The objection hence ends up attacking a straw man.

A second objection is that hypothetical intentionalism does not accord with the actual practice of interpreters and critics (Carroll, 2000, pp. 86-87). Levinson responds that he is not defending the empirical claim; instead, he is defending the norms most reasonable (Levinson,

---

[61] Levinson (1992, p. 237) adopts the disjunctive approach but apparently abandons it here.

2006, pp. 309-310). Indeed, even Carroll later concedes that theories of interpretation cannot succeed on descriptive grounds; they have to be prescriptive (Carroll, pp. 117-135). The problem becomes how to offer a convincing argument for the norm one is defending.

Another frequently expressed worry states that different ideal readers are bound to produce different best hypotheses, because it is reasonable to claim that evidence only underdetermines hypotheses (Dickie & Wilson, 1995, p. 244; Livingston, 2010, p. 92; Carroll, 2013, p. 14). There are two ways to deal with this objection. In the first place, it might happen that some best hypotheses are only best in disguise; and after closer inspection consensus can be reached to exclude them (cf. Trivedi, 2015, p. 714). Recall that the aesthetic criterion has to be used to adjudicate between epistemically plausible interpretations. Presumably, there would not be many left after the aesthetic criterion is applied.

In the second place, if we do have competing best hypotheses in the end, then Levinson's suggestion of subsumptive interpretation comes in handy. These best hypotheses would then be subsumed to form a second-order interpretation.

Some might think that introducing the notion of subsumptive interpretations reinforces only the impression that critical monism is untenable in the framework of hypothetical intentionalism. A global interpretation as suggested by Levinson appears unwieldy and contrived. I totally agree. But then one might need something more than unwieldiness and contrivance to prove Levinson's suggestion untenable. This is an issue I won't pursue. My point here is to draw attention to Levinson's suggestion, which has been largely overlooked by his critics.

The final objection goes that the theoretical apparatus of hypothetical intention is at best superfluous (Goldman, 2013, p. 80). If the motivation of this view is to make a good balance between accommodating unintended meanings and maintaining objective constraints on interpretation, the value-maximizing theory does a better job. The apparatus of hypothetical intention is superfluous when the best hypothesis matches the value-enhancing reading; it is irrelevant when it does not. However, the objector is attacking a straw man here because, for Levinson, aesthetic consideration is not a top priority in interpretation. For Levinson, making the said balance does not entail that enhancing appreciation is the central aim of interpretation. A hypothesis is best in the sense that it is epistemically best, namely, most likely to be what the author intended; aesthetic considerations kick in only when there is room. Such being the case, the hypothetical intention is neither superfluous nor irrelevant, for it does not aim for a value-maximizing reading of the work.

5.2.4 Objections to hypothetical intentionalism based on the ideal audience (2)

Now I turn to three objections that are more thoughtfully presented. Countering them demands unpacking the claims made by hypothetical intentionalism.

The first objection has it that excluding authorial intent from extratextual considerations is arbitrary (Savile, 1996; Stecker, 1997). But this is exactly the crux of the debate: whether external evidence of the author's (semantic) intention is relevant to interpretation or not. If this is the debating point then it would not be a good idea to simply say that excluding or including such evidence is arbitrary. Other arguments are needed to prove the point.

Levinson's own response is less convincing. He claims that, one of the ground rules of the literary game mandates that, to interpret, the audience does not need the author's proclamations of semantic intent (Levinson, 2006, p. 306). But this reply creates only more difficulty: it may not be empirically borne out. Is there such a rule? If there is, many readers or critics do not abide by, or are not even aware of it.

The second objection is that hypothetical intentionalism is in danger of not being concerned with the author and her achievement (Iseminger, 1996). The hypothetical intentionalist can reply that authorial intention is what is aimed at for hypothetical intentionalism: it is essentially relevant. Also, the hypothetical intentionalist locates the achievement of the author in utterance meaning.

Two additional remarks can be said about this. First, hypothetical intentionalism does not imply that the author is absolved of responsibility (moral or otherwise) for meaning best attributed to the work (Levinson, 1995, pp. 184-198). Second, we can distinguish between two senses of how an author means by her text. If a work means *p*, the author means *p* in the *broad* sense that *p* is the meaning best attributed to her by the ideal audience; but this does not entail that she means *p* in the *narrow* sense, that is, in the sense that *p* is her intended meaning (Levinson, 1996, pp. 211-212). Targeting authorial meaning in the broad sense thus still commits us to the actual author and her achievement. Levinson calls this the broadly communicative model of literary activity, in contrast to the narrowly communicative model (Levinson, 2006, p. 306).

The last objection is this. A hypothesis is devised to seek truth. If we eventually obtain access to truth that turns out not to coincide with our best hypothesis, why should we stick to the latter (Carroll, 2000, p. 83)? Levinson's reply (Levinson, 2006, p. 308) is that the literary game aims for utterance meaning, not utterer's meaning *for its own sake*, and the former is determined by the best hypothesis made by the ideal audience about the utterer's meaning. In

other words, the analogy with science is false. In literature truth is just determined by the best hypothesis based on the author-specific context.

It could be argued that the present proposal substitutes the notion of warranted assertibility for truth, and it is truth that we are eventually after. But Levinson explicitly says that warranted assertibility does not constitute truth (Levinson, 1992, pp. 250-251). To be precise, it does not constitute the truth for utterer's meaning, but it does constitute the truth for utterance meaning. The ideal reader's best hypothesis constitutes utterance meaning, even if it is designed to infer utterer's meaning. Perhaps this is why Levinson also dubs hypothetical intentionalism *constructive intentionalism* (ibid., p. 221). If we speak of construction of intent rather than hypothesis of intent, the question of truth vanishes.

5.2.5 Objections to hypothetical intentionalism based on the ideal audience (3)

The objections discussed so far are more or less surmountable. Now I get to real challenges to Levinson's view. These objections have led to fervent debates and it is not clear Levinson has successfully countered all of them. I start from objections to which Levinson has an answer to the philosophical critics that give him a harder time.

A troublesome objection goes that hypothetical intentionalism collapses into the value-maximizing theory, for when making the best hypothesis of what the author intended the interpreter inevitably attributes to the author the intention to produce a piece with the highest degree of artistic value that the work can sustain (Davies, 2007, pp. 183-185). That is, the epistemic criterion for determining the best hypothesis is inseparable from the aesthetic criterion.

In reply, Levinson claims that this objection may stem from the impression that an author normally aims for the best; however, this does not imply that she would anticipate and intend the (textually consistent and) artistically best reading of the work (Levinson, 2010, pp. 141-142). It follows that it is not necessary that the best reading is what the author most likely intended, even if she could have intended it. Davies's reply is this: still, the situation in which we have two epistemically plausible readings while one is inferior cannot arise, because we would adopt the inferior reading only when the superior reading is falsified by evidence (Stecker & Davies, 2010, pp. 308-309).

It seems to me that there is a deep disagreement here: Davies thinks that we will take the superior one, because it is always epistemically more plausible to hypothesize such; yet

Levinson thinks that this is not always the case. What Levinson says seems to accord with my intuition better, as I found it not too difficult to imagine a case he might have in mind.

For example, *Journey to the West*, one of the Four Great Classical Novels of Chinese literature, has yielded several equally plausible interpretations and there is still no consensus on which is the best. Among these interpretations, one maintains that the novel was intended to be purely cynical and humorous; on this view, the novel has no serious messages to convey and should be taken more or less at face value. Another interpretation claims that the cynicism in question should be read as ironic; it turns out that the novel, disguised as a humorous fantasy story, is actually a satire full of political symbols and metaphors. Both interpretations are widely accepted as legitimate and cohere with textual evidence, but one might think that the first is artistically inferior to the second.

The second objection is that the distinction between public and private evidence is blurry. Is public evidence published evidence? Does published information from private sources count as public (Carroll, 2000, pp. 92-93)? Levinson emphasizes that this is not a distinction between published and unpublished information (Levinson, 2006, p. 310). He then suggests without further elaboration that the relevant public context may be re-construed as what the author *appears to have wanted* the audience to know about the circumstances of the work's creation. This seems to mean that if it appears that the author has not wanted to make certain proclamation of intent known to the audience, then this evidence, even if published at a later point, does not constitute the public context to be considered for interpretation.

One thing Levinson does not make clear is whether public evidence includes the author's semantic intention publicly revealed. People who raise the present objection such as Carroll take Levinson to be saying "yes"; that is, the author's publicly (rather than privately) revealed semantic intention will be constitutive of work-meaning in the sense that it will be among relevant contextual factors on which the audience's best hypothesis is based.[62] In his reply to the objection, Levinson does not dispute this reading of him and seems to accept it. But if that is the case, Levinson simply contradicts himself here, for he explicitly maintains that semantic intention does not determine meaning. It seems that Levinson would need to exclude considerations of semantic intention entirely to make his theory consistent.

Stecker notably presents two formidable counterexamples to hypothetical intentionalism, which deserve scrutiny. The first counterexample to hypothetical intentionalism is that *W*

---

[62] Aside from Carroll, Goldman and Stecker also seem to interpret Levinson this way. See my discussion over the next objection to Levinson. An exception is Livingston, See his 1996, pp. 625-626.

means *p* but *p* is not intended by the author and the audience is justified in believing that *p* is not intended (Stecker, 2006, p. 437; 2010, pp. 159-160; see also Goldman, 1995, p. 115). Stecker's example is from the Sherlock Holmes adventures. It is famously known among readers that Dr Watson's war wound appears in two different locations. On one occasion the wound is said to be on his arm, while on another it is on his thigh. In other words, the Holmes story fictionally asserts impossibility regarding Watson's wound. But given the realistic style of the Holmes adventures, the best hypothesis of authorial intent in this case would deny that the impossibility is part of the meaning of the story, which is apparently false.

In this case the hypothetical intentionalist would not maintain that *W* means *p*, because *p* is not the best hypothesis (Levinson, 2010, pp. 143-144). She would not claim that the Holmes story fictionally asserts impossibility regarding Watson's wound, for the best hypothesis made by the ideal reader would be that Watson has the wound somewhere on his body—his arm or thigh, but exactly where we do not know. Stecker errs in presupposing that *W* means *p* without following the strictures imposed by hypothetical intentionalism to properly reach *p*.

An alternative way to deal with similar cases is this. If it is a suspected case of a typo, misspeaking or the author's mistake due to false memory, we should rectify the text before interpreting it. Therefore, in the present case the location of Watson's wound should be fixed first. Either it is on his arm or thigh, depending upon our strategy for correction. For example, if there is evidence showing that Conan Doyle intended the wound to be on Watson's arm, then the location should be changed to the arm (New, 1999, pp. 98-99). Or, if references to the arm outnumber those to the thigh, what is true in the fiction is the former; if they have equal numbers, what is true is that the wound is on one or the other but not both (Currie, 1990, p. 87). In that case, either all references to the arm are changed to the leg or the other way around.

Stecker replies that Levinson is inconsistent in specifying the semantic/epistemic conditions of fictional/actual texts (Stecker & Davies, 2010, pp. 309-310). In discussing the location of Watson's wound Levinson is inconsistent in claiming that (1) the text is in error; (2) the text asserts something (Watson has a wound somewhere), which seems to imply that there is no error. Stecker's objection fails, because, regarding (2), Levinson in his paper explicitly mentions "work" instead of "text." In other words, Levinson means that the work means something while the text is in error.

To revive the objection Stecker indicates more complications. Some inconsistencies, like visual imagination in films, reject rectification; some inconsistencies are woven into the plot and rectifying them would require extensive changes to the work. But Levinson is not saying

that we should rectify all inconsistencies; rather he is saying that in some cases such an inconsistency is apparently a mistake, and this does not rule out the possibility that in other cases it is not (Trivedi 2015, p. 719).

Finally, Stecker notes that ambiguity is another example in which the hypothetical intentionalist falsely denies that *W* means *p* when the audience is justified in not attributing *p*. An example at issue is the 2004 film *Swimming Pool*, which sustains several epistemically plausible interpretations, one of which could be what the director intended. Therefore, the work is also suitably viewed as ambiguous, which makes ambiguity another best hypothesis. Now, to adjudicate between these competing interpretations, the aesthetic criterion comes into play. Levinson claims that

> [I]t is false that the film means only what it was intended to mean by its maker, even where it can somehow be seen as meaning what it was intended to mean. The film should rather be taken to mean, ambiguously, many of the other options noted above, which are reasonably attributed to the filmmaker on both epistemic and aesthetic grounds. The film is a much richer, more satisfying, work of art when so viewed. (Levinson, 2010, p. 150).

The ambiguity interpretation hence beats other interpretations and should be taken as the correct interpretation of the film.[63]

Stecker maintains that what Levinson says is self-defeating (Stecker & Davies, 2010, pp. 310-311). In this case, Levinson apparently is prepared to say that the film is ambiguous (hence *W* means *p*) even if the director publicly reveals that he intended only one of the competing readings mentioned earlier (hence the ideal audience is justified in not attributing *p* to the director).

The point here is that Levinson is arguing under the premise that there is no public evidence about the director's semantic intention. Therefore, our best hypothesis, epistemically speaking, is the attribution of ambiguity. If the audience is not justified in attributing the intention of ambiguity to the author, then *W* will not be ambiguous.

Another way to deal with Stecker's objection (and his second counterexample, which is in a similar vein) is to stick to the point that semantic intention does not determine meaning. As

---

[63] It is not clear whether Levinson would call this ambiguous interpretation a subsumptive one. This is something he does not clarify. But in this case, the two look the same, given what he has said on subsumptive interpretations.

shown in the discussion of the previous objection, Levinson might need to ignore all evidence of semantic intention, public or private, to make his theory consistent. In the case of *Swimming Pool*, even if the ideal audience is justified in believing that the director does not intend the film to be ambiguous, this belief will not affect how the audience hypothesizes the director's intention, for semantic intention, even when publicly available, plays no role in fixing meaning.

Stecker's second counterexample to hypothetical intentionalism is the case where the audience is justified in believing that *p* is intended by the author but in fact *W* means *q*; the audience would then falsely conclude that *W* means *p* (Stecker, 2006, p. 437; 2010, p. 160). This objection is a variation on the previous one. As such I believe Levinson could circumvent it in the same way he did a moment ago. That is, what *W* (the *work*) means is determined by the ideal reader's best hypothesis based on text and context, not by what the *text* literally asserts. The meaning of the work is the product of a prudential assessment of the total evidence available. But in this and the previous objection Stecker ignores Levinson's construal of utterance meaning and presupposes what *W* means based on a notion of meaning different from the one stipulated by hypothetical intentionalism.

The final objection concerns the notion of categorial intention (Livingston, 1996, pp. 626-627; 1998, pp. 836-841; 2005a, pp. 159-165; 2010, p. 411). This is yet another troublesome challenge. Levinson claims that categorial intention has a constitutive status for work features because it seldom fails. Livingston questions this claim. Does not what Levinson says still imply that such intention is fallible? In Levinson's formulation, categorial intention covers both art-making intention and genre intention, the latter including both genre in the narrow sense and genre in the broad sense. It appears that categorial intention at any of the said three levels is fallible. We can imagine cases where someone writes a text that can be read perfectly as a poem but intends it to be a scribble; or where someone writes a short story but sincerely claims it to be a poem; or where someone directs an anti-war film but did not intend it to be such.[64] This shows that the intentional fallacy applies to categorial intention, too. One can neither safely infer from the premise that the author intends the text to be of a certain category *C* to the conclusion that the text indeed belongs to *C*, nor vice versa. With that said, there would be no principal difference between categorial and semantic intention in regard to their status.

---

[64] The commonly accepted anti-war interpretation of the 1988 Japanese war film *Grave of the Fireflies* is famously denied by the director, Isao Takahata.

Perhaps what Levinson means is that categorial intention is easier to realize. But according to Livingston, this is not true, either. Henry James claimed that he intended *The Turn of the Screw* to be a ghost story, but apparently many critics do not think that the story reads like that. An opposite case would be one in which an author intends to write something fecund of implicit contents but ends up producing something more straightforward. In contrast, many semantic intentions are not difficult to realize. No doubt it is reasonable to say that most authors successfully realize a fair number of their semantic intentions; otherwise we would have to say that their achievements are mostly based on accidents!

A second reason for giving categorial intention a superior status might be that such intention is more readily known. But this claim is again controversial. Are we really confident enough to say that when Shakespeare or Jane Austen wrote their stories, they intended them to be "serious literature" instead of sheer entertainments? The same question can be asked about many other works, not to mention that many authors do not seem to be clear about which genre they are aiming their work to be in. Discerning what the fundamental framework of a story is can be as hard as discerning what a sentence implicitly conveys.

Another unequal treatment of the two types of intention on Levinson's part appears in their implications for literary value. Levinson rejects semantic intention partly because it sometimes leads to a lower estimate of work value. But why not reject inferior categorial intention for the same reason? In regard to the implications for literary value, Levinson weighs categorial and semantic intention differently without saying why. This unequal treatment is unjustified. If it is because categorial intention plays a role in fixing the work's identity, the actual intentionalist can claim the same with semantic intention. But Levinson could reply that the debate is exactly about semantic intention, so Levinson for sure can hold different positions on the ontological and semantic level. Committing to actual intentionalism on the ontology of works does not oblige him to do the same with semantic interpretation.

I think Livingston's criticism on the status of categorial intention is not conclusive. Levinson's point is that categorial intention *virtually* cannot fail; that is, he does concede that such intention can, occasionally, fail. Livingston's criticism in this respect centers on fallibility, but this is something Levinson also agrees on. Indeed, as shown, Livingston points out that what Levinson means might be that categorial intention is easier to realize, but he does not think that this is true. This is a point the settling of which may call for empirical research, but my impression remains that people seem to have more debates about semantic intention than categorial intention. Is it not true that most papers I cite in this treatise focus on the former rather than the latter?

Moreover, Livingston overlooks the most important reason why categorial intention enjoys a superior status than semantic intention, which is a point on which Levinson puts great emphasis: categorial intention is more fundamental than semantic intention because the former indirectly shapes the latter. In other words, semantic intention is, to a great extent, subject to categorial intention. This higher-level dominance is enough to justify the superior status of categorial intention, and this is Levinson's real reason for weighing such intention more, a reason that escapes Livingston's criticism.

With all that said, Livingston seems right in saying that Levinson fails to consider inferior categorial intentions. If the aesthetic consideration is Levinson's sole reason for justifying the superior status of categorial intention, his account will be unfounded. But as I just showed, this is not the case. Levinson should have made his points without the aesthetic argument.

So far the worry is about the questionable status of categorial intention, but there is a second aspect of Livingston's objection (see also Carroll, 2002, p. 342). That is, the distinction between the two kinds of intentions is unclear, and this problem arises at different levels of categorial intention. For example, if we are committed to the view that fiction-making involves an intention to have the audience recognize the text's semantic contents, this intention would be a mixed one of categorial and semantic components. Or, imagine a case where the author intends to write a novel belonging to a trilogy:

> […] the intention, more specifically, is to create various meaningful, implicit relations between the characters in the three novels. The writer intends for the readers to think about the successive protagonists "as if" they were continuations of a single type of person. It seems hard to separate the categorial and semantic aspects of the content of such an intention, or cluster of interrelated intentions. (Livingston, 1998, p. 840)

Perhaps Levinson can reply that he never says that categorial intention is without the semantic dimension, for he does define such intention as what indirectly determines meaning. The semantic aspect of a categorial intention is only concomitant. If *The Turn of the Screw* was intended as a gothic-style ghost story, then it implies that the author also intended the existence of ghosts in the story. But this semantic content is only implied by the categorial intention in question, and is not part of the semantic contents to be hypothesized by the ideal reader. Agreeing with this point, it has been suggested that the distinction between categorial intention and semantic intention is *fluid* instead of rigid (Trivedi, 2015, pp. 710-711). For

example, the intention to satirize can be both. Livingston errs in saying the notion is not applicable rather than fluid.

