

Disagreeing about fiction

Louis Rouillé

▶ **To cite this version:**

Louis Rouillé. Disagreeing about fiction. Philosophie. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLEE064 . tel-03137346

HAL Id: tel-03137346
https://theses.hal.science/tel-03137346

Submitted on 10 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Institut Jean Nicod, UMR 8129

Disagreeing about fiction

Soutenue par

Louis ROUILLÉ

Le 02 12 2019

École doctorale n°540

Lettres, Arts, Sciences humaines et sociales

Spécialité

Philosophie

Composition du jury :

Stacie Friend Senior Lecturer in Philosophy, Birkbeck, University of London	*Présidente*
Françoise Lavocat Professeure de littérature comparée, Université Paris 3-Sorbonne nouvelle	*Rapportrice*
Emar Maier Assistant Professor Philosophy & Linguistics, University of Groningen	*Rapporteur*
Anne Reboul Directrice de recherche au CNRS, Institut Marc Jeannerod, UMR 5304	*Examinatrice*
Paul Égré Directeur de recherche au CNRS, Institut Jean Nicod, UMR 8129	*Directeur de thèse*
François Récanati Professeur au Collège de France	*Directeur de thèse*

Institut | Nicod
CNRS-EHESS-ENS

ENS DEC DÉPARTEMENT D'ÉTUDES COGNITIVES

Contents

Acknowledgments **11**

Introduction: Is it really a philosophical problem? **14**
Fiction as a metaphilosophical problem . . . 14
Parmenides, the Visitor and Plato's problem . . . 16
The problem of non-being (or the non-problem of being) . . . 16
Conceptual reconstruction of Plato's problem . . . 20
Conceptual re-reconstruction . . . 24
Brentano's problem and program . . . 24
Brentano's program . . . 28
The metaphysics of relations . . . 29
Consequences for the problem of fiction . . . 31
Content and structure of the dissertation . . . 33

I Truth in fiction: towards pretence semantics **36**

1 The problem **37**
1.1 Description of the problem . . . 37
1.2 Non-starters . . . 39
1.3 Explicitism and Intentionalisms . . . 40
1.3.1 Against explicitism . . . 41
1.3.2 Intentionalisms . . . 44
1.4 Two incompatible solutions on the market . . . 46
1.4.1 The modal account . . . 46
1.4.2 The functional account . . . 47
1.4.3 The two solutions are incompatible . . . 48
1.5 Plan of action . . . 50

2 On modally minded theorists **52**
2.1 Lewis on truth in fiction . . . 52
2.1.1 Lewis's theory of counterfactuals . . . 52
2.1.2 The counterfactual analysis of fictional statements . . . 53
2.2 On the kinds of counterarguments to the modal account . . . 56
2.2.1 Some terminology . . . 56
2.2.2 Indirect and direct arguments . . . 56
2.3 Aside on the literary minded modal theorists . . . 58

3 Indirect arguments against the modal account **60**
3.1 On the two kinds of modal theorists in the literature . . . 60
3.2 Dual solutions for dual problems . . . 61
3.2.1 Incompleteness . . . 61
3.2.2 Inconsistency . . . 62
3.2.3 Dual problems . . . 63
3.3 Doing the impossible with possible worlds . . . 64
3.3.1 Supervaluation . . . 66
3.3.2 Subvaluation . . . 66
3.3.3 Combining the two strategies . . . 67
3.3.4 The revenge of inconsistent fictions . . . 67
3.4 Impossible worlds and their limitations . . . 71
3.4.1 The unrestricted non-normal world view . . . 72
3.4.2 Berto's thesis . . . 73
3.4.3 A diagonal argument against the unrestricted non-normal world account . . . 74
3.5 Conclusions . . . 76

4 Direct arguments against the modal account **79**
4.1 Plan of action . . . 79
4.1.1 The modal intuition . . . 80
4.1.2 Two suppositions of the modal account . . . 80
4.1.3 The plot . . . 81
4.2 Fiction is not just like the other modalities . . . 82
4.2.1 Structure of the argument . . . 82
4.2.2 Non-P-modalities . . . 83
4.2.3 On what makes a *fictional* world a fictional *world* . . . 85
4.3 Evans against Lewis . . . 89
4.3.1 Evans's dazzling paragraph . . . 89

4.3.2 Explanation of Evans's argument . . . 91
4.3.3 Conclusions drawn from Evans . . . 96

5 Walton's programmatic ideas 99
5.1 Basic definitions . . . 100
5.1.1 Props . . . 100
5.1.2 Principles of generation . . . 101
5.2 Basic distinctions . . . 102
5.2.1 Imagination and fictionality . . . 102
5.2.2 Fictionality and truth . . . 104
5.3 Towards pretence semantics . . . 105
5.3.1 Normativity and fictionality . . . 105
5.3.2 Fictionality conditions . . . 107

6 Pretence semantics 108
6.1 Everett's functionalism . . . 108
6.2 Simple principles of generation . . . 110
6.2.1 Formalizing a simple pretence . . . 110
6.2.2 On distinguishing between (T1) and (T2) . . . 111
6.2.3 Technical points . . . 115
6.3 Elaborate principles of generation . . . 119
6.3.1 The Reality principles . . . 120
6.3.2 The Mutual Belief principles . . . 122
6.3.3 Other elaborate principles? . . . 123
6.4 Conclusions . . . 127

7 From pretence to fiction 129
7.1 In search for the fiction principle . . . 130
7.1.1 Genesis of the report principle . . . 130
7.1.2 The report principle . . . 131
7.1.3 Strong and Weak . . . 134
7.2 More on the secondary pretence . . . 139
7.2.1 A toy example . . . 139
7.2.2 A linguistic puzzle explained . . . 140
7.2.3 Two narratological puzzles explained . . . 141
7.3 The revenge of the reporting-the-unreported problem . . . 143
7.3.1 Virginia Woolf and the report principle . . . 143
7.3.2 Woolf-like counterexamples . . . 145
7.3.3 The appearance principle . . . 146

7.4 Conclusions 148

8 Readers as fictional truth-makers? 150
8.1 Fictions without authors 151
8.2 When readers are authors 152
8.3 A legal prequel for future fan-fictions 153
8.4 Open-ending 156

Appendices 157

A "Truth in fiction" as an unsupervised learning task 157

II The great beetle debate 163

9 Debates over disagreements 164
9.1 A logical space for disagreements 164
9.1.1 Faultless disagreements 164
9.1.2 Fictional disagreements 165
9.1.3 The logical space 167
9.2 The Great Beetle Debate 168
9.2.1 Todorov's definition of the fantastic 168
9.2.2 Application of Todorov's definition to Kafka 171
9.2.3 Nabokov's comment on *the Metamorphosis* 173

10 Imagining past each other 176
10.1 Prototypes and fictional imagination 177
10.1.1 The average reader's prototypical vermin 179
10.1.2 Nabokov's prototypical vermin 181
10.2 Imaginative resistance 183
10.2.1 The GBD as a case of descriptive imaginative resistance . . . 184
10.2.2 Interpreting imaginative resistance 186
10.3 Conclusions 188

11 The Great Beetle Debate as a faultless disagreement 190
11.1 Mistaken readers 190
11.1.1 Grading erroneous readers 190
11.1.2 Why are the Smiths so wrong? 192
11.2 Saving Private Smith_3 194

11.2.1 The way of the psychologist . . . 195
11.2.2 The way of the sceptic . . . 195
11.2.3 The way of the ignorant . . . 197
11.2.4 The way of the erudite person . . . 197
11.3 Faultless disagreements and literary disagreements . . . 199
11.3.1 The GBD is faultless . . . 199
11.3.2 Interpretation and elucidation . . . 200
11.3.3 The GBD as an interesting case of elucidation . . . 202
11.3.4 Pluralism about elucidation . . . 204

12 The fundamental indeterminacy of fiction 206
12.1 Friend on indeterminacy . . . 207
12.2 Ontologically bizarre characters . . . 209
12.2.1 Definition . . . 209
12.2.2 More examples of ontologically bizarre characters . . . 210
12.3 On the indeterminacy of Gregor . . . 213
12.3.1 Naïve epistemology . . . 215
12.3.2 Naïve metaphysics . . . 215
12.4 Finding the indeterminacy on the pragmatic nature of fiction . . . 216
12.4.1 Two readings of the reality principle . . . 217
12.4.2 Quine's notion of indeterminacy . . . 219
12.4.3 Gregor is pragmatically indeterminate . . . 220
12.5 Conclusions . . . 220

13 Conditions of possibility of great debates 222
13.1 On the non-existence of a great Nose debate . . . 224
13.1.1 The fictional detail . . . 224
13.1.2 Pairing minimally Gregor and Kovalyov's nose . . . 224
13.1.3 A debate about Gogol's nose? . . . 226
13.1.4 Aside on the explication and thematic interpretation of *the Nose* 226
13.1.5 The difference, and a pseudo-great debate . . . 228
13.1.6 Is there even a disagreement about *the Nose*? . . . 229
13.2 Further considerations about fictional imagination . . . 230
13.2.1 Two kinds of fictional objects . . . 230
13.2.2 Literary debates and thought experiments . . . 232