To conclude, the most formidable challenges to Levinson's view, based on the above discussion, have something to do with crucial distinctions he made: the distinction between public/private evidence and that between categorial/semantic intentions. I have shown that, to meet the challenges satisfactorily, those distinctions need more elaboration and clarification.

5.3 Trivedi's theory

5.3.1 The competent audience

Saam Trivedi has developed a moderate version of hypothetical intentionalism which he claims improves on Levinson's theory. I do not see that his attempt is completely successful, for he says nothing at all about the objection to the private/public distinction, and it is not clear whether he would agree with the suggestion I made in this respect. However, as just cited, he does answer Livingston's challenge to the categorial/semantic distinction and counters some further objections to hypothetical intentionalism in general.

There are several significant differences between Trivedi's account and Levinson's. First, Trivedi replaces the ideal audience with the competent audience, whose best hypothesis is fallible. This is, as he claims, a divergence between warranted assertibility and truth. Second, the ideal audience hypothesizes the author's intention, while the competent audience hypothesizes work-meaning. These two points are related. Both Levinson and Trivedi claim that they do not intend to equate warranted assertibility with truth. But they differ in the object of the audience's hypothesizing. For Levinson, the best hypothesis about the author's intention is not equivalent to that intention, but it is constitutive of work-meaning. For Trivedi, the best hypothesis about work-meaning is not equivalent to work-meaning, but it can be revised to approach verisimilitude. It seems that Trivedi leaves unanswered the constitutive question of work-meaning, as the value-maximizers do.

The competent audience attributes the best hypothesis about work-meaning to the author as what she *should have intended*, as opposed to what she did intend or could have intended. This is because the best hypothesis about work-meaning does not necessarily coincide with what the author intended, so it is not appropriate to make the attribution as such; also, the notion of what the author could have intended is not an option because it is too broad (Levinson, 2010, pp. 140-141). What is best figured out by the competent audience is

ostensibly the correct work-meaning and, hence, what should have been intended by the author. This move is very different from Levinson's account in which the audience hypothesizes the author's intention first and then attributes it to the work as what the work means.

One might object that it is odd to say that the author should have intended *p* if *p* means the opposite of what she intended, especially when this involves abandoning her deeply held convictions. It seems more reasonable to say that the author should have written a work with different features. However, Trivedi claims that if we get *p* as the correct work-meaning, this means that the best evidence suggests so. That is to say, to convey the intention the author had in mind *successfully*, she should have written a work with different features.

The third difference between moderate and robust hypothetical intentionalism is that the former grants the author's intention in such cases as irony, allusion, or satire. In other words, Trivedi holds actual intentionalism on such literary features (figurative features). The ground for making this concession is that, as said, the applicability of categorial intention should be more flexibly construed. The intention to satirize may well be considered both as categorial and semantic, and its categorial implication should not be impaired by its semantic aspect. That is, the distinction between categorial and semantic intention should be seen as fluid rather than rigid. Where an intention is a mix of both kinds of intention we may well just accept both. As suggested in § 5.2.5, Henry James' intention to make *The Turn of the Screw* a ghost story is both categorial and semantic, and the fact that it has the semantic aspect does not invalidate its categorial function. This flexible construal deflects the worry that, if there is no clear-cut distinction between the two, then the distinction is not applicable at all.

As some might claim, a problem here is that, not all cases of irony, allusion, and the like are identity-relevant. For example, whether one short paragraph in a long novel is ironic or not does not seem to affect the overall work in a fundamental way. In this case, the intention concerned is more semantic than categorial. Then, why must this kind of case be treated in terms of actual intentions? I think this might not be a big concern for Trivedi, as he can always distinguish between cases where the said literary features are identity-relevant and those where they are not (for this approach, see Davies, 2007, pp. 177-178). That is, he can maintain a stronger position with respect to fluidity: the intention involving irony and the like can be *either semantic or categorial*, or both.

The final difference is Trivedi's elaboration on communicative interest. Though Levinson also emphasizes this, he does not distinguish between cases where the author's intention is

successful and cases where such intention fails. In his account there is no difference between communications in the two cases. But intuitively there appears to be a difference. According to Trivedi, if the artist successfully realizes her intention—that is, the hypothetically attributed meaning coincides with the actual intention—we have real communication; if she fails, we at least have a hypothesized or potential communication. Of course, real communication is better, but "either way we preserve the powerful idea of art involving communication—whether real or hypothesized, actual or potential—between artist and audience […]" (Trivedi, 2015, p. 716)

The above consideration sidesteps the objection that intention in Trivedi's account is superfluous because neither actual nor hypothesized intention plays a major role in determining meaning. This objection fails, as hypothesized intention plays a role in the second step of the account: the best hypothesis about work-meaning is attributed to the author as what she should have intended. The emphasis on the artist-audience communication does not render intention superfluous.

In sum, Trivedi's account is a middle course between moderate intentionalism and robust hypothetical intentionalism. While following Levinson, Trivedi sides with moderate intentionalism not only in preserving a communicative interest closer to that in ordinary discourse but also in acknowledging the author's intention to be ironic, allude or satirize. But this does not mean that his view collapses into moderate intentionalism, for he does not target successful intention.

5.3.2 Objections to moderate hypothetical intentionalism (and to hypothetical intentionalism in general)

Now, drawing on the new resources provided by moderate hypothetical intentionalism, Trivedi is able to defend hypothetical intentionalism from some major objections.

The first objection is that appeal to authorial intent in many cases is necessary because lack of such knowledge may falsely entail rewarding interpretations (Carroll, 2000, pp. 89-90). Carroll's famous example is the 1986 film *Stand by Me*, in which the narrator turns off the computer without saving the story he is writing about some childhood trauma. Does this mean that the narrator is prepared to let it go, or that this is simply a mistake on the director's part due to insufficient knowledge of using computers? Evidence suggests the latter. However, if we ignore the artist's semantic intention as hypothetical intentionalism dictates, we would opt for the former interpretation, which is more rewarding but arguably inadequate.

Again, one can reply that what matters is what the work means, not what the author means. If the film says *p* then the meaning of the work is *p*, even if the artist intended *q*. Moreover, we can always appeal to the context of the work's generation to make a decision. In Carroll's case, we should see if the context tells us that the film viewers knew much about computers. Since the contextual evidence suggests that knowledge of computers was not prevalent at that time, the competent audience's best hypothesis of work-meaning might not be the exorcism reading (Trivedi, 2001, p. 203).[65] I say "might" because the best hypothesis made by the competent audience is fallible, which means that both interpretations are likely to be proposed by different competent audiences, depending on how they interpret evidence. Unlike Levinson, in controversial cases such as this, Trivedi is not compelled to make an absolute verdict about work-meaning.

A related objection states the opposite: appeal to authorial intent in many cases is necessary because lack of such knowledge may falsely entail superficial interpretations. A notable example is the case of Andy Kaufman (Carroll, 2002). Kaufman once engaged in insane bouts with women and a professional wrestler. His performance was perceived by many as pathological. However, recent evidence from coteries shows that his intention was to satirize professional wrestling. Only when we have knowledge of Kaufman's intent can we truly appreciate the depth of his performance.

Kaufman's example is coupled with another objection: hypothetical intentionalists never justify their aesthetic criterion for deciding "best." In some cases the aesthetic ground could be weak due to moral and intellectual concerns (ibid., pp. 342-343). In Kaufman's case, suppose the evidences for the satire reading and for the madness one are equally compelling. Hypothetical intentionalism would entail that the former is the correct interpretation because it is aesthetically superior. But this result seems morally and intellectually inadequate.

Two responses are available (Trivedi, 2001, pp. 203-204). First, moderate hypothetical intentionalism acknowledges the intention to satirize. Therefore, Kaufman's case is not a counterexample at all. Second, a competent audience can discern the author's intention to satirize; if Kaufman was really mad, this can also be known by a competent audience.

Aesthetic consideration comes into play only when there is a tie between epistemically best readings. It appears that the second response does not really meet the objection, for Carroll can press on the possibility in which the interpreter reaches two epistemically best

---

[65] Many counterexamples offered in Carroll 2011 can be handled this way. I am not going to discuss them one by one.

interpretations and one makes the work aesthetically better but morally inadequate. However, this is controversial because some people do not think aesthetic and moral considerations are independent. Many forms of aesthetic ethicism attempt to connect them. In some cases, at least, the moral failing would be an aesthetic failing because it blocks the kind of audience uptake the author was trying to achieve. Therefore, the present challenge might not hold.

Finally, Trivedi claims that the moderate hypothetical intentionalist is also able to meet Stecker's two counterexamples. The first is the case where *W* means *p*, but *p* is not intended and the audience is justified in believing so. Hypothetical intentionalism implies wrongly that *W* does not mean *p*. This is no counterexample to moderate hypothetical intentionalism, which hypothesizes work-meaning first before attributing it to the author. Therefore, in this case Trivedi's competent audience would take *p* as the work-meaning and attribute it to the author as what she should have intended. So, back to the Watson example, we should not say that the work conveys an impossibility regarding Watson's war wound, because this interpretation, from the perspective of the competent audience, is not what Conan Doyle should have intended.[66]

The second counterexample can be handled in a similar manner. Stecker claims that in the case where the author is known to intend *W* to mean *p* but ends up saying *q*, hypothetical intentionalism wrongly implies that *W* means *p*. However, since the competent audience hypothesizes work-meaning first, they will claim that *q* is the correct work-meaning.

5.4 Chapter conclusion

My biggest concern with hypothetical intentionalism is that it suspiciously collapses into another version of it. All versions of hypothetical intentionalism discussed in this chapter center on the intention of the actual author. A different species of hypothetical intentionalism could involve a hypothetical author. One wonders why Tolhurst and his followers are not referring to a different author given that this author is only the public persona of the actual

---

[66] Trivedi (2015, p. 718) invokes the principle of charity to solve this problem: "logical and interpretive charity to Conan Doyle suggest that he could not really have intended such an inconsistency […]" This sounds like a defence for Levinson, not for himself, because this quotation is about hypothesizing the author's intention, not about work-meaning. It is hard to see the connection between Trivedi's solution to the Watson counterexample and his solution to Stecker's first counterexample. One would expect him to deal with the latter in the same way he deals with the former. However, that is not the case. I believe my treatment here is what Trivedi *should have intended*.

author. In his original formulation, Levinson actually equivocates between the actual author and the hypothesized author at several places, which betrays the said conflation (for example, Levinson, 1992, p. 227, 230, 238, 251).[67] This equivocation might be what has led his critics to interpret him as targeting a hypothetical author (Davies, 2007, pp. 183-186; Livingston, 2010, p. 409).

Indeed, it is odd to say that one is still concerned with the actual author when this author is totally made up by public evidence excluding proclamations of semantic intention. Why not just say that in this case one is constructing an author like the actual one but different from him, rather that that one is discovering the aspects of the actual author relevant to the interpretation of his work? This line of thinking brings us to the second species of hypothetical intentionalism, which is the topic of next chapter.

---

[67] The fact that Levinson is suspiciously talking about a hypothesized author is pointed out by Sherri Irvin (2006, p. 126).

6. Hypothetical Authorism and the Implied Author

"The 'implied author' chooses, consciously or unconsciously, what we read; we infer him as an ideal, literary, created version of the real man; he is the sum of his own choices."

— Wayne C. Booth

The position of literary interpretation that I dub *hypothetical authorism* is often labelled a variety of hypothetical intentionalism (Carroll, 2000, p. 85; Trivedi, 2001, pp. 201-202; Irvin, 2006, p. 122; Davies, 2007, p. 168; Stecker, 2010, p. 158); however, it is sometimes viewed as a standalone version of intentionalism called *fictionalist intentionalism* (Livingston, 2010, p. 405). Although I am happy to accept that hypothetical authorism is a variety of hypothetical intentionalism, I will preserve the label "hypothetical intentionalism" for Tolhurst's (and his followers') version of intentionalism, for if a different name is not used it is often unclear which version of hypothetical intentionalism one is talking about.

The idea of an imagined author can be traced to Wayne C. Booth's account of the *implied author* (Booth, 1983), referred to and renamed by different hypothetical authorists as the *apparent artist* (Walton, 1976, 1990, 2008), *postulated author* (Nehamas, 1981, 1986, 2002), *ideal author* (Nathan, 1982, 1992, 2006), or *fictional author* (Currie, 1990).[68]

Particularly, the theoretical construct of apparent artist is mainly used to support an account of artistic style rather than of interpretation. But as its advocates admit, the account is applicable to art interpretation (Robinson, 1985). For this reason, in my discussion I will treat it as more concerned with interpretation than with style.

Generally speaking, hypothetical authorism maintains that interpretation is grounded on the intention suitably attributed by the interpreter to a hypothetical or imagined author. Two inquiries hence arise: (1) what good reason do we have to endorse any sort of interpretative authorism? (2) Why does an imagined author matter more than a real one if we do need authorism at all? We shall see the answers given by different proponents to these crucial questions in the following discussion.

---

[68] These accounts are listed chronologically by their first appearance. I will follow this order when discussing them.

Below I start with what initiates hypothetical authorism: Booth's account of the implied author.

6.1 The implied author

Booth's account can be seen as a middle course between anti-intentionalism and extreme intentionalism.[69] On the one hand, he denies that a work infallibly reflects the author's values, attitudes, intentions and other elements in her mental stock; on the other, he wants to preserve talk of the author in criticism for ethical and moral appraisals. The implied author account serves both needs. But what is an implied author exactly?

Perhaps it is easier to start with what an implied author is not. The implied author is not the *narrator* or dramatic speaker in the story. The narrator is a fictional character in the story, but seldom is this fictional character identical with the implied image of the author that lurks behind the text. Also, the implied author cannot be fully grasped by terms such as "style," "tone," or "technique." These concepts describe only selective aspects of a literary text, while the implied author is the sum of all features of the text.

Booth defines the implied author in several ways: a version of the author that the actual author creates when writing, the author's "second self," the "picture the reader gets," and the "core of norms and choices" (Booth, 1983, p. 70, 71, 74). There seem to be conflicting definitions here: the implied author is both a self-image created by the author and an author-image constructed by the reader. For the self-image of the author, Booth says:

> As he writes, he creates not simply an ideal, impersonal "man in general" but an implied version of "himself" that is different from the implied authors we meet in other men's works. To some novelists it has seemed, indeed, that they were discovering or creating themselves as they wrote. (Ibid., 1983, pp. 70-71)

But on the same page (p. 71) Booth also writes:

> However impersonal he may try to be, his reader will inevitably construct a picture of the official scribe who writes in this manner—and of course that official scribe will never be

---

[69] For Booth's view on the implied author, see especially pp.70-77 in his *The Rhetoric of Fiction*. For a nice introduction to Booth's view, see Kindt & Müller, 2006, pp. 42-61.

neutral toward all values. Our reactions to his various commitments, secret or overt, will help to determine our response to the work.

The inconsistency in the two paragraphs might be resolved by claiming that the two central claims here are actually not contraries. That all the textual features are created (in this case, caused) by the author does not entail that she has consciously or successfully turned a selective set of them into a coherent image of an implied author, which is to be constructed by the reader.

Booth offers two reasons explaining why the notion of an implied author is useful. First, if we make a distinction between the author and her implied version—which are two different things—we can avoid arguing over whether an author should be blamed simply because in real life she does not hold the same attitudes, positions, or values as expressed in her work. Not only is the need for this sort of blame in literary criticism questionable but also it deprives the author of the right of concealing her personal values and thoughts in her fiction writing.

Second, the implied author account finds a middle place between two opposing views about the author's presence in a work. At one end of the spectrum lies the claim that the author can always conceal her voice or idiosyncrasies in her narration; she should just show what happens rather than tell the reader her judgments and comments. At the other end lies the view that a work is always transparent to the author's idiosyncratic features of her mental life. Neither of these extreme views conveys the true picture. What really happens is that the author can never erase herself from the work, but what we can find out about her is not necessarily true of her, either.

Booth's theory has had a tremendous impact on literary theory and the philosophy of literature. The interpretative accounts to be discussed below all share a concept akin to that of his implied author: they have the same core in different disguises.

6.2 The apparent artist

According to Kendall L. Walton, the critic should be interested in the *appearance* of a work when considering its features. Particularly, the style of a work is understood as how it appears to have been made, which involves appeal to the agent of an action. In a related fashion, the appearance of intentionality is construed in the same way. If it appears to us that the author of a literary work intends to criticize totalitarianism, then the critic is entitled to

claim that the (apparent) author intends the work to be an attack on totalitarianism, even if the actual author did not have that intention in mind. Broadly speaking, if a work appears to have been produced with certain choices, decisions or plans (of or for characters or plots, etc.), the critic is free to consider this appearance of intention, regardless of the intention of the actual author. Specifically, appearance rather than reality should be the focus of criticism. In this case, the author we appeal to is only apparent rather than real. Walton calls this apparent genesis "apparent artist" or "apparent author."[70]

The apparent artist is fictional in nature. However, she is fictional not in the sense that she is a character in the story world but in the sense that she belongs to a fictional world different from that to which the story characters or objects belong. In the case of literature, she may be called the *storytelling narrator*, to be distinguished from the narrator who is a fictional character in the story.

Walton suggests that appeal to the apparent artist's intention is not an intentional fallacy at all, because the verdict of criticism in this case is based mostly on the work itself. Nonetheless, he also points out that not all extratextual information is irrelevant. Particularly important is knowledge of the work's historical context. Why this matters is because our beliefs would affect our perceptual experience. And some perceptions of a work are arguably incorrect or inappropriate, absent the said information. The point here is that appearance has to depend on reality, or on a contextual basis specifically.

One may ask further what this reality is, or, what sorts of contextual information are relevant. Before answering that question, a few words should be said about Walton's central thesis that our beliefs may change our experience of artworks. Walton convincingly suggests that how a work appears to us depends upon what we know about it. In other words, how a work appears is a function of both itself and context. To illustrate this point by an example given by Walton, consider the drawing *Iceskaters* by Gary Gilmore, who is a notorious American murderer. The drawing seems to be a peaceful winter scene until the viewer realizes that it was drawn by a sinister criminal. With this knowledge in mind, the drawing now appears to present a vicious and insidious mind lurking behind the serene. To give yet another example from Walton: suppose we have a painting in which the raindrops appear to have been randomly and carelessly dripped by the artist. Our perception of the painting would change after we learn that the raindrops were actually eyedropped and that the artist made great efforts to make the drops resemble those in another painting by her. This

---

[70] Walton uses the former term more often because his discussion is not limited to the literary arts.

knowledge generates in us a different viewing experience, perhaps a sense of meticulousness and order. These examples show that the audience's perception of works depends upon their beliefs about the work.

But we still need to sort out what beliefs are relevant. One immediate thought would be that authorial intent is conspicuously among the contextual factors of a work's production. But apparently knowledge of intent should not be considered if our account is to be consistent with the hypothetical authorism delineated above. This is not entirely correct.

As discussed in the previous chapter, there is an equivocation with the term "authorial intent," which, in the debate on interpretation, typically refers to semantic intention, that is, the intention an author has to mean something by a text. What Walton dismisses is semantic intention; however, he accepts categorial intention. An author's categorial intention is the intention to classify the work at a fundamental level (typically, as belonging to an artform and genre). A work misperceived in different categories may appear to exhibit different aesthetic properties. Since it is obviously wrong to perceive a work in certain categories, determining the correct category for a work will be crucial if we want to grasp the right properties it has. For example, if we perceive a bust as a statue then the bust would probably appear weird and shocking because the human figure ends abruptly at the chest. And if we read an American hard-boiled detective story as an old-fashioned puzzle mystery, the story would just turn out to be ill-constructed in plot and boring in solution. Walton argues that the only sufficient means to determine the correct category of a work is the actual author's categorial intention.[71] Therefore, categorial intention will be one part of the reality on which how a work appears to us depends.

A second category of contextual factors is common factual knowledge in the time of the work's production. For instance, even if a work produced in the seventeenth century appears to us that the author has William James's pragmatism in mind, we should not attribute this intention to the apparent artist. This is because it is common knowledge that pragmatism came into shape only in the twentieth century. Similarly, it would be inappropriate to attribute the intention of criticizing industrial societies to the apparent author even if the work in question appears so when we know that the actual author's life predated the industrial age. Walton puts this point as follows:

---

[71] For criticism and reply in this respect, see Nathan 1973 and Walton 1973 respectively. For a critical commentary on Walton, see Laetz 2010.

A great many facts about the origins of many works of art are common knowledge, especially facts about the societies in which they originated [...] So we have a great deal of common knowledge concerning what interests and attitudes are at least likely to have motivated the creation of many works, the intentions many artists could or could not reasonably be expected to have had, and so on. (Walton, 2008, pp. 246-247)

In conclusion, the motivation behind Walton's account of the apparent artist is that for art criticism and appreciation how a work appears to us is important for its own sake and should be the main concern of critics. Nevertheless, this appearance is grounded on reality, for there are apparently inappropriate perceptions of works, notably stemming from the failure in identifying the work as of its historical author. Appearance has to rely on the beliefs about the contextual factors present in the work's creation such as common factual knowledge and the author's categorial intention.

6.3 The postulated author

Alexander Nehamas's theory of literary interpretation is based on the concept of a postulated author. Briefly, the theory states that the author is construed as a figure or character the interpreter postulates to explain the greatest number of features of the text, constrained by certain contextual factors present at the time of creation.

According to Nehamas, interpretation is mischaracterized by early debaters as the attempt to search for the implicit meaning hidden behind the surface of a text. Interpretation is in truth the activity of construing movements and objects as actions and their products. For example, to interpret the movement of an arm is to construe it as, say, a greeting. In this case, what the interpreter does is not dig out a hidden meaning behind an apparent one; rather, she takes the movement to be susceptible to a certain kind of inquiry and offers an answer to it. In other words, she gives an *explanation* of the movement in question. Since actions require agents, the explaining must be done by appealing to the agent, or specifically, to intention and rationality. Likewise, to interpret a text is to construe it as the product of an action, that is, the product of writing.[72] In doing so, one subjects the text to certain inquires and offers answers

---

[72] When talking about construing a text as something, Nehamas equivocates between "an action" and "the product of an action." I find the latter more intelligible, as a text is more suitably seen as an object rather than a movement.

to them. Since the text is the product of an action, explaining it again requires the appeal to its agent. If a text can be subject to literary interpretation, it is suitably seen as a *work*.

Two points need clarifying. First, some may think that it is constrained to say that construing the movement of an arm as a greeting deserves the term "interpretation," because that construal happens almost automatically. This is true, and also true in literature. For instance, much popular fiction does not call for interpretation; the reader just gets it plainly. In this kind of case, comprehension is based on our shared conventions as background assumptions and is hardly interpretative in the strict sense. Perhaps we should use the term "interpretation" only in the strict sense. But it is also reasonable to maintain that in the above case there is still interpretation of a minimal degree, better called "understanding."

The second point is about whether we should have a general account of workhood that specifies the features responsible for making a text interpretable and hence a work. However, it seems that a theory of interpretation can still fare well without such an account. We interpret movements without difficulty despite lacking an account that explains which movements constitute actions.

Now, if we agree with what Nehamas has said so far, it seems plausible to accept that appeal to an agent is indispensable. The question becomes why the agent appealed to has to be hypothetical rather than real.

Nehamas makes an important distinction between *writer* and *author*. The writer is the actual individual firmly located in history that caused the text. Put in Aristotelian terms, the writer is the efficient cause of the text; but she is only contingently related to her text because there is the possibility that she never wrote it. The author, on the other hand, is the formal cause of the text, necessarily related to it. She is whoever arises from or manifests in the text, whoever the interpreter understands to have produced the text when interpreting it.

To give an example: there is a distinction between the Shakespeare who actually wrote *Macbeth* and the one whom we understand as having written it when engaging the text (call the second "Shakespeare*"). Shakespeare is not related to *Macbeth* in the same way as Shakespeare*, for it is not necessary that Shakespeare wrote *Macbeth*, that he is the "author" of it; someone else might have been. But there is a second sense of "author." The Shakespeare we have in mind when engaging *Macbeth* emerges from the text as soon as we set to explaining the text with the appeal to an agent. Shakespeare* is essentially connected to the text because he is whoever we make out the author to be on a textual basis; he is an effect of the text, so to speak. Construed this way, the author, unlike the writer, is not the creator of the text; rather, she is the product of the text.

Appeal to the writer is often, though not always, associated with the more general claim that interpretation aims to recapture meaning once manifest in the past. However, this view is problematic due to two concerns.

First, it is doubtful that we can recapture the work's meaning once and for all. A literary text is the product of a complicated action, and its meaning may only be gradually revealed to us as more information and developments in the world relevant to interpretation become available. We can never be confident that we have already stood on the vantage point to interpret a text; the interpretation we think is best at this moment may turn out to be incomplete as time goes by. Given this inherent corrigibility, thinking of interpretation as "capture" rather than "recapture" work-meaning may be more reasonable.

Second, the idea of recapturing could suggest recapturing a state of mind in the past, namely, the writer's intention. This creates the problem of how we can see things through a historical person's eyes; that is, how we can recapture a mental state that has already gone. For this reason, retrieving the writer's intention would be an unachievable task. A second corollary that will follow from the impossibility of replicating anyone's past consciousness is that it is impossible too to re-create the original audience's original understanding of the text. This rejects hypothetical intentionalism, which claims that the correct interpretation of a work is the best hypothesis of the actual author's intention made by the intended or ideal audience in the time of the work's production.[73]

If the writer is not the subject of criticism, then it has to be the author. Although, according to Nehamas, the author is the product of the text, it is so just to a great but not full extent. That is, the author is not completely constituted by properties of the text. A distinction needs to be drawn between the author and a fictional character in the text. A story character is totally immanent in the text: she is just what the text says she is. However, the relation between the author and the text is not exactly the same as that between a fictional character and the text. It is not an immanent relation. The author, unlike the fictional character, is not depicted or described in the text; rather, she is manifested or exemplified in it. Put another way, the author is the "representer of the representation." The relation here is described by Nehamas as *transcendental*, since the author is neither simply part of the text nor something existing straightforwardly outside it.

Nehamas elaborates how this transcendental relation develops in the process of interpretation. As said, a text becomes a work when subject to interpretative inquiries. When

---

[73] This might not be a challenge to Trivedi's competent but *fallible* audience.

interpreting, we inevitably appeal to the intention and rationality of an agent—the author, who, to be distinguished from the writer, arises from and is necessarily connected to the text. The author is *postulated* by the interpreter to explain the text: she is not only the product of the text, but also the product of interpretation, produced from and constructed in the interaction between the interpreter and the text.

One very important point to bear in mind about Nehamas's account is that the author is not only transcendental to a single text, but also to her *oeuvre*. This is an apt move to make, for interpreting a literary work typically requires knowledge of the works by the same writer. Appeal to the whole oeuvre offers clues that help produce an interpretation that explains the text under scrutiny more deeply and comprehensively. The more we learn about a writer's oeuvre, the better we explain her work, and our resultant interpretation may even change our reading of her oeuvre, which again may affect our interpretation of the work in question. This continual adjustment can be and should be extended beyond the writer's oeuvre. When interpreting a literary work, referring to the writer's oeuvre is usually not enough. For better author construction, we will need to look into the work's literary influences, which will again have their authors under construction; and for those authors' constructions, still more literary influences will need to be considered, and so on and so forth. Each new interpretation thus produced modifies our interpretations of certain works, and those interpretations give us new insight for producing a better interpretation of the initial work.

The kind of interpretation we are ultimately after is the one that coherently accounts for the greatest number of features of the text. Meaning is thus identified with this ideal interpretation. But the critical monism advocated here is only a regulative ideal, since the process of continual adjustment mentioned above may never end. What we actually have is a revisable interpretation, a mutable author-figure, or a hypothesis accepted provisionally.

The consideration of oeuvre and the influences from other works suggest that the author to be constructed is not arbitrary and has to be constrained by certain contextual factors. Specifically, the methodological constraint here is that the author has to be a historical variant of the writer: she cannot mean what the writer could not have meant. In other words, ahistorical readings of works should be excluded.

Claiming that we can give an ahistorical reading of a text implies interpretive anachronism.[74] But if we take seriously the relations a text bears to other texts (oeuvre and other literary influences and connections), anachronism would require us to revise all these relations. It is doubtful that such revisionism can be, and has ever been, achieved. Such being the case, we should make interpretation consistent with the contextual factors present in the time of production. Specifically, interpretation should be consistent with the linguistic convention, the writer, original audience, genre, possibilities of writing, and facts about the world.

To conclude, my reconstruction of Nehamas's argument of postulated author is as follows:

(P1) The view that the author can be arbitrary (ahistorically situated) implies interpretative anachronism.

(P2) A text bears different relations to an indefinite number of different texts.

(P3) Interpretative anachronism requires consistently revising all these relations.

(P4) Such revisionism is not likely to be achieved.

(P5) Interpretative anachronism is hence not workable.

(C1) Therefore, the author cannot be arbitrary (or not ahistorically situated).

(P6) If the author is historically situated we can base interpretation either on the actual author (the writer) or his variant.

(P7) Basing interpretation on the actual author ends up making the subject of interpretation the writer.

(P8) The subject of interpretation is the author rather than the writer.

(C2) Interpretation should aim at the writer's historical variant.

6.4 The ideal author

Daniel O. Nathan develops another version of hypothetical authorism, which is a sort of combination of Monroe Beardsley's conventionalism and Nehamas's postulated author.

Nathan reaffirms Beardsley's central position that extratextual information about authorial intent is dispensable. He further claims that Beardsley's conception of linguistic conventions

---

[74] Nehamas agrees that anachronism is acceptable in the case where the writer could have anticipated views that are made clear only at later times. A similar point is made by Beardsley (1981a, pp. 246-247), Davies (2007, p. 157), and (Carroll, 2009, p. 79). Here I will use the term in the sense that excludes this kind of case.

has been misconceived. Many alleged counterexamples stem from these misconceptions of conventionalism. Conventionalism has been mischaracterized by Beardsley's critics as narrow and impoverished. To be fair, Beardsley makes it very clear that the concept of linguistic convention should be richly and broadly construed in two respects (Beardsley, 1981a, pp. 122-126; 1970, pp. 25-26). These points are reconsidered in Nathan's discussion.

In the first place, *complete* rather than partial textual context must come into play. The fact that intentionalists sometimes ignore this leads to unsound counterarguments. For example, actual intentionalists such as Hirsch would claim that whether the sentence "You're such a genius" is ironic or not can be pinned down only by the speaker's intention. However, this is arguably an example of partial discourse. In daily conversation seldom do we utter this minitext out of the blue without saying more preceding or following it. And in this case it is reasonable to include the larger textual context as the proper object of interpretation. Given the complete textual context of which the minitext is part, the initial ambiguity is then duly resolved.

A different sort of counterexample suggests the need of contextual indicators on the occasion of utterance. To know that *A Modest Proposal* is actually ironic, we need to know the political atmosphere and moral values of the public at the time of the text's production. Otherwise textual evidence in this case will lead us to the infanticide reading, arguably not to the correct interpretation of the text. But again, as Nathan argues, this is not a counterexample at all because there are salient textual clues indicating irony. Intentionalists simply fail to consider the complete text produced by Swift to discern the ironic content. It follows that many counterexamples can be obviated this way. Here the objector might say that this move does not remove all counterpleas: there are many cases in which textual meaning diverges from work-meaning, and the latter can be obtained only by appeal to context. This brings us to the second restatement Nathan makes for conventionalism.

According to Nathan (and perhaps Beardsley), linguistic convention includes all the connotative richness of the words used at the time of creation and vocal intonation in the case of oral utterances. Therefore, the meaning of a text is not to be taken at face value; what it means is not purely the literal combination of word meanings, but what the words connote and suggest. The objection that linguistic convention is not rich enough to capture the implicit meaning of a text is just a misunderstanding and naïve construal of conventionalism.

The second restatement needs to be considered with another point made by Nathan. Aside from appealing to linguistic convention, Nathan advances two categories of background assumptions on the interpreter's part that he thinks suitably go with linguistic interpretation.

First off, there are nonlinguistic assumptions that all language users share. These are common facts and knowledge about our world. An example used by Nathan is the statement "We sell alligator shoes." This can mean either "we sell shoes made from the skins of alligators" or "we sell shoes worn by alligators." The second makes less sense because it is a common background assumption that people do not sell shoes worn by alligators. Were it not for this sort of factual assumption, ordinary conversation cannot even get going, for we can always understand an utterance as having an unconventional meaning. If she intends the said statement to mean the unusual, the utterer will need to provide more textual clues to defeat the factual assumption in question. Otherwise the utterance should be interpreted on the basis of shared factual assumptions.

The second sort of background assumption is metalinguistic. It is the assumption about speaker competency. Given the above requirement, it seems that we need to posit a competent speaker such that she is aware that her utterance is to be understood by public conventions rather than private mental acts. That is, we assume the speaker is competent or ideal in the sense that she uses conventions flawlessly. This is because convention is based on a long history of successful use, which gives us footing for so interpreting. For literary utterances, positing such a competent speaker amounts to imagining a hypothetical author adhering to the perfect use of convention. It follows that all features in a work are thus taken as *purposeful* and that the intention of the actual author does not trump the intention of the hypothetical one.

With the help of the above resources, if the work still remains ambiguous, then we must conclude that the meaning is ambiguous (Similarly, in conversation we may not treat an ambiguous utterance as meaning what the speaker intends it to mean; we may instead say that the utterance's meaning is ambiguous.) This again echoes Beardsley's view on ambiguous utterances.

Nathan suggests several reasons motivating the ideal author approach. Firstly, interpretation is to explain everything in the text, to see everything there as purposeful, instead of to track what was at work in the time of creation. In other words, the interpreter should aim for the *maximally explanatory* interpretation (or at least one that tries to be so), rather than the most probable one. It is not asking for the efficient cause of the text, but for the final cause.

Secondly, it makes more sense to embrace an interpretative account that allows richer interpretations than one that produces readings only confined to authorial intent. If a work can well sustain a much more interesting interpretation than the one intended by the author,

why should we dismiss it as illegitimate? Strict adherence to authorial intent in many cases forbids brilliant critics to produce aesthetically rewarding interpretations. After all, for any sophisticated work, there are bound to be interesting features the author is not aware of when creating. Actual intentionalism is not able to accommodate these marvellous accidents in art.

Thirdly, the present account shows respect to the author's intention that the audience is not to consult authorial intention. In Gricean terms, when *S* (a speaker) means p by *U* (an utterance) there is a second-order intention that *H* (the hearer) need not go beyond *U* to reach p; namely, there is no need to consult *S*'s first-order intentions to understand *U*. Similarly, when an artist creates a work for public consumption, there is a second-order intention that her first-order intentions not be consulted; otherwise it would indicate the failure of the artist. Nathan calls this the *paradox of intention*.

Finally, the ideal author view is intersubjectively based, in contrast to the subjectivist approaches such as the author-centered theory and the reader-centered theory. Both theories are subjectivist approaches in the sense that they define the correct interpretation in virtue of the mental state of some actual individual. However, this kind of position results in subjective incorrigibility.

The anti-intentionalist might question whether we still need the notion of an author if we already have a robust and sophisticated conventionalist position. Here are two additional reasons for preserving talk of the author in a theory of interpretation: without this notion works simply cannot be differentiated from natural objects; also, if to interpret is to explain, we would need an agent acting as the explanans.

The motivation of the ideal author account (and also of the postulated author account) is nicely summarized by Nathan as follows:

> The created world, literary or otherwise, will necessarily have features that its author may not have foreseen or intended […] all these unintended features no less than the intended ones become fair game for interpretation and criticism. The best interpretation then turns out to be the one that can make the best sense of the greatest number of features available in the work. The art-interpretive process appears to require no less than that. (Nathan, 2006, p. 293)

6.5 The fictional author

Gregory Currie's version of hypothetical authorism is developed to interpret fictional works using the concept of fictional author, and is an application to his account of fictional truth. For our purpose here I will concentrate on what he says about interpretation.

Since what interests Currie in the end is fictional truth, it is natural for him to limit his theory of interpretation to what he calls *story meaning*: the propositions true in the story.[75] Are there ghosts in *The Turn of the Screw*? Is the governess in the said story mad? Does Holmes's best friend Watson have his battle wound in the arm or thigh? These questions are not the same as those asking what are the themes and theses in the story, or whether there is any figurative language at play, or what are the motivations and aims for character or plot design. Apparently, story meaning is the most fundamental level of "meaning"; questions at deeper levels of meaning such as those I just suggested cannot be resolved unless we have answers to questions at the basic level. For example, if we cannot answer whether there are ghosts or not in *The Turn of the Screw*, we cannot be certain whether the work can be seen as a psychological study or not. Story meaning constitutes the missing part of the story world, from which we are able to pursue deeper interpretative inquiries.

With these preliminaries in mind, now let us consider Currie's main proposal: what is true in the story is what the fictional author believes. Two points call for elaboration. First, why does the author have to be fictional? Second, why do we explain story meaning in terms of belief? The answer to the second question will reveal that to the first.

Interpreting the fictional author is like interpreting real people: we try to build up her belief set, to make the best overall sense we can of her behaviour. Literary interpretation aims to *explain* the text, as much as belief attribution aims to explain behaviour. The point is that "It is true in the story that" functions semantically like "It is believed that." To highlight the similarities: one can believe something without believing its consequences (something can be true in the story without its consequences being true in the story); one can believe neither a proposition nor its negation (whether something is true or not in the story can be left undecided); one can believe something that is contradictory (it can be true in the story that something contradictory happens); finally, one can believe contradictory things (it can be true in the story that P & ~P).

To build a belief system it won't do just listing what the text literally asserts, because people oftentimes say things nonliterally. One needs to consider the complete textual context to ascertain what is really asserted. Also, there are implicit beliefs of the speaker that we need

[75] Currie's story meaning is similar to what Beardsley calls "elucidation," which is discussed in Chapter 2.

to infer. We do so against some background assumptions. To borrow an example from Borges, we infer many of Pierre Menard's beliefs against the assumptions about what people in Menard's community tend to believe. This qualification of the interpreter makes her an *informed reader*, a reader familiar with the relevant facts about what people in the author's community tend to believe when her work was issued.