III A problem about fictional names 237

14 Definitions and scope of the problem 238
14.1 Two orthogonal distinctions . . . 239
14.1.1 Names in fiction . . . 239
14.1.2 Fictional names and empty names . . . 239
14.2 Fictional and metafictional uses of fictional names . . . 240
14.3 How unique are metafictional uses? . . . 242
14.3.1 On a special feature of metafictional statements . . . 242
14.3.2 Complications . . . 245
14.4 Conclusions . . . 248

15 The debate between realism and anti-realism: a dead-end 251
15.1 Realism about fictional names . . . 252
15.1.1 Realism and metafictional sentences . . . 252
15.1.2 Realism and fictional sentences . . . 255
15.1.3 Conclusion about realism . . . 258
15.2 Anti-realism about fictional names . . . 258
15.2.1 Anti-realism and fictional sentences . . . 258
15.2.2 Anti-realism and metafictional sentences . . . 260
15.3 Conclusions . . . 264

16 Hybrid accounts of fictional names 265
16.1 Against the ambiguity view . . . 265
16.1.1 Presentation of the view . . . 265
16.1.2 The argument from renaming . . . 266
16.1.3 The argument from anaphora and co-predication . . . 267
16.1.4 Towards the polysemy view . . . 268
16.2 In arguing against the polysemy view . . . 268
16.2.1 Basic idea . . . 268
16.2.2 Detail of the view . . . 270
16.2.3 Argument against the polysemy view . . . 273
16.2.4 Counterproposal . . . 274
16.3 Conclusions . . . 275

17 Internal, external and mixed perspectives 277
17.1 Pseudo-assertions and assertions . . . 278
17.1.1 The Fregean origin . . . 278
17.1.2 Theorizing pseudo-assertion . . . 279

17.1.3 Shortcomings of the reduction . . . 282
17.2 Looking for a metaphysical perspective . . . 285
17.2.1 Narratological points of view . . . 286
17.2.2 Perspective in the visual arts . . . 287
17.2.3 Perspectives and conventions in the visual arts . . . 290
17.3 Metaphysical perspective in linguistic fictions . . . 296
17.3.1 External and internal perspectives . . . 296
17.3.2 Mixed perspectives . . . 297
17.4 A proposal for anti-realism . . . 299
17.4.1 Anti-realism and external perspective statements . . . 299
17.4.2 Focus on negative existentials . . . 300
17.5 Conclusions . . . 302

18 Freedom for Emma! 304
18.1 The problem with negative existentials . . . 304
18.1.1 The Naïve Analysis . . . 305
18.1.2 Classical Quantification Theory . . . 305
18.1.3 The problem . . . 307
18.2 Orthodoxy and heterodoxy on negative existentials . . . 307
18.2.1 Dropping NA: the logical orthodoxy . . . 307
18.2.2 Dropping CQT: the logical heterodoxy . . . 313
18.2.3 Down with the doxies! . . . 317
18.3 Free existence . . . 318
18.3.1 The core idea of free logics . . . 318
18.3.2 What is singular existence . . . 319
18.4 Towards free pretence semantics . . . 323
18.4.1 Choosing the right free semantics . . . 323
18.4.2 Story Semantics . . . 324
18.4.3 Combining pretence semantics and free semantics . . . 326
18.5 Conclusions . . . 329

Appendices 331

B Parafictional statements as metadata 331

C Pseudo-selves as virtual machines 334

Concluding remarks about anti-realism. Or the story of how some (sane) intuitions met some (insane) philosophical arguments in the second half of the 20th century **337**

Definitions . . . 337
Eliciting intuitions . . . 338
A personal anecdote and a joke . . . 338
Don't say no, you smiled . . . 339
Fictionalism is an Anti-realism . . . 339
Indispensability arguments . . . 340
The Varieties of Fictionalism . . . 341
Fictionalism about fiction . . . 342
Puzzling observations . . . 342
Intuitions and arguments . . . 343
Main philosophical characters in this story . . . 343

Résumé substantiel en français **346**

List of Figures

7.1 Double-structured pretence distinctive of fiction 137

9.1 Facsimile of Nabokov's lecture notes on *the Metamorphosis*. 174

10.1 *Die Verwandlung* (original cover) 182

11.1 Two spieces of cockroaches which mimic beetles 196

17.1 Stereographic projection of a sphere 288
17.2 One-point linear perspective with horizon 289
17.3 Velázquez 1656 *Las Meninas* (Museo del Prado, Madrid) 291
17.4 Drawing from Paul Klee's *Pädagogische Skizzenbuch* (1925) 292
17.5 Paul Klee 1929 *Main Path and Bypaths* (Museum Ludwig, Köln) (fragment of the original) 293
17.6 Duccio's inaccurate perspective compared to Leonoard's accurate one 295

C.1 Diagram of a virtual machine 334

List of Tables

9.1 A logical space for disagreements 168

17.1 Reduction of the distinction 282
17.2 Mixed perspectives and pseudo-reference 284
17.3 Mixed perspectives: true or fictional? 285
17.4 Final table of data: a continuum of mixed perspectives 299

Acknowledgments

Doing research taught me that the world is a space with incredibly many dimensions, which is a curse to all honest researchers. Fortunately, writing a dissertation involves mainly three dimensions: the institutional, the physical and the mental. I saw myself as a point in this three-dimensional space, as a mind, a body and a citizen, at the intersection of three orthogonal planes. The bases from which I could write this dissertation were three different people. On the institutional plane, there was Paul Égré; on the physical plane, there was Benoît Bezzina; on the mental plane, there was Dr Dardashti. They simply did their job and they were good to me. I see my crossing paths with them as a gift. Although I could never give them as much, I take this opportunity to express my deepest gratitude.

It seems that François Récanati and I have been constantly disagreeing about fiction in the past 4 years. The title of this dissertation is a tribute to all the discussions we had. I mostly remember the good laughs when trying to disentangle ideas and theories. Thanks for the whole training.

The title comes for Stacie Friend's imagination: thanks! I think of Stacie as the ideal reader, although I really fall short of the ideal author. Along with Emar Maier, thanks for your kindness and interest. You were the best comité de suivi de thèse ever.

At the IJN and LSCP, I could not possibly thank individually all the people with whom I had stimulating discussions. First, thanks to the "junior language reading group" participants and especially to Mora, Milica, Andrés and Amir for being the beating heart in this group. (I now realise I am the last to defend my dissertation!)[1] I learned a lot in a very nice atmosphere during these sessions. I also learnt a lot from the Linguae group activities: Emmanuel Chemla, Philippe Schlenker, Salvador Mascarenas, Benjamin Spector and Paul Égré were model researchers to me. Special thanks to Géraldine Carranante who was a philosophical comrade during these years at the IJN: her philosophical conversations were always friendly and her friendly

[1] I eventually got ahead of Amir by one day !

chats were always philosophical. Thanks also to Enrico Terrone, Philippo Contesi, Michael Murez, Tristan Thommen, Bianca, François Le Corre, Slawa, Dan Hoek, Mathieu Koroma, Diego and so many others to have made my stay at le 29 so nice day after day with a perfect balance of seriousness and non-seriousness.

At the philosophy department of the ENS, a big thank you to all my colleagues. Co-teaching with Agnès Grivaux was fantastic, long discussions with Dominique Lestel were wonderful, sharing ideas with Jean-Pascal Anfray was always precise and profound. Thanks to Rose-Hélène, Irlande Saurin, Mathias Girel, Dimitri El Murr, Dan Arbib and Marc Crépon for several important, informal discussions. Special thanks to Lucie Fabry who was my institutional twin sister and with whom I shared a lot! Many thanks to all the others and thanks for the informal working atmosphere which I personally enjoyed very much.

I was lucky enough to have great students at the ENS and a teaching atmosphere which was really ideal. Special thanks to all my students in the graduate courses I gave each spring. Thanks to Paul and the philosophy department for giving me priceless trust and unrestricted freedom of teach!

As for material contingency: thanks to the ED540 and the IJN for steady financial support for conferences and books; thanks to FrontCog for the contract extension which I badly needed; thanks to Paul and Emar's van Gogh grant for travel and the organisation of the workshop. Thanks also to the working women at day nursery Sainte-Lucie, for I would not have had time to write without them. The social utility of their job is without doubt, therefore I have nothing but contempt for the people in charge who think that funding childcare should be cut down. Thanks to my parents for financial support throughout.

On a personal level, a warm thank you to my family and friends for the joys of everyday life. Family goes in the two directions now, for when I was still trying to enumerate the nonexistent individuals in order to understand what I was working on, Marion was busy creating a now very much existent individual. I did not know you when I started, now we share everything, and Alvise is a good part of everything. Thanks for existing.

Special thanks to Philippe Ernoult who is the best teacher I had. I laughed so much when he taught us how to enjoy fiction and English literature, and so did we all. He then taught me many other things, like the good old common sense which is at the core of good literature, like the proverbial diamond in a crust of mud. I cannot imagine what would have happened, had he been my philosophy teacher...

Among others, a huge thanks to Olive, Hugo, Gé, Clo, Elvina, Solène, Alexis, Ayoub for being close friends. I also keep in mind all the other friends who weren't there throughout, but were clearly intentional objects to me, as Jaco would say!