It is now clear that the set of propositions true in the story is a function of the text plus relevant background assumptions. In other words, a proposition in a story is true if and only if it is reasonable for the informed reader to infer that the fictional author believes that proposition. The author in question is fictional because her belief system is constructed rather than discovered by the informed reader. No doubt what the fictional author believes in most cases will coincide with what the real author believes, but this is not always the case. Even when it is, we are not to conclude that the two are identical, for they are different entities: one is the product of interpreting, the other the cause of the text.

It would be wrong to think that the intention of the real author plays no role in Currie's account. Currie explicitly accepts the author's categorial intention, which would be one sort of background knowledge the informed reader needs to possess as she sets to work. Specifically, genre perception is at least partly determined by authorial intent and partly by standard features of that genre found in the story. The latter are not sufficient because total reliance on them might end up misconstruing the work. *Macbeth*, to give an example, has all the standard features a murder mystery has, but few critics would see it akin to *Murder on the Orient Express* or *The Hound of the Baskervilles*.

The author's *lexically specific* intentions also play a crucial role in establishing her text. This is the intention about how words are arranged, used and spelled in the text. However, the text should be corrected in accordance with linguistic convention when we do not detect the author's intention to use the words in a nonstandard manner. In the case where the author has false beliefs about how words are maneuvered, the text is also subject to correction, for in this case the author's intention to use them correctly dominates.

The fictional author account does not exclude the possibility of non-equivalent but maximally good interpretations. This consequence may lead to interpretative antirealism and relativism. Currie admits the former but rejects the latter. If we hold an antirealist position with story meaning, we would claim what is true in a story is the result reached by our best interpretative method. There is no possibility of a best supported but incorrect interpretation. In other words, there is no gap between truth and evidence. This is where the analogy

between interpreting fictional authors and real people breaks down. In the latter case there seems to be a determinate mental reality that transcends interpretative evidence.

Currie suggests that we adopt *relationalism* with incompatible interpretations: the relational statement "*P* is true in the story relative to interpretation *I*" is true if and only if *P* is one of the set of propositions *I* and it is justified that the fictional author believes *I*. That is, what is true in the story is always a relational statement of the form F(*P*, *I*). But relationalism is not relativism, which is understood either as the extreme view that the correct interpretation is relative to the reader or the less extreme that it is relative to the interpretive community. Nonetheless, relationalism denies that the correct interpretation is decided by anyone's or group's preference, and that any given proposition must be fictionally true relative to some interpretation.

However, Currie (1991b, pp. 106-108) admits that relationalism is problematic: the definition of I in F(P, I) is circular because it can only be defined as (F*(P1, I) & F*(P2, I) &…F*(Pn, I)). He modifies the account by claiming that "P is true in the story" is assertible to the degree that it is probable that it belongs to a maximal interpretation which the speaker adopts.

To summarize Currie's theory in brief: story meaning is partly based on conventional meaning (C-meaning, i.e., literal meaning of the text); however, if interpretation is seen as belief attribution, the concept of intentional meaning (I-meaning) must come into play. But the agent of action the interpreter appeals to in this case is constructed and hence fictional. As long as we abandon the realist position with interpretation, conformity with evidence, rather than the author's mental reality, defines truth. That implies critical pluralism: there can be equally plausible interpretations of the same work.

6.6 Stecker's objections to hypothetical authorism

The most severe criticisms on hypothetical authorism come from Stecker (1987) and Livingston (1998, pp. 833-835; 2005a, pp. 165-172). I critically discuss their objections in this and the next section respectively.

Stecker's first objection states that for hypothetical authorism, consideration of historical context does not adequately support the alleged interest in the work's appearance.[76] As

---

[76] In Stecker's paper these objections are raised against different authors, but I believe his main points apply to hypothetical authorism in general. Therefore, in my discussion I will treat them that way.

discussed, hypothetical authorism rejects interpretative anachronism for the reason that knowledge of historical context does not support the claim that, to use Borges's example again, Cervantes has pragmatism in mind. But when historical context plus text make it likely that Menard has pragmatism in mind in his *Quixote*, why does the hypothetical intentionalist in this case recoil from saying that historical context supports the claim that the real author, rather than the implied author, has pragmatism in mind?[77] It seems that historical context does not function in the same manner in the two cases.

A first quibble about this objection is that the function of historical context in the second case is affected by the addition of the text, but this is not so in the first case. So Stecker's reasoning by parity here does not work as an effective objection. Let us examine the matter more closely.

I think that hypothetical authorists may disagree about the extent to which historical context provides evidence for authorial intent; but one thing they agree on is that the evidentiary support in question is not very strong. If the text, as the product of intentional action, is not even good evidence for authorial intent—which the hypothetical authorist accepts and the moderate intentionalist rejects—then historical context would be more remote from decent evidence, for it is not even causally connected to the author as the text is.

But why is the text not good evidence for authorial intent? Here it is useful to appeal to Beardsley's account of evidence (1981a, pp. 18-21): if two things are distinct entities, evidence of the nature of one cannot be that for the other; however, when the two are causally connected, direct evidence for one can be *indirect* evidence for the other. But indirect evidence is less reliable and must be tested. Therefore, since the author's intention causes the text, direct evidence for what a text says will be indirect evidence for what the author intended, and vice versa. Intuitively, direct evidence for textual meaning will be the text itself; and direct evidence for authorial intention will be the author's professions of intention. Since historical context is not even indirect evidence for authorial intent, it at best functions as a defeater rather than determiner for authorial intent, as in Borges's case.

Stecker persists with the objection: if we really care about historical context, why exclude the historical author? This is identified by Nathan (2005, p. 38) as the question of *drawing lines*: it is difficult for those who are against authorial intent to draw a non-arbitrary line between extratextual evidence and evidence of authorial intent. Nathan's solution for this is

---

[77] Unless otherwise specified, I will use the term "implied author" to cover all different names for the current position.

that the second-order intention in the paradox of intention is a necessary presupposition in interpretation; in writing a public text, one basic purpose is that it be readable to people who know nothing about the artist's life. This presupposition constitutes the artistic convention that directly entails the independence of the work from authorial intent (ibid., pp. 292-293).

Stecker responds: not all artists have the second-order intention that Nathan suggests. If this premise is false then Nathan's argument becomes unsound. Even if it were true, the argument would still be invalid, because Nathan confuses the intention that the author intends to create something standing alone with the intention that her first-order intention need not be consulted. The paradox won't hold if this distinction is made (Stecker, 2010, pp. 153-154).[78]

I agree with Stecker that the public nature of works is compatible with the author's second-order intention that her first-order intention can be consulted, but compatibility does not imply that we must consult. I would claim that Stecker's counterargument amounts to a slippery slope fallacy, because the suggestion that we should consider a work's historical context does not compel us to consider authorial intent. It is unreasonable to say that if one allows one sort of extratextual information one must allow all sorts of extratextual information.

Perhaps a distinction between two different sorts of extratextual information should be made. One important criterion for distinguishing the author's intention from other contextually relevant factors is the causal relation to the text. The author's intention causes the text; if we reject the view that causality entails authority in fixing the content of a work—which is the very point of contention in the debate—we would dismiss evidence in this regard. Given this crucial difference in the nature of evidence, I do not see why one must accept evidence of authorial intent if she accepts evidence of other contextual factors.

Stecker's second objection to hypothetical authorism has it that we are mistaken about the subject of interpretation when appealing to the implied author. It is not clear why the

---

[78] Both Stecker and Davies (2007, p. 173) interpret Nathan as making the claim that all artists have the said second-order intention. However, in Nathan's paper on the publicity paradox there is a crucial typo in the following sentence on p. 40 which makes his claim hard to interpret: "the claim here is not *be merely* that, as a matter of empirical fact, most or even all artists […] have this second-order intention." (Emphasis mine). It is unclear whether the italicized part or only the word "be" should be dropped. In the latter case, the attack on the argument's soundness would not work because Nathan would accept that not all artists have the said intention. A consultation with Nathan in person revealed that he intended the former. Indeed, Nathan (2006, p. 293) does claim that "in producing an object for public consumption […] I necessarily intend to create something that will stand otherwise independent of my intentions." This indicates that this second-order intention is a conceptual necessity.

hypothetical authorist "prefers to think of hypotheses as constructing characters rather than imperfectly characterizing the writer" (Stecker, 1987, p. 266). Constructing a historical variant of the actual author sounds suspiciously like hypothesizing about her.

But there is still a difference. "Hypothesizing about the actual author," or more accurately, "hypothesizing the actual author's intention," would be a characterization of hypothetical intentionalism rather than hypothetical authorism. The latter does not track the actual author's intention but constructs a virtual one. This deflects the kind of attack launched by Carroll (2000, p. 83), that hypothetical intentionalism blindly sticks to hypotheses and ignores truth. Since hypothetical authorism does not track truth (in the sense of the actual author's intention), it is not constrained by the authorial announcement of intention, either made privately or publicly. In other words, knowledge of the author's semantic intention is in theory dispensable or irrelevant for the hypothetical authorist, but this is not the case for the moderate intentionalist. This is not only a difference in theory, but also one in practice. It may be true that hypothetical intentionalism also rejects semantic intention, but it does so at the cost of ignoring evidence that can prove the best hypothesis. Hypothetical authorism is not vulnerable to this criticism since it does not hypothesize the actual author's semantic intention.

Stecker's last criticism goes that appeal to the implied author does little work because in many cases it can be replaced either by the author or by the narrator. When the narrator of a story is not dramatized the implied author collapses into the narrator, because it would be superfluous to put another category of author between the implicit narrator and the real author. When the narrator is dramatized, it is again superfluous to postulate an implied author. This point can be elucidated by distinguishing between two cases.

First, when the dramatized narrator is reliable, it just seems reasonable to see her as the spokesperson for the implied author; the implied author thus again appears not very different from the narrator.

Second, when the dramatized narrator is unreliable, a transcendental perspective is called for. This is because in this case what we can learn from the explicit perspective will not be the real meat of the story. But if we are to find a perspective that transcends the narrator, why bother postulating an implied author instead of considering the real one?

I think all this deals no critical blow to hypothetical authorism. In the case of an implicit or undramatized narrator, the hypothetical authorist can claim that the implied author fills the role of the narrator. The two in this case merge into one, and to argue which merges into

which would be a verbal dispute.[79] In the case of an explicit narrator, reliable or not, the need for an implied author is reflected in one form of the intentional fallacy: the perspective the author endorsed when creating the work does not necessarily match that in the story. It is for this reason we need the implied author.

Stecker pursues the objection: the implied author's perspective is either adopted or just entertained by the author. As a matter of fact, there are cases where the implied author's perspective is in accordance with that of the author, and when this happens the implied author can be dropped; when they come apart, that just means that the author presents the perspective concerned to the reader for exploration, and the reader can draw on it for interpretation without attributing it to anyone. Stecker (1987, p. 269) writes: "The writer may be exploring it, trying to imagine what it would be like to actually hold it, but he does not hold it."

This sounds like Beardsley's pretence theory, which claims that one of the conditions making a work fictional is the author's intention to represent. In Beardsley's words, what Stecker says can be put as this: in writing fiction the author may pretend to perform an illocutionary act by holding a certain perspective. For example, in writing a war novel an author can pretend to hold an anti-war position and make the work read like an attack on military actions. But it is hard to imagine that an actual intentionalist like Stecker would endorse a pretence theory of fiction, which is typically adopted by the anti-intentionalist to prevent authorial intent from affecting the work's autonomy. It is more likely for the actual intentionalist to see fictional works as illocutions rather than their representations (for an example, see Carroll, 1992).

Stecker could reply that we need to separate two issues here: what is fiction and how to interpret it. One can hold a pretence theory of fiction and yet maintain actual intentionalism with respect to interpretation. In that case, the disagreement for the actual intentionalist would be on how to determine which illocutionary act is represented in the work.

This reply is still not satisfactory. To figure out what is represented we inevitably need the concept of I-meaning. It won't do for the reader to engage in interpretation without attributing the perspective presented in the work to anyone. If we choose not to attribute the perspective presented in the work to any speaker, we are not even talking about "perspective" at all, since a perspective requires somebody who takes it. As discussed in Chapter 2, even Beardsley relies on the concept of a fictional speaker, which refers either to the implicit

---

[79] Currie (1990, pp. 123-1266) seems to take this tack.

narrator or the dramatized one. On the one hand, if we attach a certain perspective to the implicit narrator, that equals attaching it to the implied author, the hypothetical authorist would say. On the other hand, it won't do if we attach it to the dramatized narrator, for what is represented might come apart from what the author intended to represent. In that case, the same objection works again at the semantic level against Stecker. It seems that the best way out is still the appeal to an implied author.

6.7 Livingston's objections to hypothetical authorism

Livingston criticizes hypothetical authorism for not being able to distinguish between different histories of creative process for the same textual appearance. For example, suppose a work that appears to be produced with a well-conceived scheme did result from that kind of scheme; suppose further that a second work that appears the same actually emerged from an uncontrolled process. Then if we follow the strictures of hypothetical authorism the criticisms we produce for these two works would turn out to be the same, for based on the same appearance the implied authors we construct in both cases would be identical. But apparently the two works have different creative histories and the difference in question seems too crucial to be ignored.

An initial reply is that the aim of art criticism is to focus on appearance. Appearance is important for its own sake, and for art criticism one should take this aesthetic point of view rather than the historical point of view (Walton, 2008, p. 248). This goes back to the debate on interpretative aims and the answer just depends on everyone's default assumption. But recall that the hypothetical authorist does not solely focus on the product of creation; she also pays attention to prior conceptions of its process where they are contextually relevant. Knowledge of the actual creative process may be crucial for our making out the appearance of a work. In that case, the work appearing well-constructed might turn out to be messy and uncontrolled in structure after we learn that the author produced it in a hasty and undeliberate manner. The objection here seems to fail to consider the subtlety of reality-dependent appearances. For example, suppose the exhibit note besides a painting tells us it was created when the painter got heavily drunk. Any well-organized feature in the work that appears to result from careful manipulation by the painter might now either look disordered or structured in an errie way depending on the feature's actual presentation. Compare this scenario to another where a (almost) visually indistinguishable counterpart is exhibited in the museum with the exhibit note revealing that the painter spent a long period crafting the work.

In this second case the audience's perception of the work is not very likely to be the same as that in the first case. This shows how the apparent artist account can still discriminate between (appearances of) different creative histories of the same artistic presentation.

But to revive the objection, Livingston can narrow the scope from criticism in general to interpretation of semantic contents. The hypothetical authorist is not able to distinguish the following three creative histories when she concludes that a work implicitly means p: (1) the author's semantic intention to mean p is successful; (2) the author did not intend to mean p but ended up writing a work meaning p; (3) the author intended p to be explicit in the work but failed.

It is not true that the hypothetical authorist is not able to distinguish the said creative processes, because hypothetical authorism does not forbid consultation with evidence of authorial intent (Nehamas, 1981, p. 145, 147); what she claims is only that this sort of evidence is not definitive in fixing the meaning of the work. This reply is ad hoc, the objector might say. Even so, the present objection still fails, for here Livingston seems to shift to a different topic: whether interpretation should reveal the actual creative history of the work. But this again assumes the historical perspective on criticism. The central point of contention in the debate has been how to determine what a work means. Distinguishing between different creative processes could be relevant, but it need not be. One way it could be relevant is where this distinction can help us correctly identify the work. For instance, whether a piece is impromptu or not could matter a lot with respect to its categorization that would resultantly lead to different verdicts of the work, in contrast to the case in which the work is otherwise viewed in regard to how it was produced.

6.8 Chapter conclusion

Before concluding this chapter, I quickly address two frequently raised worries for hypothetical authorism.

Some philosophers think that hypothetical authorism is not an adequate version of intentionalism at all, for an imagined author lacks the nature and force a real author has and hence the interpreter would just end up positing a causally inefficacious entity unable to explain the text (Carroll, 2000, p. 85; Davies, 2007, p. 186).[80] But hypothetical authorism

---

[80] Note that Carroll and Davies use the term "hypothetical intentionalism" to cover what I call hypothetical authorism and (Tolhurst's/Levinson's version of) hypothetical intentionalism. Davies sees the latter as a version

does not aim to give causal explanations; rather, it is normative explanations that she aims for (Irvin, 2006, p. 124). Alternatively, one can draw a distinction between causal explanation and interpretative justification (Nathan, 2006, pp. 282-283). The implied author is then a means by which one justifies her reading of a work.

Relevant to the previous one, another worry has it that hypothetical authorism ends up postulating phantom entities (hypothetical authors) and phantom actions (their intendings). In other words, the postulation is ontologically unnecessary (Trivedi, 2001, p. 202). To this the hypothetical authorist can claim that she is giving descriptions only of appearances instead of quantifying over hypothetical authors or their actions (Walton, 2008, p. 229). Indeed, most critics or philosophers drawing on this theoretical apparatus are not committed to the view that these fictional entities do exist. Even if they were, it would call for another occasion to justify the ontological necessity needed.

Now we have answers to the two questions I put forth at the beginning of this chapter. First, the reason why interpretative authorism is appealing is that a text is the product of an intentional action. To explain it, appeal to the action's agent is necessary. Even anti-intentionalism cannot avoid talk of fictional persona or appearances of intentionality. Hypothetical authorism does a better job than extreme anti-intentionalism in drawing our attention explicitly to I-meaning and makes it central in interpretation.[81] Second, the reason why the author we appeal to is virtual rather than real is that, if we want to preserve the centrality of the text, we need to abandon the authority of the real author. The text is a complex linguistic entity and is bound to have many interesting features the author is not aware of or cannot anticipate. To limit interpretation to authorial intent is somehow to turn critics into historians.

In the next chapter I will try to defend hypothetical authorism in more detail. The temporary conclusion is that this theory seems to me a convincing middle course between extreme anti-intentionalism and actual intentionalism.

---

of the former; that is, for him, all these variations involve a hypothetical author. Therefore, he is actually using "hypothetical intentionalism" for "hypothetical authorism."

[81] By "extreme" I mean the traditional interpretation of Beardsley's position, that is, an absolute formalist position rejecting contextual considerations and talk of the author. I have stated that this is not Beardsley's positions.

7. Defending Hypothetical Authorism

In this chapter I defend hypothetical authorism as the most convincing theory of literary interpretation. The defence consists of two parts.

In the first part, I argue that, from a descriptive point of view, the implied author account of interpretation best describes critical practice following the current positions based on the utterance model. That is, people who interpret in accordance with these positions end up interpreting an implied author. If that is the case, that *might* give us a reason for endorsing hypothetical authorism as a reasonable interpretative policy.

In the second part, I offer some arguments showing that hypothetical authorism is the best interpretative position to take.

7.1 Hypothetical authorism defended from the descriptive point of view [82]

7.1.1 The implied author argument

Theories of interpretation developed in the analytic philosophy of literature tend to be normative or prescriptive accounts. Specifically, these theories aim to offer a recommendation or policy for how we *should* interpret instead of a description of how interpretation actually proceeds. The main reason for taking this normative route is that a descriptive project is thought to be doomed to failure. This is because critical practice does not seem to be united by one single approach to interpretation (Davies, 2007, p. 13; Carroll, 2011, pp. 132-133).

As seen in the previous chapters, the normative route is usually combined with a theory of literary meaning and the ideal is to identify the interpretative aim with the search for work-meaning (Stecker, 2013, p. 314). That is, the aim of the project is to demonstrate that we should interpret in accordance with what a work *actually* means. Carroll even goes so far as to claim that the metaphysical aspect of a theory of interpretation is indispensable (Carroll, 2013, p. 10). However, this is not always the case. For example, it is unclear whether the value-maximizing theory, one of the contenders in the debate, has offered an explicit account of meaning. Nevertheless, the maximizing view is ostensibly defended on normative grounds.