Introduction: Is it really a philosophical problem?

Fiction as a metaphilosophical problem

> I hope you will consider what I arranged, but be skeptical of it.
>
> John Berger 1972, "Ways of Seeing", Episode 1

Writing a dissertation in philosophy about *fiction* puzzles many people. Most non-philosophers are genuinely surprised and ask: "is it really a philosophical problem?" The easiest response to such astonishment is to explain that the "problem of fiction" is a first cousin of the long standing philosophical problem of nonexistence, non-being or nothingness. At this point, an appeal to a known authority such as Plato or Sartre will convince anyone that non-being is indeed a problem, and a philosophical one. Hence, fiction, as a species of non-being, inherits the status: it is a philosophical problem.

But this is not a fair response. I think the *real* problems with fiction are local, relative as opposed to general, absolute theoretical problems. They are not as foundational as the problem of non-being. The non-philosopher's surprise can be made more subtle if one remarks that, among the species of nonexistent entities, fictional entities enjoy a peculiar status. Indeed, fictional entities seem to occupy many of our ordinary discourses and thoughts without apparent conflict or tension with real entities. This is obviously not the case of other nonexistent entities created by philosophers such as the round square, universals, Nothingness, etc. which are rarely in people's mind and often difficult to think about. Hence, a fairer response to the non-philosopher is to explain that what is puzzling about fictional discourse is the apparent *absence* of problem about it, given the fact that *there is* a long standing philosophical problem about nonexistence.

By contrast, philosophers have a very different reaction from the layman. My experience in informal discussions with fellow PhD candidates about one another's dissertations led me to meditate on what the Ugly would say to the Good: "There's two kinds of philosophers in this world my friend, those who see a problem with fiction and those who don't". Unfortunately, these informal discussions never seem to change anybody's mind and the latter group of philosophers always left me with the unpleasant impression that I was working on a non-issue. I thus came to wonder: Why do philosophers disagree about fiction to the point that they do not even agree that *there is a philosophical problem* about fiction? In trying to answer this question, I had to reflect on the relative nature of the concept of fiction. In this introduction, I will defend a metaphilosophical thesis which can be put in the following catchphrase: *Fiction is not a philosophical problem, but it is a problem for a kind of philosophy.* I will show in this introduction that the paradigmatic kind of philosophy which has

a problem with fictional discourse is so-called analytic philosophy, at least in its beginnings.

I do not think this metaphilosophical point can be made for all the notions which philosophers write their dissertations on. Indeed, if the key concept of your research is a notion like "truth" or "the self" or, for that matter, "nonexistence", I think no philosopher will doubt that you work on a substantial philosophical concept. Of course, they will find a way to disagree about the content of the theories you want to defend or criticise: that's what philosophers do. The contrast with the notion of fiction is thus striking. This suggests that the concept of fiction is related to other general theoretical commitments in a way these other concepts are not. I will argue for such a claim.

One might legitimately think that the origin of this fundamental disagreement among philosophers lies in the (lack of) definition of the term "fiction". As such, a conceptual clarification should make everyone see eye to eye. [Renauld 2013] contributes to this clarification. In her dissertation, she defends that there are rival definitions of fiction in the philosophical literature such that we should drop the term "fiction" altogether and use more precise idioms like "story", "narration", "representational arts", etc. depending on the context. No doubt such wise rephrasing would avoid many philosophical misunderstandings.

However I think the metaphilosophical point I am about to make is somehow independent of one's preferred precise definition of fiction. What I am after is whether philosophers think that the concept of fiction, as they like to define it, yields philosophical problems or not. For the same reason, I will not consider the subtle, interesting differences between the possible media of fictions. As a starting point for a general definition of "fiction", I think we should find these two general features which are really uncontroversial: fictional entities are a species of nonexistent entities and the practice of fiction is a collective practice of imagination.[2] Nothing in the following will hinge on the specifics of the definition of fiction. So I will remain neutral on the detail.

There are many ways one can take to make the metaphilosophical point that fiction is a problem for a group of philosophers only. In this introduction, I will comment on and criticise [Rorty 1981]. In "Is there a problem with fictional discourse?", Rorty defends the thesis I want to defend, although I disagree with him about the detail. He traces back the philosophical commitments at issue to Parmenides while I will suggest we should go back to Brentano instead. My suggestion is very much inspired by two works on the origin of analytic philosophy which I should make overt:

[2]There are substantial controversies about each claim in metaphysics and philosophy of mind, but as for the debate over the definition of "fiction", these controversies are either verbal or marginal.

(*i*) A conference by Graham Priest "A Lecture on Frege's Legacy";

(*ii*) A book by Jocelyn Benoist, *Représentations sans objets: aux origines de la phénoménologie et de la philosophie analytique* [Benoist 2002], which is, to my knowledge, not translated into English.

Choosing Rorty as a guide is perhaps as counterproductive as it is efficient. It will be efficient, because Rorty does the main conceptual work and arrives at the metaphilosophical claim I want to consider: so why not benefit from his work? But it may be counterproductive, because it is not sure that this article should be read seriously. First, it is rather clear that Rorty is being very ironical in this article and his irony is so pervasive that it may affect the central metaphilosophical claim I want to get from him. Second, Rorty's agenda in this article is political and institutional, thus not unbiased and for that reason he is clearly and intentionally failing to meet the common standards of philosophical argumentation.

Indeed, this article was reprinted in his [Rorty 1982] which collects a number of articles aiming at giving an answer to the following question: should philosophy be part of the human sciences or the natural sciences? He campaigns for bringing back philosophy into the human sciences. He claims that the American academy is against him and that the root of evil, so to speak, comes from pragmatism which is the dominant philosophical doctrine in those days. Nevertheless, with these *caveats* in mind, I think following Rorty into a historico-conceptual reconstruction of the philosophical problems surrounding fictional discourse is worth the trouble.

Parmenides, the Visitor and Plato's problem

According to Rorty, the philosophers who follow the tradition initiated by Parmenides have a problem with fictional discourse. Indeed, fictional discourse is a kind of discourse about nonexistent entities. But, according to Parmenides's doctrine non-being cannot be; *a fortiori*, it cannot be an object of discourse. Therefore, for Parmenides's followers, the mere possibility of fictional discourse is a problem.

This line of reasoning is well-known in the tradition. It is argued for in Plato's dialogue *the Sophist*. It is useful to go back to Plato's text for a start and reconstruct Rorty's conceptual point before we can appreciate his claim.

The problem of non-being (or the non-problem of being)

The context of the relevant passage in Plato's dialogue is the following: the Visitor and Theaetetus just agreed that two things should be distinguished. One is

the faithful images or representations (*eikon*) and the other is the false images or representations (*eidolon*). But the Visitor immediately points out that this reasonable distinction contradicts the teaching of Parmenides. Indeed, false representations presuppose that non-being is, in some sense:

> Visitor: My friend, the fact is that the investigation we are involved in is an extraordinarily difficult one. This whole matter of appearing, and seeming, but not being, and of saying things but not true things, has always caused puzzlement and confusion in the past, and it still does. It's extraordinarily difficult to grasp, Theaetetus, how one is to come out with the claim that it really is possible to say or believe things that are false, and express this without being caught up in contradiction.
>
> Theaetetus: How so?
>
> Visitor: Such a claim already dares to assume that what is not is; only on that assumption will a false thing said or believed turn out to be something that is. But, my boy, from the time I was a boy the great Parmenides never stopped testifying against it, whether expressing himself in prose or in verse: "For never shall this prevail," so his lines go, "that the things that are not are; / keep you your thought, as you search, back from that path." So we have his testimony, but most of all the statement itself will demonstrate it if we subject it to moderate crossexamination.[3]

In order to make this point, the Visitor proceeds with an argument taken from philosophy of language, for discourse is one kind of representation. The argument is the following: if we take "false discourse" seriously, then it entails that non-being is, in some sense. He also makes explicit the presuppositions of such a philosophy of language:

> Visitor: Speech, when there is speech, must necessarily say something *of* something; it's impossible for it to say something of nothing.
>
> Theaetetus: Agreed.
>
> Visitor: And must it not also be of a certain sort?
>
> Theaetetus: Obviously.
>
> Visitor: Then let's turn our attention to each other.
>
> Theaetetus: We certainly should!
>
> Visitor: So I'm going to say something to you, putting together a thing with an action using name and verb. You're to tell me what exactly the speech is speaking of.

[3] [*Theaetetus and Sophist*], 237a.

THEAETETUS: I'll do it as best I can.

VISITOR: "Theaetetus sits." There, that wasn't a long speech, was it?

THEAETETUS: No, not too long!

VISITOR: Your task is to tell me what it's speaking about and of what.

THEAETETUS: Clearly, it's about me and of me.

VISITOR: And what about this one?

THEAETETUS: What one?

VISITOR: "Theaetetus, with whom I'm presently conversing, flies."

THEAETETUS: Nobody could possibly claim anything else about this one either – it's of me and about me.

VISITOR: But now we say that each and every instance of speech must necessarily be of a certain sort.

THEAETETUS: Yes.

VISITOR: So of what sort must each of our two be declared to be?

THEAETETUS: The second, presumably, false, the first true.