---

[82] This section is an earlier version of my recently accepted paper "Interpretation and the Implied Author: A Descriptive Project" in *Southern Journal of Philosophy*.

In what follows, I take neither route. My aim is to explore the path less trodden: to attempt a descriptive account of interpretation. This account is likely to pave the way for a normative theory of interpretation but it does not need to. I will postpone its link to the interpretative policy to §7.2 and occupy myself with the descriptive aspect in the present discussion. Nevertheless, the account to be developed is not meant to be global. To be precise, I limit the applicability of the account to current theories based on the *utterance model*, which suggests we interpret a literary work in the same way we interpret an utterance. As discussed in §3.2.1, one of the underlying claims of the utterance model is that we interpret a literary work *as of its author*.

Almost all the positions discussed so far are based on this model, which claims that literary meaning is utterance meaning. This has the implication that interpretation should be constrained by a work's identity-relevant factors in its context of production because a work, like an utterance, acquires its identity and content in part from its relations to that context. The utterance model is thus a contextualist model for literary interpretation.

This model is representative in the contemporary analytic philosophy of literature since it is the basis for most theories of interpretation under debate (Davies, 2007, pp. 166-168). My contention is that people who follow the interpretative recommendations based on the utterance model are actually interpreting an implied author; that is to say, how interpretation proceeds in these apparently different cases can be *described* by a general account of the implied author. From now on, the domain of interpretative practice will thus be restricted to that based on the utterance model.

Hypothetical authorism, in its normative version, argues that interpretation should be guided by the intention of the author implied by the work. Below I try to show that in its descriptive form, hypothetical authorism best describes what is going on in the process of interpretation. Here is a precise formulation of my argument:

The Implied Author Argument (IAA)

(P1) The implied authors of literary works are the authors we can make out from those works.

(P2) Appeal to an author is necessary for working out the right interpretation of a literary work.

(P3) In working out the right interpretation of a literary work, we need to appeal to an author whom we can make out from that work.

(C) Appeal to a work's implied author(s) is necessary for working out the right interpretation of a literary work.

There does not seem to be much to be argued about the validity of this argument. I shall thus focus on the premises and examine them one by one in the sub-sections to follow.

7.1.2 The implied author revisited

Let us examine the first premise: *the implied authors of literary works are the authors we can make out from those works.*

In the previous chapter, I discussed different forms of Booth's original theoretical concept, such as the apparent artist, postulated author, ideal author, and fictional author. But all these variations share a common ground: the author concerned in interpretation is mainly constituted by work features (textual features plus context). I mention "work" and "context" because all the philosophers supporting hypothetical authorism are contextualists. This is something not characeristic of Booth's account. Hereafter I will use "work features" to mean textual features understood in the context of the work's production.

Simply put, the implied author in the present discussion is the author implied by a work, rather than by a text.[83] To be precise, such an author is constituted by a reasonably selected set of work features. The psychological traits expressed in a work, such as values, attitudes, thoughts, feelings, intentions and the like, when coherently grouped, constitute an author-image that does not necessarily coincide with that of the real author. Such divergences stem from accident, failure to realize intentions, or the actual author's intention to pretend.

Obviously, the first premise of IAA is true by definition. The implied author is the author the interpreter makes out from the work, and this is the standard construal of what an implied author is.

One point needs to be clarified before we move on. Hypothetical authorism does not imply interpretative monism. One can be a hypothetical authorist and yet hold critical pluralism in regard to interpretation. Better put, the account does not need to exclude the possibility of

---

[83] Though the implied author is designed to account for fictional works, what counts as such works remains controversial. Tentatively, I incline to the view that nonficitonal works also have implied authors, but will not argue for that here. Therefore, in my argument I will use "literary work" instead of "fictional work." Even if this stronger view fails, my central argumentation still covers fictional works, which constitue a major part of literary works.

non-equivalent but equally plausible interpretations (Currie, 1990, p. 101). The plausibility of an interpretation partly depends on the interpreter's value concerns and background assumptions (Currie, 1993, pp. 424-426). This means that a work could have more than one plausible interpretation based on different implied authors. This point will prove useful when I address a worry in §7.1.4.

7.1.3 Interpreting an utterance

The utterance model assimilates a literary work to an utterance, namely, a text *used* on a particular occasion. Literary works, according to this model, are texts of various lengths used by authors in different contexts. This model implies a contextualist position: the identity and content of a work are in part determined by relations it holds to its context of production.

This contextualist assumption is often used to support the suggestion that interpretation should not misidentify a work; that is, the interpreter should pay attention to the work's contextual factors that are identity-relevant, such as the work's title, genre, style, and facts about the author's oeuvre, social elements to which the work makes reference, and the artistic and linguistic conventions in place when the work was produced. Ignoring these identity-conferring factors leads the interpreter to consider a different work; in that case, the interpreter is not treating the work as authored.[84] Nonetheless, theorists disagree about whether or not the actual author's semantic intention is among the relevant factors. This is a debate between actual intentionalism on the one hand, and anti-intentionalism, hypothetical intentionalism and the value-maximizing theory on the other.[85]

Obviously, actual intentionalism and hypothetical intentionalism agree that the author's intention is *relevant* to interpretation. For the former, interpretation should be guided by the actual author's intention; for the latter, interpretation should be guided by the audience's best hypothesis of the author's intention. In either case, appeal to an agent or subject in determining the right interpretation is indispensable. By contrast, the anti-intentionalist seeks the interpretation best secured by linguistic convention, and the value-maximizing theory advises us to look for the interpretation that maximizes the work's value. In neither suggestion is the author explicitly invoked.

---

[84] When the distinction between work and text is drawn, "work" already implies contextual considerations in place and hence an author. I will use "work" to mean "work as authored" or "work as of its author" hereafter.

[85] As discussed in Chapter 2, I will treat anti-intentionalism as a contextualist position.

Now, the second premise of IAA asserts that *appeal to an author is necessary for working out the right interpretation of a literary work.* Note that this claim says nothing about the nature of this author, so the answer is left undecided for the moment. Apparently the claim holds with actual intentionalism and hypothetical intentionalism, but does it hold with anti-intentionalism or the value-maximizing theory? My impression is that in literary criticism most of the time the author is explicitly referred to. I claim that even if the author is not explicitly mentioned, the implied author is nevertheless implicitly invoked to determine the correct or acceptable interpretation. Below I will prove my claim, starting with the case of anti-intentionalism.

7.1.3.1 Interpreting an utterance: anti-intentionalism

To give an example, what follows is a paragraph from a critical piece on Su Shi's song lyric "Meditation on the Past at Red Cliff":

Internal evidence from the song itself, however, supports a different interpretation. The gentler, intellectual side of the image of the young general, in contrast with the image of nature depicted as "rugged" and "angry" in the first stanza, foregrounds the vulnerability of humankind. Human life is beautiful yet evanescent, created only to be swept away. The juxtaposition of the young general so vividly called forth in the poet's reflections of the ancient hero, long dead, with the living yet rapidly decaying gray-haired poet lamenting the past (line 17) *expresses* the *poet*'s perplexity over the inscrutable and devastating power of time. (Lian, 2008, p. 273, emphases added)

At first glance, the critic here seems to interpret the work by adopting anti-intentionalist strictures, since he bases his interpretation on internal evidence. However, the second half of the passage reveals that the meaning inferred is still attributed to an agent—the poet, who, in this case, is reasonably treated as an apparent author, since the critic is clearly talking about the work's appearance.

The anti-intentionalist may suggest that to turn this paragraph into a perfect example of non-intentionalistic criticism, we should omit at least the last reference to the poet. This way the intentionalistic reference can be removed (Beardsley, 1981a, pp. 26-28). But this approach strikes me as overly calculated. I doubt most critics would refrain from referring to

the author in this way, because avoiding reference to the agent does not seem to be a natural way to talk about an object intentionally made.

I believe there are perfect examples of non-intentionalistic criticism, but would maintain that in these cases the subject is at least implicitly posited. Even the anti-intentionalist concedes that the meaning of a literary work is "personal," which entails a subject or speaker responsible for it (Wimsatt & Beardsley, 1946, p. 470).[86] This in large part stems from the move of treating a literary work as an utterance. For instance, if we detect an anti-war attitude in a story, it does not seem to be a good move to say that the story, instead of the author, holds such an attitude, since the predication in question is inadequate. To solve this problem, the anti-intentionalist would need a "postulated" intention to explain why there is utterance meaning (Knapp & Michaels, 1985, p. 103).

Briefly speaking, the key lies in his presumption that every fictional work must have an ultimate speaker to whom meaning inferred from the text should be attributed (as discussed in §2.2). This claim is almost the core of an implied author theory of interpretation.

A difficulty in classifying Beardsley's view as a version of the implied author position is that his characterization of the story's presenter might apply better to the story's narrator than to its implied author. We can test this by examining different types of narrative modes to see whether the fictional speaker merges with the implied author in each of these scenarios.

The first factor to consider for classifying narrative modes is whether the narrator's presence is explicit or implicit. The narrative scenario in which the narrator is implicit can be further divided into two sub-types: either the story is told from an omniscient viewpoint or centers on the experience of a third-person character. In either case, the story is not told by any of the characters in the story; rather, it is told by an implicit speaker whose words the work purports to be. It seems reasonable to identify this fictional speaker with the implied author, for both function as the subject to which textual meaning is attributed.

As for the narrative mode in which the narrator is explicit, this involves first-person narratives. In these, either the narrator is reliable or unreliable. When the narrator is unreliable, a transcendental perspective is required in determining the text's meaning, because what is said ultimately in the work is not equivalent to what is literally said by the unreliable narrator. It follows that an implicit speaker has to be assumed and she again coincides with the implied author.

---

[86] Though it is widely believed that what the anti-intentionalist has in mind in this case is the narrator rather than the author, I argue that the former merges into the latter, as we shall soon see.

Where the narrator is reliable but textual meaning transcends what is literally expressed, an implicit speaker is at play again. This narrative scenario is thus better classified as a case in which the narrator's presence is implicit. This leaves us with the narrative scenario in which the narrator is a reliable spokesperson for the implied author.

The identification of the narrator with the implied author in the case last mentioned is controversial. The crucial difference between them is that the former is dramatized in the story while the latter is not. I am happy to accept that the narrator here is not happily called an implied author, though they are similar to each other in many respects.

It might be objected that the narrator and the implied author reside in different fictional worlds and this is what makes their merging impossible. But it is questionable whether this is a defining feature of the implied author; moreover, the interpreter can take the implied author to be an instrumentalist concept and hence avoid talk about the ontological status of fictional entities.[87] This latter reply is apparently what most literary critics have in mind.

The conclusion is that, Beardsley actually has a continued emphasis on appeal to a fictional agent of intentionality when dealing with the issue of literary interpretation, which is often overlooked in the philosophical literature on the topic. If this makes sense, the second premise of IAA would then be true for anti-intentionalism as well; that is, anti-intentionalism, a theory of interpretation based on the utterance model, invokes the concept of an author in seeking the right interpretation of a work.

My argument here can be further backed up by how Beardsley's followers defend his position. For example, Nathan has put a great emphasis on assuming an (implied) author for interpretation. Nathan's stance is broadly anti-intentionalist, but he argues that we must appeal to an utterer when interpreting an utterance. Otherwise, the project of interpretation cannot even get under way (Nathan 1992, 195). This agrees with Beardsley's emphasis on presupposing an ultimate speaker for every fictional utterance.[88]

7.1.3.2 Interpreting an utterance: the value-maximizing theory

Now, let us consider the case of the value-maximizing theory, according to which, to seek the interpretation(s) that maximizes the value of a work, the interpreter should consider the literary meanings a work can sustain within the limits set by convention and context. I do not

---

[87] Beardsley (1981b, 292-293) explicitly says that he does not intend to solve ontological issues when discussing fiction and fictional speakers.

[88] This section is a short summary of the argument I present in my 2018 paper.

think there would be a big difference in meaning if, in the previous sentence, we replace "can sustain" with "appears to have." If so, the maximizer might not be able to avoid assuming an *apparent* author.

To articulate my point, let us see how the maximizer characterizes her position: "[…] our interest in art rests on a recognition of its *intentionality*, but not necessarily as a result of a recognition of the artist's intentions." (Davies, 1982, p. 66, my emphasis) In another key passage, Davies writes: "[…] the maximizing theory can interest itself in the *appearance of intentionality* whether or not the relevant features were intended in fact. In this, it falls into line with hypothetical intentionalism." (Davies, 2006, p. 124, my emphasis)

As noted, by "hypothetical intentionalism" Davies also means "hypothetical authorism," which focuses exactly on the appearance of a work. But the appearance of intentionality seems to presuppose the appearance of an author; otherwise the agent of intentionality would be lost. In other words, one cannot talk about an apparent intention without presuming its agent.

Perhaps the proposal of invoking apparent intentions applies only in the case of irony, quotation, allusion, metaphor, symbolism, etc., namely, features that many people think need to be analyzed intentionalistically. Indeed, Davies claims that the above features must be so analyzed (Davies, 2007, pp. 177-178). He suggests we distinguish between two cases: the case where the said features are identity-relevant and the case where they are not. In the former, those features are more categorial than semantic; therefore, they will be accepted as involving categorial intentions. However, when the said features are not identity-conferring, the interpreter is free to consider their appearance.

Since the maximizer holds thin conventionalism, she is not committed to the view that a work can appear in only one salient manner. Take irony for example. If a work, constrained by convention and context, appears to be ironic, then irony will be an acceptable interpretation of the work, regardless of authorial intent; however, for the same work, if it is also reasonable to say that, within the constraints imposed by convention and context, it does not appear ironic, then a non-ironic reading would be legitimate. The point here is that, for the maximizer, the impact exerted by convention and context on a work's appearance is not robustly construed; therefore, a work can be taken to appear in more than one way. Then, saying that appearance should be analyzed intentionally would mean that, to use the above example again, there is an apparent intention that the work be ironic, or an apparent intention that it be not.

If we agree that, in the case of irony and the like, appearance should be analyzed intentionalistically, it is hard to see why the same does not apply in other cases of appearance. If a work appears to criticize capitalism, then we should say that there is an apparent intention to criticize capitalism. It would be inconsistent to say that in this case appearance should be analyzed non-intentionally while maintaining the opposite with cases of irony and the like (where these features are not identity-conferring).

The issue here is whether we can avoid any notion of an author when interpreting. It is true that the maximizing view acknowledges the author's categorial intention, but that is a move concerning work-identity rather than interpretation. As Davies claims, the maximizing position is intentionalist on the ontology of literary works, but anti-intentionalist on literary interpretation (ibid., p. 15). Indeed, from Davies' characterization of his position, the author and her intention do not seem to be a necessary ingredient in interpretation:

> The *value-maximizing theory* maintains that the conventions of language and art in place when the work was created are sufficient to secure its meaning. This last variant focuses on "utterance" rather than "utterer's" meaning, while recognizing that the meaning intended by the actual author usually identifies at least one of the meanings the work will sustain. Typically, more than one meaning can be put on a work within the limits set by the relevant conventions and practices, but meanings that cannot be reconciled to these cannot be attributed legitimately to the work. (Ibid., p. 151, emphasis original)

According to this construal, convention (plus context) is sufficient to secure legitimate interpretations, among which there is *usually* one intended by the author. This means that appeal to the notion of an author is not necessary when deciding legitimate interpretations, which is the core claim of conventionalism. But the maximizing position, as characterized above, construes works as utterances. Since for any utterance there has to be an utterer, I question whether utterance meaning can be determined without reference to an utterer.

Note that the maximizing view seeks what an utterance could mean. In Stecker's terms, this will be the epistemic sense of "could," namely, there is a possibility relative to the evidence we have (Stecker, 2003, p. 66). In other words, the maximizer seeks *possible* meanings an utterance bears, given textual and extratextual evidence.

Now, it seems to me what an utterance could mean is just another way of saying what the utterer could mean by that utterance. Recall the claim made by Knapp and Michaels, that

intention and meaning are inseparable. According to them, even recognizing a sequence of symbols as a sentence,

> […] we must already have posited a speaker and hence an intention. Pinning down an interpretation of the sentence will not involve adding a speaker but deciding among a range of *possible* speakers. […] as soon as we attempt to interpret at all we are already committed to a characterization of the speaker as a speaker of language. We know, in other words, that the speaker intends to speak; otherwise we wouldn't be interpreting. (Knapp & Michaels, 1992, p. 53, my emphasis)

On the same page they continue: "[in the case where] we have less information about the speaker […], […] the relative lack of information has nothing to do with the presence or absence of intention."

If the point they make is convincing, the maximizer cannot reach meanings without reference to a speaker and her intention. And since it is an apparent/possible intention that is under scrutiny, the agent responsible for this intention will be an apparent/possible speaker. I conclude that the maximizer would need to presuppose an apparent author responsible for the possible meanings her literary utterance bears.

Finally, it is of interest to note that, even if we shift to other models that are not based on the analogy between utterances and works, it is doubtful that the intentional aspect can be totally avoided. For example, according to the model given by Nehamas, to interpret a text is to construe it as the product of an action; in doing so, one subjects the text to certain inquiries and offers answers to them. Since the text is the product of an action, explaining it again calls for appeal to its agent (see §6.3).

7.1.4 Interpreting the implied author

It seems that there are only two options regarding the nature of the author invoked in interpretation: either she is the implied author or the real author. The third premise of IAA asserts that the author in question will be an implied author.

In the case of anti-intentionalism and the value-maximizing theory, the apparent author implicitly assumed is just the implied author because what is implied in a work is what the work appears to be (see Chapter 6; see also Walton, 1990, p. 370). Since both positions focus on a work's appearance and both involve presupposing an apparent author, followers of these

positions are actually interpreting an implied author. But how do things stand with actual intentionalism and hypothetical intentionalism that target the actual author? Let us first examine the case of hypothetical intentionalism.

7.1.4.1 Interpreting the implied author: hypothetical intentionalism

An interpreter practicing hypothetical intentionalism will hypothesize the actual author's intention with the help of publicly available information about the author found in the context of the work's production. The best hypothesis in this case is mainly based on textual and contextual evidence. As discussed in §5.2, this may require ignoring all evidence about the author's semantic intention. Then there is virtually no difference between the evidence that constitutes the author for hypothetical intentionalism and that for hypothetical authorism, for the interpreter following either position is constructing an intention based on the same resources.

Some might feel tempted to say that there will be virtually no difference between the evidence that constitutes the author for hypothetical intentionalism and that for hypothetical authorism, for the interpreter following either position is constructing an intention based on the same resources. This claim seems problematic because not all hypothetical authorists put emphasis on the actual author's oeuvre as Levinson does, which produces a crucial difference in interpretative resources by which the interpreter constructs the intention to be attributed to the author. [89] However, some hypothetical authorists such as Nehamas do include consideration of the actual author's oeuvre as a constraint on interpretation, and there is nothing in my argument requiring that hypothetical authorism as a general label must refer to a specific version.

At this juncture, hypothetical intentionalism faces the challenge that it is arbitrary to exclude *only* evidences about the author's semantic intentions from all the contextually

---

[89] Currie and Nathan are hypothetical authorists in this respect. A discussion on how hypothetical authorists construct authors bearing different degrees of resemblance to the actual author can be found in S. Davies, 2007, p. 180.

relevant factors. The hypothetical intentionalist does not seem to offer any argument justifying the present theoretical choice. If it is the actual author whose intention we hypothesize, then why should we ignore her semantic intentions, while we consider all other evidences about her, such as that concerning her categorial intention and oeuvre? Of course we have the claim made by the hypothetical intentionalist that semantic intention does not determine meaning. This is because hypothetical intentionalism derives from the theory of meaning that bases meaning on the audience's uptake. In that case, meaning is not to be analysed by reference to the actual speaker's intention. The question becomes whether the theory of meaning endorsed here is convincing. But that is something on which consensus can hardly be reached. Hypothetical authorism can get away with all this because it does not target the actual author. Evidence about the actual author's semantic intention can be ignored because the subject of interpretation is the implied author, not the actual author.

As suggested at the end of Chapter 5, hypothetical intentionalism is in danger of collapsing into hypothetical authorism. Here I cite passages explicitly indicating Levinson's commitment to the implied author position. When elaborating hypothetical intentionalism, Levinson says: "a work of art says what, on the basis of the work contextually construed, it would be reasonable to impute to its artist as a view that he or she both significantly held and was concerned to convey." (Levinson, 1995, p. 188) This is the standard characterization of hypothetical intentionalism. On the same page, Levinson continues: "Some have thought to capture the drift of the preceding reflections in terms of the notion of an *implied artist* or *author* […]" (emphases original). On the next page Levinson soon distinguishes between two kinds of implied author. The first kind is called the "thinly" implied author: "this is roughly the mind you would infer is behind a work on the basis of the work alone, given only general information about the period of creation, genre conventions, and prevailing language involved." Obviously, Levinson here is referring to Beardsley's position, and his interpretation of Beardsley supports my argument in §7.1.3.1. That is, anti-intentionalism is actually hypothetical authorism.

The second kind of implied author Levinson has in mind is the "thickly" implied author: "this is roughly the mind you would infer is behind a work given the work and the aforementioned minimal historical context, but also the author's previous works, his public self or image, and the specifics of his situatedness in relation to his surrounding culture and society […] So does a work say what it seems to when we project its implied author thinly or thickly? The latter, I suggest." The interpretative method described here is precisely the standard version of hypothetical intentionalism offered by Levinson himself, except that the

"author" referred to is construed as an implied author. This confirms that the author that the hypothetical intentionalist has in mind is the implied author, not the actual author.

In his latest defence of hypothetical intentionalism, Levinson insists that the position targets the actual author, rather than the hypothetical or implied author (Levinson, 2010, pp. 140-141). However, in their reply, Stecker and Davies seem right to say:

We do not think talk of hypothetical authors is out of place in this context given that it is not the full-blooded author (private diaries and all) who is considered. It is appropriate to talk of a hypothesized author who shares with the actual author his public persona. And this is more plausible than insisting that the interpretation considers the actual author, given that HI rejects as irrelevant the actual author's intentions at the point where they come apart from what his public persona is most likely to have intended. (Stecker & Davies, 2010, p. 308)

Indeed, it seems that the concept of "appearance" mobilized by hypothetical authorism explains the relevant interpretative resources better than the concept of "hypothesis" endorsed by hypothetical intentionalism. After all, there is nothing implausible in saying that what hypothetical intentionalism targets is the apparent author's intention (what is apparently said by the work as made out by the audience), as Levinson himself admits. Levinson's uneasy shifts between actual and hypothetical authors show only that the distinction he intends to draw may not be as legitimate as he expected.

7.1.4.2 Interpreting the implied author: actual intentionalism

The case of actual intentionalism is complicated by the fact that there are two versions of it: the extreme version and the moderate version. The only advocates of the extreme version seem to be Knapp and Michaels, who hold that work-meaning is to be identified with the author's intended meaning. Their account is usually read as a Humpty-Dumpty theory of meaning: the author can mean whatever she intends the work to mean (Carroll, 2016, p. 311). The upshot is that failed intention is impossible on this account, which creates cases where the third premise of IAA will not hold. This is because the presumable case of failed intention is a case in which the author intends her work to mean something the work cannot possibly appear to mean. It follows that the author appealed to in this case cannot be the implied author.

I do not dispute this counterclaim, but I suggest that extreme intentionalism be excluded from the present discussion. There are two reasons for this. First, it is doubtful whether this position is a viable theory of interpretation. Given its outlandish claims, I doubt that any sensible reader or critic would put this position into practice. As a theory of meaning, it does have its theoretic points to make; but it is difficult to imagine people doing Humpty-Dumpty criticisms. If extreme intentionalism is not viable in critical practice, I do not see any reason to include it in my project, which attempts to describe how people interpret.

Second, we may want to view the theoretic contribution made by Knapp and Michaels in a different way, given that it is not always crystal-clear that they really aim for a Humpty-Dumpty theory of meaning.[90] They have been more successful in showing that *intentionality* is a necessary condition for a sequence of marks being a piece of language, namely, being something *meaningful*. Their argument is more convincing and makes more sense when construed as arguing for intentionality as a necessary condition for something being language (and hence meaningful), than when construed as arguing that intentionality is both a sufficient and necessary condition for work-meaning, and so infallibly confers meaning. I believe this is a reasonably charitable and constructive way to rethink their argument. Such being the case, our focus will be on moderate actual intentionalism.

Most of the time, the moderate intentionalist interprets in the same way that a hypothetical intentionalist does, that is, by making hypotheses about the author's intentions; however, a best hypothesis will be abandoned when there is independent evidence that overturns that hypothesis.

The most salient feature of modest intentionalism is that the author's intentions, to be taken seriously, are only the successful ones. A point to note first is that moderate intentionalists agree that, when the author's intention fails, meaning is fixed by convention and context alone. This is an anti-intentionalist move, because it is to acknowledge the possibility of failed intentions. However, it seems that in this case there still exists a sense of authorial relevance, for the interpreter will not know that the author's intention is unsuccessful, so as to appeal to convention plus context instead, unless she appeals first to external evidence of the author's intention. This crucial difference does not exist in the account offered by the anti-intentionalist or the maximizer.

It follows that, to work out the right interpretation of a work in the case of failed intention, the interpreter still needs to appeal to the author, because what she aims for is what the author

---

[90] Such an interpretative hesitation can be seen in, for example, Carroll 2016, 311, and Lamarque 2007, 125-126.

ends up saying. Once this subject is in place, the meaning reached in this case will be the same as that reached by the hypothetical authorist, for in both cases meaning is construed in terms of how the work appears to be.

Next let us examine the case of successful intention. Although opinions differ about how to define whether an intention is successfully realized in a work, the success condition is typically measured by the strong or weak standard. Consider the strong standard first. This means that the intended meaning is the one apt to be recognized by the audience with the relevant contextual knowledge (Stecker, 2003, p. 14). In this case, the intention that meets the success condition is no doubt that of the implied author. If an anti-war attitude is eminent in a work then it is going to secure the best uptake from the audience and hence qualifies as one part of the implied author. Every intention taken to be successful by the moderate intentionalist adopting the strong standard is already implied in the work.

The weak standard, roughly put, claims that an intention is successful just in case it is compatible with the work (Carroll, 2016, p. 310). Carroll never elaborates this compatibility, which is done instead by Livingston in terms of the meshing condition.

The requirement of meshing rules out the strange case in which the intention in question could not possibly be discerned by any well-prepared audience, whereas it is not explicitly rejected by textual evidence. Livingston's example is this. Suppose the Japanese writer Natsume Sōseki actually intended the main character in his realist trilogy novels to be a Martian in disguise. Though no textual evidence explicitly rejects this reading, it does not meet the weak success condition because "[…] if no features of the novels' characterizations resonate with the Martian intention, the latter should be discounted." (Livingston, 2005a, p. 155) Since the meshing condition requires a high degree of coherence between the author's intention and textual structure, it seems plausible to assume that the successful intention could be discerned by an appropriately backgrounded audience without independent knowledge of authorial intent. If so, the intended meaning would be salient enough to be implied in the work. As Carroll himself admits, a successful intention is by definition within a work and discoverable by the reader (Carroll, 1992, p. 160). The fact that the author is made up of work features makes her implied.

Some might object that the interpretation reached by the modest intentionalist adopting the weak standard might not be one that will be targeted by the hypothetical authorist. If so, hypothetical authorism cannot be said to really describe the interpretative practice suggested by moderate intentionalism. This is because there are two possibilities with respect to the strength of uptake: either the intended meaning is the most apt to be recognized by the

audience or it is less ascertainable (but not indiscernible, as indicated in the previous paragraph). It seems that the second case would escape hypothetical authorism.[91]

I doubt that the distinction between strong and weak uptake is absolute. What appears most saliently to one person may not appear so to another. This links back to what I said at the end of §7.1.2: different background assumptions and value concerns result in differently plausible author-images. One would assume that the modest intentionalist would not want to say that the interpretation reached by the weak standard could be implausible. Then such an interpretation could well be plausible for a hypothetical authorist with a particular set of assumptions and value concerns in mind.

7.1.5 Section conclusion

In the first half of this chapter, my aim has been to show that hypothetical authorism best describes the critical practice suggested by the utterance model. All the positions based on this model in the current debate end up appealing to an implied author when working out the right interpretation of a work. In the case of literature, it is doubtful that the interpreter can ever avoid interpreting the implied author, for it seems that a reasonable balance between utterer's meaning and utterance meaning would always commit us to an implied author.

7.2 Hypothetical authorism as an interpretative recommendation

7.2.1 Some remarks on the descriptive account's prescriptive power

If my argument (IAA) so far is convincing, probably all the positions based on the utterance model will end up interpreting the implied author. In the case where the actual author is thought to be invoked, it is actually the implied author that is in play; in the case where no author is explicitly appealed to, there is an implied author assumed. The question now is whether IAA, if sound, serves as a reason for us to favor hypothetical authorism as the best interpretative recommendation.

---

[91] Here I point out one inconsistency in Carroll's account. He holds that the work is the best evidence for authorial intent (Carroll, 2000, p. 77). If this claim is to be taken seriously, it is hard to reconcile it with the weak standard, because the weak standard may render the work *second* best evidence for authorial intent.

There is some degree of support for this view, I believe. If the interpreter is convinced that she is actually interpreting an implied author when following another interpretative recommendation, she might also be convinced that she should just adopt the implied author position all along. For example, if an interpreter following actual intentionalism comes to believe that the author she has been appealing to is an implied author, then there seems to be some plausibility in making things simpler, that is, having just one sort of author involved in the process of interpretation.

I agree that this will not be the chief support for favoring hypothetical authorism as the best interpretative policy. The main reason is this. All that IAA claims is that all the major interpretative positions based on the utterance model end up interpreting an implied author; it does not make the further claim that these positions will arrive at the same interpretation(s) as hypothetical authorism does. For example, though the interpretation that maximizes a work's value is suitably attributed to the implied author, it is not necessarily the interpretation reached by the hypothetical authorist interpreting the same work. In other words, IAA does not hold that the interpretation made by other positions endorsing the utterance model *must* coincide with that made by hypothetical authorism. The account at most implies possibility, but not necessity. That is, every interpretation reached in the former case is likely to coincide with that reached by hypothetical authorism (particularly when it accepts critical pluralism). This is the best IAA can claim in this respect. And the possibility of divergence in the results of interpretation could undermine the support IAA gives for hypothetical authorism as a prescriptive account.

To strengthen my defence, below I offer some more arguments in support of hypothetical authorism as the ideal interpretative recommendation. My line of reasoning goes as follows. First, I reinstate the need to preserve talk of the author in interpretation, which gives us a reason to place less priority over versions of anti-intentionalism that do not pay explicit attention to the concept of an author. Second, I argue that actual intentionalism is not a viable position with respect to interpretation, mainly because its strictures are almost impossible to carry out. Finally, I show that hypothetical authorism does a better job than hypothetical intentionalism in achieving a balance between objective interpretations and the reader's interpretative freedom.

7.2.2 The author and her place

In this section, I briefly restate the need for preserving talk of the author in interpretation. Here I agree with Nehamas and Nathan: referring to a subject responsible for a product intentionally produced is the most natural way to talk about that product. If we do not posit an agent, it is hard to see how we can explain work features, or the "doings" (in Stecker's term) in the work.

I suspect most critics can avoid talking about the author when it comes to interpretation. I say "interpretation," not "criticism," because the anti-intentionalist's suggestion is not that we should avoid talking about the author entirely, but that we should refrain from using intentionalistic phrases such as "the author intended to express" when considering work-meaning. But work-meaning here is loosely construed as a variety of artistically relevant features and doings. Given this, it would be an annoyance for the critic to constantly shift between intentionalistic and non-intentionalistic phrases. This would not be a natural flow of writing: such a demand could keep interrupting the critic's train of thought when writing. Let us examine another example from Lian's critical pieces. When interpreting Li Qingzhao's most famous poem "One Beat Followed by Another, a Long Tune," Lian writes:

The whole piece can be summed up by one word, "sorrow." As we have seen, the poet *emphasizes* this at the beginning by repeating the idea of sorrow six times in line 3. Now, at the end of the song, she tells us that the word "sorrow" just cannot express what she has tried so hard to say. Suddenly the poet sounds like Zhuangzi (ca. 369–286 B.C.E.), the language skeptic who wished that he could "have a word" with someone "who had forgotten words." However, the poet has also *inherited* Zhuangzi's dilemma. She has no other medium but language, ironically, even when she *wishes* that her readers would bypass language. Her best hope is that some kind of unmediated grasp of her sentiment can be achieved by those readers who are willing to go beyond language and try to experience what her words attempt to convey. Seen in this light, her unconventional use of sounds at the beginning of the song can be read as a direct appeal to readers' sensual, rather than simply intellectual, perception. (Lian, 2008, p. 276, emphases added)

What I have italicized is the intentional doings of the author. To convert the relevant sentences into non-intentionalistic terms, we would need to eliminate references to the author, and probably also change the verbs to avoid odd predications. For example, it would be inadequate to say that the poem *wishes* something. But doing so means that we might also need to change other parts of the passage to make the language flow naturally. I am not

169

saying this kind of revision is impossible, but this writing strategy could lead to a laborious job. I suspect that the most natural way of interpreting is to refer to the author and her intentional doings all the time. This concern prompts me to conclude that (the traditional version of) anti-intentionalism is not the best interpretative policy for interpretation.

The above concern also figures in the case of anti-intentionalism in the broad sense, namely, the value-maximizing theory. It seems that the maximizer will need to attribute the interpretation that maximizes the value of the work to some agent. Note that the interpretation in this case is one that would make a work an artistic achievement; and an achievement demands an achiever.

The maximizer can object that we praise or blame what one achieves even if what is achieved is not what one intended to achieve. For example:

> Tom insulted Mary, although he did not intend to. Despite his lack of intention, the insult might be blameworthy. Ron set the world record for one-hundred meters, although he only intended to win the heat and had no thought at all of the record, before or after. Surely the race he ran should still be an object of praise and of great value in the world of track. If anything, such cases seem even more obvious in the artworld, where artworks exist as public objects to be fully appreciated in their own right. Even if we grant that artistic value lies only in what an artist achieves, uncovering such value might not lie in discovering the artist's intentions. (Goldman, 2013, p. 31).

The examples here strike me as inadequate. It seems right to blame Tom for accidentally insulting Mary, and it also seems fine to praise Ron for setting a world record. But it seems *odd* to praise James Joyce for *accidentally* producing a great work. Typically, we would not expect an admirable work to be produced by accident; however, we do expect that, from time to time, people could by accident say something offensive. Given that the maximizing view is tolerant of multiple interpretations among which the author's intention *could* name one, it seems to entail the odd claim that authors very often produce valuable works by accident.

The said problem stems from the improper subject to which the maximizing position attributes the artistic achievement. As argued in §7.1.3.1, there should be an apparent author behind the appearance endorsed by the maximizer. The postulation of an apparent author would then remove the accidentalness found in the problematic claim in question.

The maximizer might object that she does not claim that the interpretation arrived at by her position necessarily renders the work valuable; rather, what she means is that the

interpretation in question is intended to make the work as valuable as possible within designated confines, and this does not entail that the work must be valuable in an absolute sense when so construed. However, the following passage does not seem to strongly support this rejoinder:

[…] what is valuable in the work will vary according to the presupposition the interpreter adopts. Even if only one interpretation matching each initial premise maximizes the work's value, a change of premise will likely lead to quite different interpretations some of which may present the work as no less valuable, though this value will result from a different combination and balance of virtues. (Davies, 2007, p. 189)

Such being the case, one wonders whether the standard for value-satisficing (a term the maximizer thinks is more adequate) would be too liberal. In other words, it might not be too difficult to arrive at an interpretation that is able to impute artistic merits to a work. Note that in the maximizer's account, no story is ever told about whether we would have a work to which no artistic values can be imputed at all. The impression remains that it is possible to find features that satisfy artistic values in any given work, or at least, most works.

This might not be a serious problem, for it is possible for the maximizer to say more about artistic value (for example, Goldman, 1995) and perhaps accept that there can be cases in which no interpretations will make the work valuable. However, the latter option appears unappealing, as it would leave plenty of cases where no legitimate interpretations can be determined.

By contrast, hypothetical authorism is able to bypass the above difficulties related to value. In spite of that, some might object that hypothetical authorism fares no better in falsely attributing artistic achievements to an imagined entity, rather than to the real author. An initial reply is that, as discussed, the artistic achievement is value-related, but the hypothetical authorist does not seek value-conferring interpretations. So this criticism makes more sense in the case of the value-maximizng view when it presupposes an apparent author.

Still, my reply does not satisfactorily resolve the issue of how we can evaluate a work in terms of a fictional author's performance without considering the actual author. But I do not see any faults with evaluating a work this way. If my conclusion in §7.1 is convincing, that all the current positions based on the utterance model end up interpreting an implied author, it does not seem implausible to say that we are constantly evaluating an implied author when adopting these positions.

Alternatively, it is possible for hypothetical authorism to evaluate a work based on the actual author's performance. Recall Booth's dual characterization of the implied author. On the one hand, the implied author is a version of the actual author, in the sense that she is created, wittingly or unwittingly, by the actual author; on the other hand, the implied author is constructed by the reader, in the sense that she is constituted by a set of work features grouped by the reader to form a coherent author-image. Then, on the first construal, we can at least evaluate the actual author's doing indirectly. For example, if the intention we attribute to the implied author makes the work a splendid one, we might praise the actual author on the basis that the implied author is part of the actual author in the above sense. This account is rather sketchy, but indicates a possible direction worth exploring.

7.2.3 The unavailability argument revisited

So far I have offered reasons for why the interpreter should favor positions that maintain talk of the author. In this section, I will give a further reason to exclude actual intentionalism in general. In other words, I intend to argue that any position that *recommends* us to settle interpretative disputes by appealing to the actual author's intention is not a good interpretative recommendation. I will utilize the unavailability argument developed in §4.4 and say more about it.

The main idea of Beardsley's unavailability argument is precisely that the fact that the author's intention is normally unavailable is a reason good enough for us to seek textual meaning (work-meaning if we interpret Beardsley as a contextualist) instead of authorial meaning (the author's intended meaning). The premise that authorial intent tends to be unavailable seems quite convincing, because, under normal circumstances, the author will not reveal her intention unless asked to. It is true that with the advent of social media such as Facebook or blogs, authors may feel more motivated than before to talk about their works publicly, but I still doubt that this phenomenon is prevalent enough for us to confidently claim that, in our time, the author's intention is *normally* available. I will strengthen this claim later. For now, let us examine the actual intentionalist response to the unavailability argument.