VISITOR: And the true one says the things that are, as they are, about you.

THEAETETUS: Obviously.

VISITOR: Whereas the false one says things that are different from those that are.

THEAETETUS: Yes.

VISITOR: In which case it says the things that are not as if they are.[4]

This passage ends on a useful definition. A *false discourse* is a discourse which "says the things that are not as if they are". The paradigmatic example being: "Theaetetus is flying".

The Visitor's definition is thus much broader than what common sense would suggest as a definition of "false discourse". It focuses on the intrusion of non-being into natural language in general. "False discourse" as common sense would understand it is erroneous or mistaken discourse, but according to the Visitor's definition such discourse is only one kind of false discourse. Indeed, considering the speaker's intentions, we can easily distinguish between three kinds of false discourse in the Visitor's sense:

[4] [*Theaetetus and Sophist*], 262e-263b.

1. A *mistake* (or an illusion): the speaker says "Theaetetus is flying" with the intention of communicating what they believe to be real, therefore the speaker is deceived by an illusion of some sort.

2. A *lie*: the speaker says "Theaetetus is flying" with the intention of deceiving their interlocutor into believing that Theaetetus is really flying (so that they look up, for instance).

3. A *fiction* (or a *mimesis*): the speaker says "Theaetetus is flying" with the intention of having their interlocutor imagining that Theaetetus is flying (and he would meet a bird, and ...): nobody is deceived.

All three kinds of discourse "say the things that are not as if they are". But the intentional structure in each case is different, of increasing complexity. A fiction, or fictional discourse, or *mimesis* is defined as presupposing the most sophisticated intentional structure. Namely that of an overt, public, collective imaginative practice over and above the utterance of a false discourse (in the Visitor's sense).[5]

The Visitor also makes explicit the principles of his philosophy of language which clash with Parmenides's doctrine, namely that a sentence must be "of something" and "of a certain sort". Without going into the exegesis of these principles in the context of Plato's general philosophy, one can infer a principle we might call representationalism. *Representationalism* says that discourse represents something as being, or occurring, or as being the case. In other words, discourse says, at least, that something is of something.

Note that Rorty uses the expression "'picture picture' of language" to talk about this principled representative capacity of ordinary language. He takes this expression from [Wittgenstein 1921], as hinted at in the following:

> 2.0211 If the world had no substance, then whether a proposition had sense would depend on whether another proposition was true.

[5]Note that it is not clear which kind of false discourse the sentence is in the context of Plato's dialogue. It is presented as if it were a mistake, but neither the Visitor nor Theaetetus are under any illusion; it is not really a fiction, since it is not to be imagined; as for a lie, you never really know with Plato's dialogues.

This situation closely resembles Manuel Garcia-Carpintero's claim that he is not asserting anything in a strict sense when he does philosophy. This claimed he made, for instance, in a series of lectures at the IJN in 2017: http://www.institutnicod.org/seminaires-colloques/archives/cycle-de-conferences_archives/2016-2017-1000/manuel-garcia-carpintero/?lang=fr.

2.0212 In that case we could not sketch any picture of the world (true or false).

This is the first occurrence of the word "picture" in Wittgenstein's treatise, and the concept is developed throughout section 2. However, the thesis that ordinary language is a *picture*, i.e. that language by definition represents reality (the "picture picture" of language) is not stated until section 4. Nowhere is it more clearly stated than in the following:

4.01 A proposition is a picture of reality. A proposition is a model of reality as we imagine it.

4.011 At first sight a proposition – one set out on the printed page, for example – does not seem to be a picture of the reality with which it is concerned. But neither do written notes seem at first sight to be a picture of a piece of music, nor our phonetic notation (the alphabet) to be a picture of our speech.

And yet these sign-languages prove to be pictures, even in the ordinary sense, of what they represent.

Interesting though the expression might be, I prefer the expression "representationalism", for I think the way Rorty uses the expression "picture picture" conflates two distinct principles as I will show in the following section.

Conceptual reconstruction of Plato's problem

> Both literature and mathematics proceed from postulates, not facts; both can be applied to external reality and yet exist also in a "pure" or self-contained form. Both, furthermore, drive a wedge between the antithesis of being and non-being that is so important for discursive thought.
>
> [Frye 1957], p. 351.

Plato's problem as a paradox

The problem can be put in the form of a paradox. This reconstruction of Plato's problem as a paradox is hinted at in [Rorty 1981], but not presented as such. The labels are mine for I think Rorty's labels are confusing as I will show in this section.

Here are three claims which can be argued for independently but lead to a contradiction if held together.

(Parmenides) In reality, being is and non-being is not.

(Representationalism) Language represents reality.

(Sophism) There are false discourses.

Assuming (Sophism), given the above definition of false discourses, it follows that in language sometimes non-being is. Hence, by (Representationalism), in reality sometimes non-being is. But this amounts to contradicting (Parmenides). Therefore, the three claims cannot be held together. As usual, one can resolve the paradox by denying at least one of these three claims.

But the three thesis are not equally likely to be denied. It appears that (Sophism) is by far the least likely to be denied, given that we have clear uncontroversial examples of mistaken discourses as well as lies and linguistic fictions in the everyday use of language. Denying (Sophism) would lead to the conclusion that such discourses are actually not linguistic discourses but mere noise. This line of thinking seems hopeless. I do not know of any philosopher who seriously tried to deny (Sophism).

Denying (Parmenides), in contrast, is a very well-entrenched metaphysical position in philosophy. Indeed, at the time when Parmenides was teaching his doctrine, he was arguing against Heraclitus who defended the opposite thesis, according to which reality is a mixture of being and non-being. Heraclitus was just as influential as Parmenides and the debate between the two forefathers of Western philosophy arguably continues to this day.[6] But there are many other ways of denying (Parmenides) than that of embracing Heraclitus's world view. The Visitor in Plato's *Sophist* is indeed led to reject (Parmenides) and offer a complex ontology where being and non-being do not exhaust the ontological space. This route is usually called the "parricide", since the Visitor was a student of Parmenides.

Finally, denying (Representationalism) is probably the most intuitive way out of the paradox. It can be done in different ways and Rorty invites us to consider one very influential denial of (Representationalism) as put forward in the work of the late Wittgenstein on the concept of "language-games". Here is what Rorty writes:

> The language-game approach of the *Investigations* abandons this "picture picture" of language precisely in allowing that whether a sentence has

[6]This is, for instance, one of Nietzsche's metaphilosophical thesis that the metaphysical debate between Parmenides and Heraclitus is definitional of Western philosophy, as put forward in [Nietzsche 1908].

> sense (i.e., can be intelligible and true or false) *may* be dependent upon whether another sentence is true. Since this possibility is paradigmatically actual in the case of fictional discourse [...], philosophical problems about fiction simply do not arise once the picture picture is dropped. Nor, for parallel reasons, do sceptical problems about life being a dream, nor the problem of how scientific theories may be "philosophically" distinguished from poems.
>
> The common root of all these problems is the fear that the manifold possibilities offered by discursive thought will play us false, will make us "lose contact" with the real. This fear is, as Heidegger has argued, definatory of the Western philosophical tradition.[7]

Here is the crux of Rorty's insight: there is a philosophical problem about fictional discourse only if one accepts (Representationalism). But one can deny it, e.g. Wittgenstein's theory of language-game does so. So, there is no "problem of fictional discourse" in general, only relative to one's commitment to (Representationalism).

Shortcomings of Rorty's claim

One can already see that focusing on (Representationalism) only is an arbitrary narrowing of the logical space. I suggest that we should consider *all* the necessary commitments for it to be a problem with fictional discourse.

In fact, in [Rorty 1981] he conflates (Parmenides) and (Representationalism) into a single thesis called the "Parmenidean Picture Picture". Roughly, he targets under this label all the semantics whose aim is to give a theory of "how words relate to the world", as he himself puts it. I think what he has in mind, given what he says about it, is a referentialist semantics in which the basic building blocks involve direct reference to real individuals. In today's language, he would target the idea of a truthmaker semantics.[8]

Consequently, Rorty claims that both (Parmenides) and (Representationalism) shall stand or fall together. But this is not true. Indeed, one can deny (Parmenides) while holding on to (Representationalism). The Visitor arguably does so in *the Sophist*. Another somewhat extreme example of such a strategy can be found in Montaigne's essay "Du repentir" (III, 2) where he famously declared: "Je ne peins pas l'être. Je peins le passage." In this essay, Montaigne defends a Heraclitean

[7] [Rorty 1981], pp. 129-130.

[8] For a philosophical discussion of the so-called "truthmaker principle", see [Armstrong 1973], p. 11. See also Kratzer's 2019 Leverhulme Lectures for a linguistically motivated version of truthmaker semantics.

ontology in which everything is constantly changing and thus every being is always becoming non-being. But, in the same breath, he defends the view that language can represent this ever changing world.

One can also deny (Representationalism) while holding on to (Parmenides). In fact, Wittgenstein's language-game theory can be interpreted in this way: it says that the truth and falsity of linguistic discourse does not depend on how language represents reality; but it says nothing about the ultimate nature of reality. It is perfectly compatible with (Parmedides) according to which reality is wholly on the side of being and involves no mixing of being and non-being.

The second shortcoming of Rorty's claim is that are different ways of denying (Representationalism) which he does not consider. Actually, his choice of the late Wittgenstein can be seen as a very bad choice. Indeed, the late Wittgenstein is interesting because he criticises the project of a correspondence theory of truth for natural language. The idea that a theory of truth should be build on the notion of correspondence to reality is a very old and well-entrenched idea which was first forcefully defended as such in the work of Aristotle.[9] Rorty claims that all the philosophers who accept the idea of a correspondence theory of truth should be grouped together and labelled: the Permenidean philosophers. Moreover, he interprets the notion of "correspondence" so broadly in [Rorty 1981] that Russell and Meinong happen to both be Parmenidean. Consequently, Rorty heroically succeeds in reconciling Russell and Meinong (against the second Wittgenstein): both have a problem with fictional discourse.

One can now see that Rorty's definition of the "Parmenidean philosophers" is much too coarse to even mean anything, for Russell and Meinong are clearly *not* on the same side. This suggests that Rorty very likely presents this result with his tongue in his cheek, although the irony is very subtle. Be it as it may, I think the reasonable conclusion is that the theses in Plato's paradox are vague. Rorty's text, unfortunately, does not help to clarify the concepts involved. Indeed, my reconstruction of Plato's argument, following Rorty's hints, is more precise than Rorty text. So in order to make things more precise, I need to change the framework and historical background and shift from Plato to Brentano.[10]

[9]For a general presentation of such theories, see David, Marian, "Correspondence Theory of Truth", *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2016/entries/truth-correspondence/.

[10]Jean-Marie Schaeffer, in the first chapter of [Schaeffer 1999], also traces back the problem of *imitation* to Plato. However, he focuses on the philosophical problem concerned with the *value* of fictional discourse and other kinds of *mimesis*, for Plato famously proposed to deport the poets out of his ideal state (in *Republic X*). So one can trace back to Plato's work two distinct, though related problems about fictional discourse: that of the *possibility* and that of the *desirability* of

Conceptual re-reconstruction

The missing notion in the previous reconstruction is that of reference. This notion puts us in a very different conceptual background, which is that of Brentano. Here is a restyling of the above paradox using the notion of reference:

> (Referentialism) Reference presupposes existence.[11]
>
> (Intentionalism) Linguistic (*qua* mental) representations refer.
>
> (Emptyism) There are empty representations, i.e. whose object they refer to does not exist.

These theses lead to a contradiction when held together. One can see the similarities with the previous paradox. However, (Intentionalism) is quite obscure if left unexplained and unmotivated.

First I will motivate (Intentionalism) calling in Brentano. Then I will show that what is at the heart of the issue is a metaphysical controversy about the notion of relation. Finally, I will show how this relates to the problem of fictional discourse.

Brentano's problem

Definitions

In [Brentano 1874], one can find the famous thesis that *intentionality* is the characteristic feature of mental phenomena. That is, every mental representation (*a fortiori* linguistic ones) is defined as "having an object":

> Every mental phenomenon is characterised by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as an object within itself, although they do not so in the same way. In presentation, something is presented, in

fictional discourse. The first problem points to the semantics and the other to the ethics of fictional discourse. Both are related, in Plato's work, through a presupposed ontology of fictional objects. I am about to show that the conceptual framework for the semantics of fictional discourse can be fruitfully updated if one shift to Brentano. Interestingly, Schaeffer argues, there is no updated framework for thinking about the ethics of fictional discourse.

[11] Another way of putting it, following Rorty, is: "Whatever is referred to must exist."

> judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.
>
> This intentional in-existence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves.[12]

But this definition, as appealing as it sounds, comes with a serious problem. "Having an object" is intuitively thought of as relational in nature: the intentional object is related to the cognitive agent *via* the mental act (presentation, judgment, love, etc.). At the same time, intentionality cannot be a *real* relation, because some mental representations are empty, meaning that they relate the cognitive agent to, literally, nothing:

> What is characteristic of every mental activity is, as I believe I have shown, the reference to something as an object. In this respect, every mental activity seems to be something relational. [...] In other relations both terms – both the fundament and the terminus – are real, but here only the first term – the fundament – is real. [...] If I take something relative, [...] something larger or smaller for example, then, if the larger exists, the smaller one exists too. If one house is larger than another house, the other house must also exist and have a size. Something like what is true of relations of similarity and difference holds true for relations of cause and effect. For there to be such a relation, both the thing that causes and the thing that is caused must exist. [...]
>
> It is entirely different with mental reference. If someone thinks of something, the one who is thinking must certainly exist, but the object of his thinking need not exist at all. In fact, if he is denying something, the existence of the object is precisely what is excluded whenever his denial is correct. So the only thing which is required by mental reference is the person thinking. The terminus of the so-called relation does not need to exist in reality at all. For this reason, one could doubt whether we really are dealing with something relational here, and not, with something somewhat similar to something relational in certain respect, which might, therefore, better be called "quasi-relational".[13]

Brentano's problem thus concerns the articulation between two fundamental notions: relation and existence. Given the data he considers, there are two very different

[12] [Brentano 1874], p. 68.
[13] [Brentano 1874], p. 211-2.

cases.

In the first case, the relation comes with a *full* ontological constraint: both relata must exist for the relation to hold.[14] Brentano's examples are the relation "being smaller than" and "being the cause of". We can think of more naïve examples like: "spilling one's coffee on", "kissing", etc. Let us call them the *natural* or *extensional* relations.

In the second case, the relation comes with a *half* ontological constraint: only one of the relata must exist, i.e. the cognitive subject. Brentano's example is the relation "deny the existence of". One can think of other intuitive examples like "dreaming about", "imagining", etc. Let us call them the *intentional relations* or the mental ones.

Important distinction between direct intentionality and derived intentionality

Direct intentional relations: As suggested above by Brentano's light, the paradigmatic intentional relation is a relation holding between a cognitive agent and an object of cognition. All cognitive states require the existence of the cognitive agent, by definition. Interestingly, *prima facie*, some cognitive state require the existence of the object of cognition and some others do not. For instance, perceiving, remembering or having a singular thought are usually regarded as putting an ontological constraint on what is perceived (it must be in the vicinity of the cognitive agent), remembered (it must have been experienced by the cognitive agent), singularly thought (the cognitive agent must be able to deploy a causal informational channel to the object of cognition). In this sense, these cognitive states behave like *natural* relations. By contrast, as seen above, many other cognitive states do not put any constraint on the existence of the object of cognition and are thus paradigmatic cases of intentional relations.

So the mental lives of cognitive agents is a complex mixture of the two kinds of relations. The pressure a philosopher might feel in considering that all mental states are of one kind justifies their calling all mental relations intentional. But, eventually, one should have a more fine-grained picture of the mind. This is part of Brentano's program: explain how intentional and natural relations are integrated in one's mental lives.

[14] More rigorously the constraint on the relation is: if the relation holds in reality, then both relata must exist *qua* objects in reality. In Brentano's quoted passage he puts the constraints on the *objects* and not on the relation when he says: "if the larger exists, the smaller one exists too". This is part of Brentano's mistake, that was later corrected by his students, as seen below.

Derived intentional relations: But there is another very important class of intentional relations which do not involve any cognitive agent. It is the relation between a representation and that which is represented. In this case, one usually talk of *derived* or *borrowed* intentionality, since it is usually acknowledged that representations are created and used by cognitive agents.

Classic examples are: maps, pictures, descriptive sentences, etc. The intentional relation, in this case, is a relation which holds between a physical object and the object of representation. The physical object must exist by definition, because otherwise there is no representation. But the object of representation need not exist, for maps can chart imaginary territories, one can draw nonexistent individuals, etc.

Just as in the case of intentionality of cognitive states, some representations impose an ontological constraint on the object of representation, because of the structure of the representation. The main contrast between a drawing and a photograph, for instance, is that the object of the photograph must have been causally connected with the photograph, so it must (have) exist(ed); not so for the object of a drawing. Again, one might want to put the photographic relation in the natural relations instead of the intentional ones or rather insist on representations as being of one kind.

Following the tradition, such borrowed intentionality of representations is usually called "signification". It is because representations are traditionally called *signs*.[15] Here is, for instance, the definition of a sign in the *Port-Royal Logic* which precisely corresponds to a representation as defined above. The borrowed intentionality can be defined as a "prompting relation", to borrow Arnauld and Nicole's expression:

> When we consider an object in itself and in its own being, without carrying the view of the mind to what it can represent, our idea of it is an idea of a thing, such as the idea of the earth or the sun. But when we view a certain object merely as representing another, our idea of it is an idea of a sign, and the first object is called a sign. This is how we ordinarily think of maps and paintings. Consequently the sign includes two ideas, one the thing which represents, the other of the thing represented. Its nature consists in prompting the second by the first.[16]

[15] The tradition goes back (at least) to Augustine's work who defined the notion of *signum* in this manner (see *De doctrina christiana* – II, i, 1).

[16] [Arnauld and Nicole 1996], Part 1, Chapter 4, p. 35. Here is the original of the last two phrases, in which one finds the beautiful, suggestive expression defining a sign:

> Ainsi le signe renferme deux idées, l'une de la chose qui représente, l'autre de la chose représentée; et sa nature consiste à exciter la seconde par la première.