As Carroll suggests, the actual intentionalist can maintain that the work is the best evidence for authorial intent. Therefore, one can engage with the work in order to seek authorial meaning. This means that evidence for authorial intent is *always* available. However, here Carroll seems to conflate two kinds of evidence.

Certainly, when Beardsley says that the author's intention is normally unavailable, he means that *evidence* for such intention is unavailable; and clearly, in this context he excludes textual evidence. As discussed in §6.6, the work is only *indirect* evidence for authorial intent; by contrast, the author's proclamations of her intentions will be *direct* evidence for her intentions. What Beardsley disputes is the availability of the latter kind of evidence, not the former.

I have noted in §7.1.4.2 that Carroll fails to give a coherent account of the work as the best evidence for authorial intent. Here the same problem presents itself. If the work is indeed such evidence, then Carroll is basically giving up his weak standard for successful intention, which assumes the pervasiveness of textual ambiguity. If a literary text tends to be ambiguous, how can it be the best evidence for authorial intent? Such being the case, Carroll would accept the strong standard as Stecker does. But this cannot be, as it will completely ruin Carroll's theory of interpretation, according to which authorial intent functions as a meaning-disambiguating tool.

Even if we ignore the incoherence on Carroll's part, accepting the work as the best available evidence for authorial intent is still not a good move. Is the work really a better piece of evidence than the author's notebook, correspondence, or interview? If so, why does the actual intentionalist strive to argue for the use of biographical evidence? It is true that a casual proclamation of intention from the author may not be good evidence for her intent; nonetheless, a serious interview, or her notebook that details writing plans and relevant thoughts, may well be credible evidence for what she intends to say in the work. The best-evidence claim clashes with the defence of biographical data.

It seems that, the work is the best evidence for authorial intent *only if* we accept that the phenomenon of unintended meanings is not pervasive. That is, we accept that the possible intention we grasp from the work has a good chance of being a successful intention. This is a very controversial claim, and I am inclined to think this is probably false. In the main, I agree that very often authors successfully realize their intention; otherwise, authors would seem to be very incompetent. However, it seems false to say that every unintended meaning results from a failed intention. If we agree that multiple interpretability is a common phenomenon in literature, it seems fair to claim that a large number of unintended meanings come by accident. From the discussions in Chapter 4, the impression remains that this is a point which actual intentionalists do not dispute.

Suppose the actual intentionalist concedes that the work is unreliable or indirect evidence for authorial intent. The problem I just raised would thus evaporate. Beardsley seems

comfortable with this potential reply, and does not pursue the unavailability argument any further.

As I see it, the unavailability argument itself still has the potential to dispose of actual intentionalism. My point is that the search for authorial meaning in the work cannot be justified by the work's weak evidentiary support for such meaning. If the work is bad evidence for authorial intent, seeking it in the work could be a futile task. Now we must ask: under what circumstances would the work be a piece of bad evidence for the author's intention? Two possibilities come to mind: either the intention is not there in the work, or it is there but hard to discern. In the former case, the author's intention is not available in the work at all; the interpreter's *hypothesis* about the author's intention will be based on inadequate evidence and never hit the target. In the latter case, the interpreter is more likely to focus on an unintended meaning that is more apparent to the eye than the intended meaning. This means that the hypothesis the interpreter makes about the author's intention will very likely be false. Though, strictly speaking, in this kind of case the author's intention is available in the work, textual evidence does not function as an effective guide to the existence of the intention in question. In both cases, searching for authorial intent in the work would result in the interpreter's making false hypotheses.

At this stage, the real force of the unavailability argument begins to surface. The thrust of the argument lies not in the unavailability of authorial intent, but in the *status* of the hypothesis made by the interpreter about such intent. As shown in the previous paragraph, in the case where there is no evidence of authorial intent otherwise obtained (call this the *internal* case), the interpreter would end up making *wrong* hypotheses. To make things worse, there is no *additional* evidence that can change the status of the hypothesis; the hypothesis will stay wrong and unproved probably forever. Such a consequence makes actual intentionalism a very unappealing recommendation for interpretation.

Now, as mentioned early on in this section, the actual intentionalist might object that in the internet era we have more chances than before to know what the author was thinking about when creating her work. Therefore, cases where independent evidence of authorial intent is available (the *external* case) are not as rare as the opponents of actual intentionalism expect. I have expressed doubts about this reply; below I show this reply is actually untenable.

Recall the claim that work-meaning is a complex (call this the *meaning-complex thesis*). Implicit meanings abound in different proportions of a work: words, sentences, paragraphs, chapters, and the work as a whole. Normally, work-meaning refers to a reasonably selected set of these sub-meanings; even so, it is still a rich complex that a single proclamation of

intention cannot capture. Given the meaning-complex thesis, how do actual intentionalist justify their hypothesis for every bit of this complex? It just seems false that, for *all* literary works, *all* of the author's intentions for a particular meaning-complex are available to the reader. I believe that my claim remains plausible even if we replace "all" with "most."

It follows that, in the external case, the actual intentionalist's interpretation will consist of proved hypotheses and unproved ones, an unfortunate mix. Such being the case, the interpretation will be a conjunction of interpretative statements highly likely to be false. This does not seem to be a favorable result for the interpreter.

To sum up, the recommendation that the literary interpreter should concern themselves with authorial meaning is unappealing, because the unavailability of authorial intent will lead the interpreter to produce a significant number of unprovable hypotheses. This updated unavailability argument might be better renamed the *unprovability argument*, so as to reflect the frustrating situation in which the interpreter, in seeking truth, can live only with hypotheses. This is why Beardsley concludes that "the *general* and *essential* task of the literary interpreter cannot be the discovery of authorial meaning." (Beardsley, 1970, p. 33, emphases added)

7.2.4 Interpretative freedom and constraints

7.2.4.1 Hypothetical authorism vs. hypothetical intentionalism

Based on the discussions so far, hypothetical authorism seems to be an adequate middle course between (extreme) anti-intentionalism and actual intentionalism. On the one hand, it preserves intentionalistic criticism for the critic, a more natural form of criticism as argued; on the other hand, it avoids the stringent demand on interpretation imposed by actual intentionalism.

A middle course between anti-intentionalism and actual intentionalism also means a balance between interpretative freedom and objective constraints, which is notably a merit of hypothetical intentionalism. On the one hand, the audience plays a part in constructing work-meaning; on the other, her play is constrained by objective factors. However, hypothetical authorism has the same merit. Such being the case, further reasons are needed to qualify hypothetical authorism as a better recommendation than hypothetical intentionalism.

My opinion is that talk of the work's appearance preserves and explains interpretative freedom better than talk of hypothesizing the actual author's intention. This is because

interpretative freedom means that the interpreter's reading of the work is not to be bound by indications of the actual author's intention. In that case, it would be inappropriate to characterize what the interpreter does as hypothesizing about the actual author's intention. Here an important claim is that our psychological tendency is to hold onto the interpretation that we think best explains the work, but this does not mean that what we are doing must be hypothesizing the actual author's intention. We might resist hypothesizing the actual author's intention because we are aware that our reading might not be what the author has in mind.

Consider the case in which a critic reviews a novel and writes a review for his readers. There are two important facts to note. First, it is very unlikely for the critic, when writing the review, to contact the author and ask for his or her intentions; that is just not what a critic normally does. Second, it is also very uncommon for the critic to publicly correct her interpretation when there is evidence indicating that the actual author did not intend the meaning reached by the critic. I think most—if not all—critics are perfectly aware of these two facts. It is thus hard to see why and how critics (let alone normal readers) would accept the recommendation offered by hypothetical intentionalism that the interpreter should target and hypothesize the actual author's intention. Should they put that recommendation into practice, they probably would come to realize at some point that they do not want to describe what they are doing as hypothesizing the actual author's intention. Below I use a cinematic example to demonstrate this point.

7.2.4.2 A case study: *Open Your Eyes*

Open Your Eyes (*Abre Los Ojos*) is a 1997 Spanish psychological thriller film directed by Alejandro Amenábar, starring Eduardo Noriega and Penélope Cruz. There is a 2001 Hollywood remake retitled *Vanilla Sky*, starring Tom Cruise and again, Penélope Cruz. Since the plot of the remake is almost the same as the 1997 film, I will treat the remake as a different version of the same work. Below I will focus on the original film but my point applies to both. Some might think the remake is a different work, but the ontological issue does not really matter here.

Below is a short but accurate summary of the film (spoiler contained):

César is living a perfect life. He is rich and handsome, and has a way with women. After having an affair with Nuria, César meets the beautiful and fascinating Sofia. Even though his best friend clearly has a crush on her, César cannot help falling in love with Sofia.

Nuria, raging with jealousy, offers César a ride home and commits suicide by crashing their car into a wall. While César survives the crash, his face is completely disfigured. Still in love with Sofia, he decides to look her up and take her out to a club. While the evening ends miserably—Sofia and his best friend take off and César ends up falling asleep on the sidewalk—his luck seems to change the next morning when Sofia wakes him up and declares her love for him. When, in addition, plastic surgeons are suddenly able to restore his face, happiness seems to be coming his way.

One night, César wakes up and finds himself in bed with a woman he takes to be Nuria. Convinced that she has kidnapped Sofia, he kills her. When everybody keeps telling him that the woman he has killed is in fact Sofia, César no longer understands the increasingly surreal world that surrounds him and starts to lose his mind completely. He ends up recounting his past to a psychiatrist in a mental institution. At the end of the movie, it becomes clear to César what exactly happened. Everything César experienced after the night at the club was actually a dream, that turned into a nightmare. In real life, César woke up on the sidewalk, never saw Sofia again and grew more and more depressed. In the end, he decided to kill himself, after signing a contract with *Life Extension*, a company that agreed to freeze him after his death, erase the last part of his memories and replace it by what the company calls a *Lucid Dream*. This *virtual* reality (in which he is with Sofia, has his face rebuilt, kills Nuria and ends up in the mental institution) is not real at all. It is all in the mind of César, whose dead body is lying in a freezer and whose brain is hooked up to a machine that provides the sensory input needed for César to experience his *Lucid Dream*. (Engelen, 2010, pp. 33-34, emphases original)[92]

What is the theme of this film? What comes to my mind first is Nozick's experience machine. As Engelen indicates in his discussion of the film:

The most prominent philosophical issue that is raised in both *Abre Los Ojos* and Nozick's thought experiment concerns the nature of happiness, which is arguably the ultimate goal in most people's lives. As Nuria asks César right before she crashes their car into the wall, the real question is: 'what is happiness to you?' Is it all about enjoying yourself and having pleasurable experiences? What is happiness, if not simply the feeling of being happy? (Ibid., p. 35)

---

[92] I edited the text slightly to make it flow more fluently in this abstract.

Nuria's question for César is a very crucial clue. Most films involving simulated reality can be seen as bearing connections to Nozick's experience machine. However, to be precise, the major theme of these films is actually *external world scepticism*; in other words, the epistemological issue in these films stands out more than the ethical issue. No doubt the latter is Nozick's main concern, although he does mention in passing that the former would be a related issue (Nozick, 1974, p. 45). Such being the case, to interpret *Open Your Eyes* as reflecting or implicating the experience machine argument seems a reasonable move to make.

The film's connections to the experience machine argument is also highlighted by César's decision to plug into and unplug from the machine. César chooses to use the machine to stay in a beautiful dream, and decides to wake up from that dream in the end. Toward the denouement of the story, César is presented with two options: stay with Sofia in the dream forever or wake up in reality 150 years later. He chooses the latter, which agrees with Nozick's conclusion that people tend to value reality more than fake experiences even if the latter can bring us pleasure and joy.

Nozick's argument focuses more on the decision to plug into the machine than to unplug from it. The second scenario has been used as a counterargument to Nozick's argument by the experimental psychologist Felipe De Brigard (2010). De Brigard's argument is the reverse version of the experience machine: suppose one day you are told that you have been living in a simulated reality because you have been plugged into the experience machine, will you choose to remain plugged or wake up in the actual world, which you know nothing about? This scenario is roughly the same as that confronting César in the film *Open Your Eyes*. De Brigard claims that most people would remain plugged because people have a *status quo bias*, that is, a preference for the current state of affairs. Such being the case, Nozick's intuition that most people would resist plugging into the experience machine would be wrong. His attack on hedonism thus fails.

Although *Open Your Eyes* focuses more on the unplugging from the experience machine, it would be more convincing to interpret the film as implicating Nozick's argument, rather than its reverse version. The reason is simple. The *apparent* conclusion of the film simply disagrees with De Brigard's claim.[93] It appears to me that the film engages less with César's

---

[93] I say "apparent" because that interpretation stands out most compared to others. There are clues indicating a different interpretation but the support is weak. Therefore I disagree with Engelen's alternative interpretation of the film. See his 2010, pp. 37-38.

decision to sign the *Lucid Dream* contract simply due to dramatic effects: the contract-signing part has to be concealed in order for the ending of the film to come as a real surprise; moreover, the decision to unplug presents a more powerful dramatic moment than that to plug. (Compare the scene in *The Matrix* where the main character Neo did not choose to be plugged in.) The film thus appears more coherent when interpreted along Nozick's line of thinking.

Now, suppose my interpretation is at least plausible, if not *true*. A question to ask is: how applicable is the hypothetical intentionalist's recommendation in this case? It seems to me that, should I have adopted hypothetical intentionalism, I would have concluded that the above interpretation is not the best hypothesis of the director's intention. This is because the public evidence I obtained—especially the information about the film and the director on major websites such as Wiki and IMDb—reveals no connection between the director and philosophy backgrounds or interests. Neither does the public evidence confirm that Alejandro Amenábar studied philosophy (instead, the director's short biography on IMDb says he studied cinema), nor does it say that he entertained philosophical intentions when making this film. No evidence I have found shows that Amenábar has ever read Nozick. All the references to Nozick in this film were found in the *interpretations* of it.

Adopting hypothetical intentionalism, I would think that presenting or dramatizing the experience machine argument is not the best hypothesis of Amenábar's intention. Rather, it is more likely that he was influenced by sci-fi fiction involving simulated reality or by external world scepticism, both of which had long been there before 1997. It is true that Amenábar could have read Nozick, since the film came out in 1997 and the experience machine argument was published in 1974. But that hypothesis does not seem to be the best based on publicly available evidence.

Compare the case of *The Matrix*. The film has explicit philosophical references. For example, in one scene, a copy of Jean Baudrillard's book *Simulacra and Simulation* is used to conceal disks; moreover, one of the phrases in the book is quoted by Morpheus, the subordinate protagonist. These are glaring clues that the director might have intended the film to convey Baudrillard's philosophical ideas. That could be the best hypothesis of the director's intention with respect to the theme of the film. This hypothesis is further supported by contextual information: in the documentary *The Matrix Revisited* Keanu Reeves, who plays the character Neo, reveals that he was required by the directors to read *Simulacra and Simulation*. However, this kind of evidence does not seem to exist in the case of *Open Your*

Eyes, which makes it less plausible to say that presenting the experience machine argument is the best hypothesis about the director's intention.[94]

Why hypothetical authorism does a better job than hypothetical intentionalism in balancing interpretative freedom and constraint is now apparent. In the case of *Open Your Eyes*, the interpreter following either position utilizes the same interpretative resources, that is, internal evidence plus public information about the work and the director. Both positions, then, constrain the interpretation of the work in the same manner. Nevertheless, the role played by the reader in the story told by hypothetical intentionalism is actually subordinate, for the reader's interpretation in this case is, in the abstract, still a reconstruction of the actual author. By contrast, the interpreter adopting hypothetical authorism does not reconstruct anything; rather, she constructs an author from the work. It is in this sense that the hypothetical authorist enjoys more interpretative freedom than her kin, while constrained by the same interpretative data.

7.2.5 Section conclusion

In this section I have offered some arguments showing that the major contenders in the debate are less convincing than hypothetical authorism as the best interpretative recommendation. Versions of anti-intentionalism are unappealing because they underestimate the importance of the concept of an author; actual intentionalism is even less appealing because of its stringent demands on the interpreter; finally, hypothetical intentionalism fails to give as much interpretative autonomy as hypothetical authorism and hence is a less successful middle course between actual intentionalism and anti-intentionalism. All the weaknesses mentioned above are not to be found in hypothetical authorism. I thus recommend this position to the reader as the best interpretative policy.

---

[94] Note that a hypothetical authorist without a philosophy background may not interpret the film in the same way as I do. As said, different backgrounds and value concerns generate different implied authors. The hypothetical authorist without philosophy training might well arrive at the interpretation that coincides with that made by the hypothetical intentionalist in the case of *Open Your Eyes*. However, this result does not conflict with my argument in this section.

Conclusion

I conclude this thesis by making two observations and suggesting two interesting follow-up directions.

From my discussion we can notice two major trends in the debate. First, most participants are committed to the contextualist ontology of literature (and perhaps the arts in general). The textualist or formalist conception of art and literature is less followed nowadays. The relevance of art's historical context, since its first philosophical appearance in Arthur Danto's 1964 essay (Danto, 1964), continues to influence analytic theories of literary interpretation. There is no sign of this trend diminishing; in Carroll's most recent survey article on interpretation, the contextualist basis is still assumed (Carroll, 2016). The underlying assumption that a literary work is an utterance is only occasionally challenged (by, as discussed, Goldman, Lamarque and Olsen), and probably will continue to be accepted for quite a while.

Second, actual intentionalism remains the most popular position among all. The most representative proponents—Carroll, Livingston, and Stecker—have all produced substantial monographs in this century to defend the position (Stecker, 2003; Livingston, 2005a; Carroll, 2009). This intentionalist prevalence probably results from Grice's influence on the philosophy of literature. And again, this trend, like the contextualist vogue, is still ongoing. And if we see intentionalism as an umbrella term that encompasses not only actual intentionalism, but also hypothetical intentionalism and probably hypothetical authorism, the influence of intentionalism and its related emphasis on the concept of an author will be even stronger. This presents an interesting contrast with the trend in Continental philosophy that tends to downplay authorial presence in theories of interpretation.

There are a couple of interesting directions worth a brief look. As we can see, analytic theorists of interpretation tend to produce a wholesale account that is to be applied to *all* kinds of literary works, indifferent to literary categories. (The only distinction ever made seems to be the divide between fiction and nonfiction, with the emphasis on the former.) But some people have thought it plausible to devise different theories for different literary genres or categories. For example, Jukka Mikkonen holds that actual intentionalism is the best interpretative policy for philosophical fiction (Mikkonen, 2010). This is because this kind of fiction is typically a literary presentation of the author's philosophical ideas; without knowing the author's philosophical views, the reader might not be able to understand the work

properly. This line of thinking pays attention to the subtle differences between genres and advocates that theories of interpretation should be tailor-made. In fact, as early as 1993, similar thoughts were proposed by Berys Gaut, who offered a patchwork theory: "[…] *no* global theory of interpretation, intentionalist or otherwise, is correct. Instead, one needs a series of local theories of interpretation." (Gaut, 1993, p. 597, emphasis original)

Another interesting (but largely neglected) line of thinking draws our intention not to different literary categories, but to different levels of meaning. As briefly mentioned in §2.2.3, Christopher New recommends that we distinguish four levels of meaning. The most basic level is the vocabulary level at which we establish the correct text for interpretation; the second level is the locutionary level that deals with word or sentence meaning; the third level is the illocutionary level that deals with the force (or the purported force if we treat fictional works as representations of illocutions) borne by the text; the final level is the suprasentential level at which we consider what is implied by the work viewed as a linguistic whole. There is reason to think that the author's intention is relevant to different degrees at these different levels; therefore, a global theory of interpretation indifferent to different levels of meaning is an oversimplification.