Brentano's program

Brentano thus presents us with an original philosophical problem which can be summed up in the following question: what is the unified theory of these two kinds of relations? His logical space was simple: there are extensional and intentional relations. One should have a general theory of relations which explains this distinction.

Brentano's program is a way of answering this question. It comes with the presupposition that the extensional relations are genuine relations whereas intentional relations fail to be genuine relations. As shown in the quotes above, he calls them "quasi-relation" The program thus consists in trying to explain or derive the intentional from the extensional. This program clearly indicates Brentano's empirical stance.

But Brentano's logical space is incomplete and underwent crucial modifications when taken up by his students. Indeed, one can see that Brentano's empirical program suggests a dual idealistic program. It consists in taking intentional relations to be primitive and deriving natural relations from intentional relations. This dual program is idealist in that it reduces all relational phenomena to mental phenomena. This duality invites us to extend Brentano's original logical space of relations.

Instead of considering only genuine relations and quasi-relations, one should consider three kinds of relations, depending on ontological constraints on the relations. First, one can impose a *full* ontological constraint on a relation. A full ontological constraint imposes that *both* objects in relation exist. This corresponds to what Brentano has in mind when he talks about extensional relations. Second, one can impose an *empty* ontological constraint on a relation. An empty ontological constraint imposes that *neither* objects in relation exist. This is a kind of relation Brentano did not consider. Third, one can impose a *half* ontological constraint on a relation. A half ontological constraint imposes that *only one of* the objects in relation exist.[17] This corresponds to what Brentano has in mind when he talks about quasi-relations.

In this broaden logical space, Brentano's problem can be rephrased thus: how is it possible to impose a half ontological constraint on a relation so as to get the intentional relations? Among Brentano's students, those who see the intentional as half-full are the founders of analytic philosophy, whereas those who see it as half-empty are the founders of phenomenology.

[17] One could then distinguish between two kinds of half ontological constraint, depending on which of the "fundament" and the "terminus" bears the ontological constraint, to take Brentano's terminology. This fine-grained distinction will not be relevant for the following discussion though.

The metaphysics of relations

In order to answer Brentano's question "what is a quasi-relation?" one is thus led to consider the metaphysical question: What is a relation? Given the logical space given above, two theories of relations stand out and prove very useful to understand what a quasi-relation is.

The combinatorial interpretation

There is what I will call the *combinatorial* interpretation of relations. It says that a relation between two objects is a combination of the two objects. How one should interpret "combination" can be spelled out in different terms. But, if one follows the now standard logic of relation, one can define the combination in general as an ordered pair.

Indeed, in the logic of relations systematised by Russell and Whitehead in [Whitehead and Russell 1912], one can define relations extensionally. According to the extensional definition, a *relation* over two sets of individuals is a subset of the Cartesian product of the two sets. A *Cartesian product* of two sets is a mathematical operation which extracts all the ordered pairs from two sets of individuals. In set-theoretical idiom:

$$A \times B =_{def} \{(a, b) : a \in A \ \wedge \ b \in B\}$$

Following this definition, the objects related are ontologically independent of the relation. In loose talk, we say that, according to the combinatorial interpretation, the *relata* pre-exist or precede the relation.

The foundational interpretation

The dual interpretation is what I will call the *foundational* interpretation. It says that a relation ontologically grounds the *relata*. In other words, in this case, the objecthood of the *relata* is strictly speaking relative. The objects related are ontologically dependent upon the relation. In loose talk, we say that, according to the foundational interpretation, the relation pre-exists or precedes the *relata*.[18]

[18]One might want to call it the *structural*, or the *differential* interpretation. Indeed, Jakobson's contribution to structural linguistics led him to define phonological relations precisely in this way, first and foremost in [Jakobson 1929]. This idea was later generalised by Levi-Strauss, defining an original notion of *difference*, as in [Lévi-Strauss 1949]. Although this structuralist movement arguably first made a scientific use of this interpretation, I think the philosophical interpretation was available before, in the work of Husserl, especially in [Husserl 1913] (§135), where he describes the constitution of mental objects. This interpretation can also be found in many other philosophical

Here are some examples so as to help having an intuition about the foundational interpretation. Given a sphere in rotation around an axis, one can define cardinal points: the North and the South poles. These points exist only because all the points on the sphere are related to the axis by the rotation movement. The poles are defined as the intersections of the axis with the sphere. In this sense, the rotation seen as a relation precedes the polar points.

Phonemes in natural language are usually defined differentially in the acoustic space. The phonemes exist only because they are in relation with each other in a complex phonetic system and this is why all natural languages have different phonetic systems although they share the same acoustic space.

Roles in conversations are dependent upon the relations which hold between the participants: nobody is essentially a speaker or essentially an interlocutor.

There are also examples taken from the mental life which are of importance to this foundational interpretation. One can find in [Merleau-Ponty 1945] a theory of perception as sketching the object of perception which is a case in point. He developed a theory of perception according to which the perceiver and the object of perception emerge as subject and object from the relation of perception and its structural features.

Other examples can be found in the literature on so-called existential emotions. Sartre's analysis of shame in [Sartre 1943], for instance, makes it the case that the subject and the object of shame exist only relative to someone else's look on the shameful situation.

Historical ground

Historically, one can say that analytic philosophy in its beginnings adopted the combinatorial interpretation of relations, whereas phenomenology in its beginnings adopted a foundational interpretation of relations. Both philosophies are the offspring of Brentano's program. They simply answer differently the metaphysical question: what is conceptually primitive, the relation or the *relata*?

At this point, I have a counter-proposal to Rorty's claim. I think that fictional discourse is a problem for philosophers who endorse a combinatorial interpretation of relations only. In other words, fictional discourse is a problem for analytic philosophers only. I will now show how this counter-proposal fares better than Rorty's original claim.

---

works of the phenomenologist tradition. I use the term "foundational" to be as neutral as possible regarding the scientific and philosophical doctrines which breed on this interpretation.

Consequences for the problem of fiction

Consequences for combinatorialism

Let us go then, back to our paradox:

> (Referentialism) Reference presupposes existence.
>
> (Intentionalism) Linguistic (or mental) representations refer.
>
> (Emptyism) There are empty representations, i.e. whose object they refer to does not exist.

Given the combinatorial interpretation of relations and the fact that reference is a relation, all three claims are true; hence, we have a problem with fictional discourse.

Indeed, (Referentialism) is true, because according to the combinatorial interpretation of relations, for it to be a relation of reference, the *relata* of the relation of reference must exist. Hence, making reference to something presupposes that that something exists. (Intentionalism) is true, because linguistic representations, *qua* mental representations, are signs. Signs, given the definition given above, have indirect intentionality. So, given Brentano's theory of intentionality, they refer. (Emptism) is empirically true.

There are thus three ways of blocking the paradox, which consists in denying one of the three claims. Each solution has been explored by Brentano's students:

- Meinong denied (Referentialism) in claiming that one can refer to nonexistent objects.[19]
- Frege denied (Intentionalim) in claiming that there is a notion of sense without reference, so that linguistic representations need not refer to have meaning.[20]
- Twardowski denied (Emptism) in claiming that every representation has a real "intentional act" as its (primary) object.[21]

Russell followed Frege here. As a result, one can see how, in this picture, Meinong and Russell are indeed close cousins. But they are also not in the same basket at all, because they solve the problem of fiction in a radically different manner. Consequently, following this proposal, Rorty's original idea is still valid, for it is true that fictional discourse is not a philosophical problem but a problem for a kind of philosophy. But the landscape is more fine-grained so that we avoid Rorty's shortcomings.

[19] See [Meinong 1904].
[20] See [Frege 1892].
[21] See [Twardowski 1912].

Consequences of the foundational interpretation

By contrast, assuming the foundational interpretation of relations and the fact that reference is a relation, it is not the case that the three claims are should be taken to be true together; hence, there is no problem of fiction.

On the one hand, (Referentialism) is not true, and it is even impossible by definition. If reference is a genuine relation which grounds the existence of its *relata*, then it should not presuppose the existence of its *relata* on the pain of circularity. On the other hand, (Emptyism) is a weird claim to make for the foundationalist, since, by definition the objects of a representation *never* exist independently of the representation; *a fortiori*, (Emptyism) is trivially true. (Intentionalism), given Brentano's theory of intentionality, is also true.

Note the closeness of Meinong's solution to the problem and the foundationalist position. They both reject (Referentialism): the former using an argument within the combinatorial framework, the latter by appealing to metaphysical considerations about the nature of relations. This, I think, explains the fact that Meinong is very often labelled a phenomenologist and why he takes so much time to differentiate himself from the existing phenomenological account in [Meinong 1904]. But I think the present overview allows for useful, subtle distinctions.

Dual problems

If the foundationalist does not have a specific problem with fictional discourse, they have a problem understanding non-fictional discourse. Indeed, they have a dual problem which consists in explaining how discourses about reality are possible. For if the object of discourse is not independent of the intentional relations which grounds it, then it seems that explaining the intuitive existence of the objects of, say, perception or knowledge, is a challenge.

For instance, in the work of Husserl, the problem of theorizing natural or extensional relations was very salient. Indeed, following Descartes, Husserl could easily ground the independent existence of the cognitive agent from the relations it entertains with objects in the external world. But grounding the "objectivity" of these intentional objects is really difficult. Of course, one pressing worry for Husserl was to account for discourse produced by the empirical sciences which is one kind of discourse for which it seems that the object precedes the relation we, cognitive agents, entertain with it. I think it is fair to say that this has been (and perhaps still is) one of the central problems of phenomenology.[22]

[22]Here is how [Rivenc 1995] very clearly describes the problem, commenting on [Husserl 1913] (§36) where Husserl makes it clear that intentionality is not a relation in the combinatorial sense:

Hence, just like there is a "problem of fiction" for analytic philosophy, there is a "problem of reality" for phenomenology.

When it comes to giving a theory of language which accounts for both fictional and non-fictional uses of language, one is thus led to choose between two different strategies. Phenomenology starts from the idea that we know what is going on when we read fictional texts, and then proceeds in questioning what is going on when we engage with non-fictional representations. The analytic strategy is converse. One starts with the idea that we know what is going on when we engage with, say, scientific discourse, and then proceeds in questioning what is going on when we read fictional texts.[23]

This dissertation is an instance of the analytic strategy.

Content and structure of the dissertation

Analytic philosophy has thus developed a very specific interest in fiction. There are many problems surrounding fictional representations which have been tackled by analytic philosophers in the last century, which are still very much debated and, for most questions, open problems. Among these problems, one can find questions like: How come one can learn things about the real world through reading fiction? How come we can be moved by fictional characters while knowing that they do not exist? And many others which are grouped under the label "philosophy of fiction".

In this dissertation, I will contribute to one goal of the philosophy of fiction by providing a semantics for fictional discourse. In particular, I will focus on the two central notion of semantics which are truth and reference. I will be looking for a theory of truth and reference in fiction.

I will defend a view I call "functionalism" which originates in the work of Kendall Walton. I am not going to defend Walton's philosophy of fiction as a whole.[24] In-

(my translation)

> That intentionality is not a relation in the sense that a century of logical-mathematical culture has accustomed us to understand this word, it follows from the very position of the problem that this problem is intended to solve: the constitution within the subject's life of the ideality of all objectivity. That there is a whole enigma here for those who remain in the naïve position of scientific and philosophical realism, Husserl was fully aware of this. However, the analysis of the noematic structures should, in [Husserl's] opinion, allow us to see more clearly into these problems of constitution.

[23] For a recent illustration of this dual strategy around fiction, see [Woods 2018], §3.2 (pp. 57-58).
[24] Guillaume Schuppert recently defended such a doctoral thesis on Walton's philosophical system.

stead, I aim at constructing a new version of functionalism whose debt to Walton's ideas will be repeatedly acknowledged throughout. I will be focusing on linguistic fictions for reasons of space and competence. In this respect, I will follow and use [Everett 2013] which is an application and development of Walton's idea in order to solve problems within linguistics and philosophy of language.

Walton's theory of *mimesis* does not originate in the solving of semantic problems. Its original concern came from aesthetic problems and the philosophy of art. Philosophers of art look for a general theory of what aesthetic experience is, i.e. what it means to engage with a work of art. Walton forcefully proposed that there is a continuity between children's games of make-believe and aesthetic practices, and develops a general theory of make-believe which culminates in his *magnum opus* [Walton 1990]. From the start, philosophers of language stated their interest in Walton's theory. Among others, Gareth Evans pioneered in taking up Walton's programmatic ideas into a full-fledged semantics for fictional discourse.

When it comes to truth and reference, functionalism met already existing theories and heated debates started right on. The central tenet of this dissertation is that functionalism can meet the challenges posed by rival theories and solve the difficult problems raised by linguistic data. Thus I will engage on the two fronts awaiting functionalism.

On the problem of "truth in fiction", the rival account is a possible-world account of truth in fiction with all its ramifications. The main figure of this rival account is David Lewis. I will argue against the possible-world account of "truth in fiction" and propose instead what I call a pretence semantics in the wake of [Evans 1982] and [Everett 2013]. This is part 1.

On the problem of reference in fiction, the rival account is that of the realist in all its different forms, from the hard-core neo-meinongians to the mildest artefactualists about fictional characters. Realists hold that fictional characters exist in some sense, while functionalists are, to the bone, anti-realists holding that fictional characters do not exist in any sense. I will argue that pretence semantics as developed in part 1 can be naturally combined with a causal theory of reference to explain the complex linguistic data raised by realists. This is part 3.

Part 2 does not engage in any philosophical debate with rival views. Instead, I will go into the detail of the functionalist account. The devil being in the detail, it will be interesting to fine-tune the theory developed in the previous part on a hard case. This part focuses on a case study, recently unearthed by Stacie Friend, called the "Great Beetle Debate". In trying to understand what literary critics are arguing

See [Schuppert 2019]. He nicely shows how Walton's ideas are both incredibly encompassing and, in some detail, unsystematic.

about when they argue, I will show that the notion of fictional "truth" as defined by the functionalist meets a particular form of indeterminacy which is not linguistic, epistemic or ontological but, I will argue, pragmatic.

At the end of each chapter, I put appendices. They are digressions and analogies inspired by contemporary works in computer sciences and mathematics. It is useful to set them aside, because they would drive the reader away from the philosophical problems discussed in the dissertation. But I think they are important to the dissertation for they shed interesting lights on the new philosophical notions which I introduce. To be honest, the "novelties" I present are really nothing new. I simply borrowed ideas from these other disciplines and transformed them so as to contribute to existing debates in the philosophy of fiction. The presence of these appendices should make this plain.

Moreover, it so happens that inspiration from computer science in the studying of fiction is a long story. Marie-Laure Ryan, in her thought-provoking [Ryan 1991], identifies five "contributions of AI to literary theory":

> The potential contributions of AI to literary theory and textual semiotics fall into five categories: a fundamental belief; a lesson in methodology; a set of questions defining a particular approach; a repertory of analytical tools; and a source of analogies.[25]

I think the appendices of this dissertation fall into the last two categories. When Ryan talks of analogies, she has a certain framing metaphor in mind:

> For years, the leading analogy in literary thinking has been the linguistic one: literature is a language (as well as, metonymically, an artefact made of language); the text is a "system of signs". Involvement with AI and computer science suggests another metaphor: the text as a machine. Or if the text is a language, why not a computer language?[26]

This framing metaphor does not come from Ryan's work. Among others, Calvino made decisive contributions towards such a framing metaphor in [Calvino 1980]. As computer science greatly changed in the past forty years, it is useful to update the metaphor with the recent results and problems one can find in computer sciences.

[25] [Ryan 1991], p. 5.
[26] [Ryan 1991], p. 8.

Part I

Truth in fiction: towards pretence semantics

Chapter 1

The problem

1.1 Description of the problem

The so-called problem of "truth in fiction", or fictional "truth", consists in explaining the contrast between pairs of sentences like the following:

(1) Hamlet is a human being.

(2) Hamlet is a crocodile.

There is a contrast between (1) and (2). In order to introduce this contrast, it is usually[1] said that (1) is intuitively *true* in *Hamlet* whereas (2) is intuitively *false* in *Hamlet*.[2] This suggests that the contrast between (1) and (2) should be reflected in the *truth-value* of the two sentences. Ay, there's the rub! (Hamlet would say). Indeed, the claim that the above contrast should be reflected in the truth-value of (1) and (2) is a highly controversial one. In this part, I will present and engage in this controversy about the theorizing of fictional "truth". Before I provide technical definitions, I will flesh out what the above contrast is about.

The play *Hamlet* is a set of sentences written by Shakespeare in England around year 1600. It so happens that neither (1) nor the denial of (2) is one of the sentences of *Hamlet*. But one relevant sentence Shakespeare wrote is that Hamlet is a Prince; indeed, it is even clear from the entire title *The Tragedy of Hamlet, Prince of Denmark*. With this piece of information in mind, it is very easy to explain the above contrast following this line of reasoning: Hamlet is a prince, and princes are human

[1] See in particular [Lewis 1978], [Walton 1990], [Currie 1990].

[2] I take Shakespeare's *Hamlet* to be a purely fictional play. Please choose your own example if you think this is not the case.

beings, hence (1) is true in *Hamlet*; and (2) is false in *Hamlet*, since (1) is true in *Hamlet*. So the problem is about inferences.

But, in fiction, inferences can go astray. In some fictions, princes are crocodiles. In such fictions, (2) would be intuitively true... Of course, *Hamlet* is not such a story. But how do we know this?

In general, then, in order to predict such contrasts as that between (1) and (2), we need to define an inference relation for the relevant fiction against which the sentences are interpreted. Such an inference relation should, in principle, extract all the intuitive "truths in fiction" from the set of sentences which constitute the relevant fictional text. This what is called the "problem of truth in fiction". A theory of fictional "truth" should thus explain how actual readers draw the inferences they draw when they read a text of fiction. Such a theory should therefore at least predict contrasts like the above one.