These two directions are often underestimated but I believe they would shed new light on the debate. Exploring them awaits another occasion, and my conclusion at this stage remains that hypothetical authorism is the ideal policy for interpretation.

List of References

Austin, J. L. (1962). *How to Do Things with Words.* Oxford: Clarendon Press.

Auxier, R. (2014). Virtual Intentionalism or Actual Intentionalism? *Sztuka i Filozofia, 44*, 34-36.

Beardsley, M. C. (1953). Interpretation. In J. T. Shipley (Ed.), *Dictionary Of World Literature* (pp. 232-233). New York: Philosophical Library.

Beardsley, M. C. (1970). *The Possibility of Criticism.* Detroit, MI: Wayne State University Press.

Beardsley, M. C. (1973). The Concept of Literature. In F. Brady (Ed.), *Literary Theory and Structure: Essays in Honor of Willaim K. Wimsatt* (pp. 23-39). New Haven, CT: Yale University Press.

Beardsley, M. C. (1977). The Philosophy of Literature. In G. Dickie, R. Sclafani, & R. Roblin (Eds.), *Aesthetics: A Critical Anthology* (pp. 420-430). New York: St. Martin's Press.

Beardsley, M. C. (1978a). Intending. In G. A. I., & J. Kim (Eds.), *Values and Morals* (pp. 163-184). Dordrecht: Springer.

Beardsley, M. C. (1978b). Aesthetic Intentions and Fictive Illocutions. In P. Hernadi (Ed.), *What is Literature* (pp. 161-177). Bloomington, IN: Indiana University Press.

Beardsley, M. C. (1979). Verbal Style and Illocutionary Action. In B. Lang (Ed.), *The Concept of Style* (pp. 149-168). Philadelphia: The University of Pennsylvania Press.

Beardsley, M. C. (1981a). *Aesthetics, Problems in the Philosophy of Criticism* (2 ed.). Indianapolis, IN: Hackett.

Beardsley, M. C. (1981b). Fiction as Representation. *Synthese, 46*, 291-313.

Beardsley, M. C. (1982). Intentions and Interpretations: A Fallacy Revived. In M. J. Wreen, & D. M. Callen (Eds.), *The Aesthetic Point of View: Selected Essays* (pp. 188-207). Ithaca, NY: Cornell University Press.

Booth, W. C. (1983). *The Rhetoric of Fiction* (2 ed.). Chicago: University of Chicago Press.

Carroll, N. (1992). Art, Intentions, and Conversations. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 97-131). Philadelphia, PA: Temple University Press.

Carroll, N. (1993). Anglo-American Aesthetics and Contemporary Criticism: Intention and the Hermeneutics of Suspicion. *Journal of Aesthetics and Art Criticism, 51*(2), 245-252.

Carroll, N. (1997). The Intentional Fallacy: Defending Myself. *Journal of Aesthetics and Art Criticism, 55*(3), 306-311.

Carroll, N. (2000). Interpretation and Intention: The Debate between Hypothetical and Actual Intentionalism. *Metaphilosophy, 31*(1-2), 75-95.

Carroll, N. (2002). Andy Kaufman and the Philosophy of Interpretation. In M. Krausz (Ed.), *Is There a Single Right Interpretation?* (pp. 319-344). University Park: Pennsylvania State University Press.

Carroll, N. (2009). *On Criticism.* New York: Routledge.

Carroll, N. (2011). Art Interpretation. *British Journal of Aesthetics, 51*(2), 117-135.

Carroll, N. (2013). Criticism and Interpretation. *Sztuka i Filozofia, 42*, 7-20.

Carroll, N. (2014). Criticism and Interpretation Redux: Responses to My Commentators. *Sztuka i Filozofia, 44*, 37-45.

Carroll, N. (2016). Interpretation. In N. Carroll, & J. Gibson (Eds.), *The Routledge Companion to Philosophy of Literature* (pp. 302-312). New York: Routledge.

Currie, G. (1990). *The Nature of Fiction.* Cambridge, England: Cambridge University Press.

Currie, G. (1991a). Work and Text. *Mind, 100*(399), 325-340.

Currie, G. (1991b). Interpreting Fiction. In R. Freadman, & L. Reinhardt (Eds.), *On Literary Theory and Philosophy* (pp. 96-112). London: Macmillan.

Currie, G. (1993). Interpretation and Objectivity. *Mind, 102*(407), 413-428.

Currie, G. (2005). Interpretation in Art. In J. Levinson (Ed.), *The Oxford Handbook of Aesthetics* (pp. 291-306). Oxford: Oxford University Press.

Danto, A. C. (1964). The Artworld. *Journal of Philosophy, 61*, 571-584.

Davies, D. (2007). *Aesthetics & Literature.* London: Continuum.

Davies, S. (1982). The Aesthetic Relevance of Authors' and Painters' Intentions. *Journal of Aesthetics and Art Criticism, 41*(1), 65-76.

Davies, S. (1991). *Definitions of Art.* Ithaca, N.Y.: Cornell University Press.

Davies, S. (1996). Interpreting Contextualities. *Philosophy and Literature, 20*(1), 20-38.

Davies, S. (2005). Beardsley and the Autonomy of the Work of Art. *Journal of Aesthetics and Art Criticism, 63*(2), 179-183.

Davies, S. (2006). *The Philosophy of Art.* Oxford: Blackwell.

Davies, S. (2007). *Philosophical Perspectives on Art.* Oxford: Oxford University Press.

Davies, S. (2014). Modest Actual Mentalism: Questions and Comments. *Sztuka i Filozofia, 44*, 13-17.

De Brigard, F. (2010). If You Like It, Does It Matter If It's Real? *Philosophical Psychology, 23*(1), 43-57.

Dickie, G. (2005). Beardsley, Monroe. In D. M. Borchert (Ed.), *Encyclopedia of Philosophy* (2 ed., Vol. 1, pp. 508-510). Detroit, MI: Macmillan Reference.

Dickie, G. (2006). Intentions: Conversations and Art. *British Jpurnal of Aesthetics, 46*(1), 71-81.

Dickie, G., & Wilson, W. K. (1995). The Intentional Fallacy: Defending Beardsley. *Journal of Aesthetics and Art Criticism, 53*(3), 234-250.

Engelen, B. (2010). Open Your Eyes? Why Nozick's Experience Machine Does Not Refute Hedonism. In D. Shaw (Ed.), *Film and Philosophy* (General Interest ed., Vol. 14, pp. 33-46).

Freeland, C. (2014). Abductive Reasoning, Intentionalism and Meanings in Artwork. *Sztuka i Filozofia, 44*, 18-24.

Gaut, B. (1993). Interpreting the Arts: The Patchwork Theory. *Journal of Aesthetics and Art Criticism, 51*(4), 597-609.

Goldman, A. H. (1991). Response to Stecker. *Journal of Aesthetics and Art Criticism, 49*(3), 246-247.

Goldman, A. H. (1995). *Aesthetic Value.* Boulder: Westview Press.

Goldman, A. H. (2013). *Philosophy and the Novel.* Oxford: Oxford University Press.

Goodman, N., & Elgin, C. Z. (1986). Interpretation and Identity: Can the Work Survive the World? *Critical Inquiry, 12*(3), 564-575.

Gracyk, T. (2009). Ontological Contexualism. In S. Davies, K. M. Higgins, R. Hopkins, R. Stecker, & D. E. Cooper (Eds.), *A Companion to Aesthetics* (pp. 449-453). Malden: Blackwell.

Grice, P. (1957). Meaning. *Philosophical Review, 66*(3), 377-388.

Hirsch, E. D. (1967). *Validity in Interpretation.* New Haven: Yale University Press.

Hirsch, E. D. (1976). *The Aims of Interpretation.* Chicago, IL: University of Chicago Press.

Huddleston, A. (2012). The Conversation Argument for Actual Intentionalism. *British Journal of Aesthetics, 52*(3), 241-256.

Irvin, S. (2006). Authors, Intentions and Literary Meaning. *Philosophy Compass, 1*(2), 114-128.

Iseminger, G. (1992). An Intentional Demonstration? In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 76-96). Philadelphia, PA: Temple University Press.

Iseminger, G. (1996). Actual Intentionalism vs. Hypothetical Intentionalism. *Journal of Aesthetics and Art Criticism, 54*(4), 319-326.

Iseminger, G. (1998). Interpretive Relevance, Contradiction, and Compatibility with the Text: A Rejoinder to Knight. *Journal of Aesthetics and Art Criticism, 56*(1), 58-61.

Jannotta, A. (2014). Interpretation and Conversation: A Response to Huddleston. *British Journal of Aesthetics, 54*(3), 371-380.

Kindt, T., & Müller, H.-H. (2006). *The Implied Author: Concept and Controversy.* New York: de Gruyter.

Knapp, S., & Michaels, W. B. (1982). Against Theory. *Critical Inquiry, 8*(4), 790-800.

Knapp, S., & Michaels, W. B. (1985). A Reply to Our Critics. In W. J. Mitchell (Ed.), *Against Theory: Literary Studies and the New Pragmatism* (pp. 95-105). Chicago: University of Chicago Press.

Knapp, S., & Michaels, W. B. (1987). Against Theory 2: Hermeneutics and Deconstruction. *Critical Inquiry, 14*(1), 49-68.

Knapp, S., & Michaels, W. B. (1992). The Impossibility of Intentionless Meaning. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 51-64). Philadelphia, PA: Temple University Press.

Laetz, B. (2010). Kendall Walton's 'Categories of Art': A Critical Commentary. *British Journal of Aesthetics, 50*(3), 287-306.

Lamarque, P. (2002). Appreciation and Literary Interpretation. In M. Krausz (Ed.), *Is There a Single Right Interpretation?* (pp. 285-306). University Park: Pennsylvania State University Press.

Lamarque, P. (2007). *The Philosophy of Literature.* Malden, MA: Blackwell.

Levinson, J. (1985). Titles. *Journal of Aesthetics and Art Criticism, 44*(1), 29-39.

Levinson, J. (1992). Intention and Interpretation: A Last Look. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 221-256). Philadelphia, PA: Temple University Press.

Levinson, J. (1995). Messages in Art. *Australasian Journal of Philosophy, 73*(2), 184-198.

Levinson, J. (1996). *The Pleasures of Aesthetics : Philosophical Essays.* Ithaca, NY: Cornell University Press.

Levinson, J. (2006). *Contemplating Art: Essays in Aesthetics.* Oxford: Oxford University Press.

Levinson, J. (2010). Defending Hypothetical Intentionalism. *British Journal of Aesthetics, 50*(2), 139-150.

Levinson, J. (2016). *Aesthetic Pursuits: Essays in Philosophy of Art.* Oxford: Oxford University Press.

Lian, X. (2008). Ci Poetry: Long Song Lyrics (Manci). In Z.-Q. Cai (Ed.), *How to Read Chinese Poetry: A Guided Anthology* (pp. 262-285). New York: Columbia University Press.

Lin, S.-Y. (2016). Beardsley on Literature, Fiction, and Nonfiction. *Journal of Aesthetics & Culture, 8*. doi:10.3402/jac.v8.29208

Lin, S.-Y. (2018). Beardsley and the Implied Author. *Journal of Literary Theory* 12(1), 170-192.

Lin, S.-Y. (forthcoming). Beardsley's Contextualism: Philosophical and Educational Significance. *Journal of Aesthetic Education*.

Lin, S.-Y. (accepted). Interpretation and the Implied Author: A Descriptive Project. *Southern Journal of Philosophy*.

Lintott, S. (2002). When Artists Fail: A Reply to Trivedi. *British Journal of Aesthetics, 42*(1), 64-72.

Livingston, P. (1996). Arguing over Intentions. *Revue Internationale de Philosophie, 198*(4), 615-633.

Livingston, P. (1998). Intentionalism in Aesthetics. *New Literary History, 29*(4), 831-846.

Livingston, P. (2005a). *Art and Intention: A Philosophical Study.* Oxford: Oxford University Press.

Livingston, P. (2005b). Intention in Art. In J. Levinson (Ed.), *The Oxford Handbook of Aesthetics* (pp. 275-290). Oxford: Oxford University Press.

Livingston, P. (2009). *Cinema, Philosophy, Bergman: On Film as Philosophy.* Oxford: Oxford University Press.

Livingston, P. (2010). Authorial Intention and the Varieties of Intentionalism. In G. L. Hagberg, & W. Jost (Eds.), *A Companion to the Philosophy of Literature* (pp. 401-419). Malden, MA: Wiley-Blackwell.

Lyas, C. (1992). Wittgensteinian Intentions. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 132-151). Philadelphia, PA: Temple University Press.

Lyas, C., & Stecker, R. (2009). Intention and Interpretation. In S. Davies, K. M. Higgins, R. Hopkins, R. Stecker, & D. E. Cooper (Eds.), *A Companion to Aesthetics* (pp. 366-369). Malden: Blackwell.

Margolis, J. (1979). Literature and Speech Acts. *Philosophy and Literature, 3*(1), 39-52.

Meiland, J. W. (1978). Interpretation as a Cognitive Discipline. *Philosophy and Literature, 2*(1), 23-45.

Meiland, J. W. (1981). The Meanings of a Text. *British Journal of Aesthetics, 21*(3), 195-203.

Mikkonen, J. (2010). Literary Fictions as utterances and Artworks. *Theoria, 76*, 68-90.

Mitchell, W. J. (Ed.). (1985). *Against Theory: Literary Studies and the New Pragmatism.* Chicago: University of Chicago Press.

Nathan, D. O. (1973). Categories and Intentions. *Journal of Aesthetics and Art Criticism, 31*(4), 539-541.

Nathan, D. O. (1982). Irony and the Artist's Intentions. *Journal of Aesthetics and Art Criticism, 22*(3), 245-256.

Nathan, D. O. (1992). Irony, Metaphor, and the Problem of Intention. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 183-202). Philadelphia, PA: Temple University Press.

Nathan, D. O. (2005). A Paradox in Intentionalism. *British Journal of Aesthetics, 45*(1), 32-48.

Nathan, D. O. (2006). Art, Meaning, and Artist's Meaning. In M. Kieran (Ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art* (pp. 282-293). Oxford: Blackwell.

Nehamas, A. (1981). The Postulated Author: Critical Monism as a Regulative Ideal. *Critical Inquiry, 8*(1), 133-149.

Nehamas, A. (1986). What an Author Is. *Journal of Philosophy, 83*(11), 685-691.

Nehamas, A. (2002). Writer, Text, Work, Author. In W. Irwin (Ed.), *The Death and Resurrection of the Author?* (pp. 95-115). Westport, CT: Greenwood Press.

New, C. (1999). *Philosophy of Literature: An Introduction.* London: Routledge.

Nozick, R. (1974). *Anarchy, State, and Utopia.* New York: Basic Books.

Olsen, S. H. (1982). The "Meaning" of a Literary Work. *New Literary History, 14*(1), 13-32.

Olson, E. T. (2015). Personal Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Retrieved from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/archives/sum2017/entries/identity-personal/

Percival, P. (2002). Can Novel Critical Interpretation Create Art Objects Distinct from Themselves? In M. Krausz (Ed.). University Park: Pennsylvania State University Press.

Robinson, J. M. (1985). Style and Personality in the Literary Work. *Philosophical Review, 94*(2), 227-247.

Savile, A. (1996). Instrumentalism and the Interpretation of Narrative. *Mind, 105*(420), 553-576.

Shusterman, R. (1992). Interpreting with Pragmatist Intentions. In G. Iseminger (Ed.), *Intention & Interpretation* (pp. 167-182). Philadelphia, PA: Temple University Press.

Shusterman, R. (2009). Text. In S. Davies, K. M. Higgins, R. Hopkins, R. Stecker, & D. E. Cooper (Eds.), *A Companion to Aesthetics* (pp. 565-562). Malden, MA: Blackwell.

Sirridge, M. (1978). Artistic Intention and Critical Prerogative. *British Journal of Aesthetics, 18*(2), 137-154.

Spoerhase, C. (2007). Hypothetischer Intentionalismus. Rekonstruktion und Kritik. *Journal of Literary Theory, 1*(1), 81-110.

Stecker, R. (1987). Apparent, Implied, and Postulated Authors. *Philosophy and Literature, 11*(2), 258-271.

Stecker, R. (1991). Goldman on Interpreting Art and Literature. *Journal of Aesthetics and Art Criticism, 49*(3), 243-246.

Stecker, R. (1993). The Role of Intention and Convention in Interpreting Artworks. *The Southern Journal Of Philosophy, 31*(4), 471-489.

Stecker, R. (1994). Art Interpretation. *Journal of Aesthetics and Art Criticism, 52*(2), 193-206.

Stecker, R. (1997). *Artworks: Definition, Meaning, Value* . University Park, Pa.: Pennsylvania State University Press.

Stecker, R. (2003). *Interpretation and Construction: Art, Speech, and the Law.* Oxford: Blackwell.

Stecker, R. (2006). Moderate Actual Intentionalism Defended. *Journal of Aesthetics and Art Criticism, 64*(4), 429-438.

Stecker, R. (2008). Intention and Interpretation. *Journal of Literary Theory, 2*(1), 35-50.

Stecker, R. (2010). *Aesthetics and the Philosophy of Art : An Introduction.* Lanham: Rowman & Littlefield.

Stecker, R. (2013). Interpretation. In B. Gaut, & D. M. Lopes (Eds.), *The Routledge Companion to Aesthetics* (3 ed., pp. 309-319). New York: Routledge.

Stecker, R., & Davies, S. (2010). The Hypothetical Intentionalist's Dilemma: A Reply to Levinson. *British Journal of Aesthetics, 50*(3), 307-312.

Tolhurst, W. E. (1979). On What a Text is and How it Means. *British Journal of Aesthetics, 19*(1), 3-14.

Tolhurst, W. E., & Wheeler III, S. C. (1979). On Textual Individuation. *Philosophical Studies, 35*(2), 187-197.

Trivedi, S. (2001). An Epistemic Dilemma for Actual Intentionalism. *British Journal of Aesthetics, 41*(2), 192-206.

Trivedi, S. (2004). Artist-Audience Communication: Tolstoy Reclaimed. *Journal of Aesthetic Education, 38*(2), 38-52.

Trivedi, S. (2015). Surplus, Authorial Intentions, and Hypothetical Intentionalism. *College Literature, 42*(4), 699-724.

Walton, K. L. (1970). Categories of Art. *Philosophical Review, 79*(3), 334-367.

Walton, K. L. (1973). Categories and Intentions: A Reply. *Journal of Aesthetics and Art Criticism, 32*(2), 267-268.

Walton, K. L. (1976). Points of View in Narrative and Depictive Representation. *Noûs, 10*(1), 49-61.

Walton, K. L. (1990). *Mimesis as Make-Believe.* Cambridge, Mass.: Harvard University Press .

Walton, K. L. (2008). *Marvelous Images: On Values and the Arts.* Oxford: Oxford University Press.

Wilson, W. K. (1997). Confession of a Weak Anti-Intentionalist: Exposing Myself. *Journal of Aesthetics and Art Criticism, 55*(3), 309-311.

Wimsatt, W. K., & Beardsley, M. C. (1946). The Intentional Fallacy. *The Sewanee Review, 54*(3), 468-488.

Wreen, M. (2014). Beardsley's Aesthetics. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Retrieved from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/beardsley-aesthetics/