This talk of drawing inferences suggests readers handle *propositions*. I will say that a proposition is *fictional* when it is expressed by a fictional sentence, like (1).[3] Following the above reasoning, I will call *fictional sentences* the sentences which constitute a work of (linguistic) fiction or a story. The focus of this part is thus clearly on linguistic fictions. Everything I will say apply to non-linguistic fictions, though, provided one has a theory about the semantic content of non-linguistic fictions.

However, it is arguably problematic to talk about fictional *propositions*. Indeed, a sentence like (1) arguably contains a name which does not refer, namely "Hamlet". So, given compositionality, the sentence (1) cannot express a complete proposition.[4] One way of thinking about propositions is to say that they are about real *facts*. By definition, fictional events are not real facts, though. This also suggests that the term "proposition" is misplaced in the context of fiction.

On the other hand, it seems clear that fictional sentences have semantic content. Fictional sentences are not mere noise, but grammatical sentences. Fictional sentences can be interpreted. Stories can be translated, adapted into other media, and summed up. The semantic content of a story is what survives such transformation; it is expressed by the story. The situation is strictly analogous to semantic content of a non-fictional piece of writing. Some would say that given a strict definition of what counts as a proposition, one should hold that the semantic content of a story is not a set of propositions. Nothing I will say hinges on this debate on the

[3]The terminology comes from [Walton 1990]. I will use this technical terminology whenever possible and put quotation marks around "truth" whenever I use the expression fictional "truth" so that I avoid all suggestion that fictional "truth" is a kind of truth.

[4]This thesis originates in [Frege 1892]. See also Russell's formulation defence of "logical atomism" in [Russell 1918].

nature of propositions.[5] Therefore, I will adopt a definition of proposition which is loose enough to encompass fictional propositions.For simplicity, I will use the term "fictional proposition" to designate this semantic content expressed by a fictional sentence.[6]

A couple more definitions: I will say that a fictional proposition is *explicit* when it is expressed by a fictional sentence. I will say that it is *implicit* when it is inferred from another fictional proposition. We can now rephrase the problem of fictional "truth" in this manner: a theory of fictional "truth" should give a systematic way of retrieving all the fictional propositions, as opposed to the non-fictional ones, given a fictional text. In particular, such a theory should say how the implicit fictional propositions are derived from the explicit ones.

1.2 Non-starters

A straightforward correspondence-theory of truth will not work.[7] Some have even argued against such a theory of truth on the ground that it does not square with our intuitions about fictional "truth".[8] I suggested that this is taking the word "truth" too literally here. Thinking of fictional "truth" as a contrast between fictional and non-fictional propositions shows that the problem is not here.

In fact, the above contrast shows precisely that the problem is not about *reference*. So that the usual tricks to circumvent singular terms which do not refer to any real individual, such as "Hamlet", is clearly a non-starter. Take, for instance, Russell's method of paraphrasing an empty singular term.[9] Then, both (1) and (2) will end up false, since both paraphrases contain the assertion that there is a real individual which is called "Hamlet", and this is false, by definition. On the other hand, if we try to appeal to Strawson's idea of a presupposition failure, then it follows that both (1) and (2) are truth-value-less.[10] Such strategies are thus missing the above contrast.

[5] In particular, some people want to say that they are *incomplete* or *gappy* propositions, as first argued for in [Braun 1993]. I am neutral on this.

[6] The expression *fictional content* was proposed in [Stock 2017] for the same purpose. It may be preferred, though I find it less natural. Besides, the expression "fictional proposition" is widespread in the philosophical literature.

[7] The root of such theories comes from Aristotle. The more contemporary version of the idea of "correspondence" is the idea of a truth-maker semantics. See the SEP entry on The Correspondence Theory of Truth for a comprehensive presentation of such theories and further references.

[8] See [Rorty 1981]

[9] See in particular [Russell 1919].

[10] See [Strawson 1950].

At this point, one can think that the above contrast is pragmatic in nature, and appeal to the idea of pragmatic enrichment. Suppose, for instance, that we think Grice's mechanism is at work. But, given that (1) and (2) are about a fictional character, both sentences run counter to Grice's maxim as expressed in [Grice 1975]: "Do not say that for which you lack evidence." Indeed, the only "adequate evidence" one has for the "truth" of (1) (and for the falsity of (2)) is the inference made explicit above. But this is precisely what we are trying to get. So, putting aside such evidence for fear of circularity, Grice's story would have it that both (1) and (2) are infelicitous. Again, the above contrast is missed.[11]

What this shows is that it is important to focus on the *intuitive contrast* between (1) and (2) and not rush into trying to predict the *intuitive truth* of (1) alone. The problem of "truth in fiction" is a problem about *inference* not about *reference*.[12]

Note on Grice: Interestingly, it seems that this apparent violation of the maxim of quality does not happen in a context where the overarching principle of cooperation fails. Indeed, an overtly fictional utterance of (1) is not uncooperative in a way a lie would be. On the contrary, it seems that the hearer has to be *more cooperative* so that they do not reject this talk about a nonexistent prince. I think this surprising violation of Grice's maxim of quality in a cooperative fashion is precisely what Coleridge has in mind when he coined the famous expression: "willing suspension of disbelief [...] which constitutes poetic faith."[13]

1.3 Explicitism and Intentionalisms

I want to dispose of two recent minority views which appeared in the literature, before I consider the debate between the two competing views which is the main subject of this part. I will first argue against explicitism. Then I will present the

[11] Interestingly, it seems that (1) *is* infelicitous in certain contexts. For instance, if one reads (1) in a serious history book, one is likely to frown. By contrast, it seems that this is not the case in an "ordinary" conversation about Shakespeare's play. This shows that the uttering of such fictional sentences are very context sensitive. Again, this remark is beside the point, since we should focus on the *contrast* between the two fictional sentences once the context has been fixed.

I am caricaturing Grice's position here, for one can refine Grice's view so as to get something very close to functionalism. See for instance [Currie 1990] and [Stock 2017]. The point I am trying to make is that the problem of truth in fiction should be to predict the above *contrast* and not merely the "truth" of (1). Grice's mechanism, if taken at face value, treat both (1) and (2) in the same manner: they are infelicitous.

[12] On the problem of fictional "truth" as a general problem of inference, see appendix A.

[13] [Coleridge 1817].

different views which fall under the label "intentionalism" without discussing them in detail for these views are not meant to address the problem of fictional "truth" as defined above.

1.3.1 Against explicitism

One can find a defense of *explicitism* about fictional "truth" in the recent [D'Alessandro 2016]. Here is his definition of the view:

> Explicitism, as I've defined it, is the thesis that the set of true propositions in a given fiction is a subset of the explicit propositions in that fiction.[14]

Implicitism is the denial of explicitism thus defined.

D'Alessandro's defense of a form of explicitism is a heroic and somewhat desperate attempt, as he himself confesses. Here is the way he describes his own defense:

> [...] no philosopher that I know of has seriously questioned implicitism or attempted to defend any version of explicitism. My second and primary goal here is to challenge this orthodoxy. I hope to show that implicitism is much less obviously true, and that explicitism is much less obviously false, than the standard wisdom suggests.[15]

Indeed, "standard wisdom" has it that explicitism faces two serious problems: it both undergenerates and overgenerates fictional "truths". It obviously undergenerates, since we have the intuition that there are some implicit fictional propositions like (1).[16] It also overgenerates, because it sometimes happens that the narrator is unreliable so that fictional sentences should not be taken at face value: in this case, the expressed explicit fictional propositions are intuitively not "true in the fiction". For instance, Humbert in *Lolita* explicitly asserts that Lolita was willing to have sex with him and that he did not, but that he eventually gave in to temptation. Given many narratological cues to be found in *Lolita*, the reader is to question such a description of the fictional events and rather adopt as a fictional "truth" that Lolita is a victim of paedophilia and did not give her consent.

This point was made over and over again. Here is, for instance, Byrne dealing away with explicitism:

[14] [D'Alessandro 2016], §4.
[15] [D'Alessandro 2016], §1.
[16] Maybe (1) follows deductively from what is explicitly stated in Shakespeare's play. One can easily change the example as is done below.

> Why not identify what is true in a fiction with what is explicitly said in the fiction (or follows deductively from what is explicitly stated)? Well, in some fictions there are deluded narrators, and so they speak falsely. Therefore the proposal doesn't give a sufficient condition. But it doesn't give a necessary condition either. There are many truths in fiction which aren't explicitly stated, and aren't entailed by what is explicitly stated. It's true in the Holmes stories – as Lewis pointed out – that Holmes doesn't have a third nostril, and that he never visited the moons of Saturn. However, neither of these propositions is explicitly stated in the stories, or entailed by what is explicitly stated.[17]

D'Alessandro responds to the overgeneration problem by stressing that in his definition, he explicitly says that the fictional propositions are *a subset* of the explicit propositions. Precisely, this subset which are reliably said by the narrator. I think this response is insufficient. First, it follows from this that a fiction in which the narrator was utterly unreliable would contain no "truths" at all. Indeed, suppose, for the sake of argument, that the narrator was such that everything he says is "false" in the fiction. Then, by definition, none of the explicit fictional proposition are "true" in the fiction. Although it seems very intuitive to say that one can easily get (some of) the "truths" in the story by negating the statements expressed by the narrator. D'Alessandro's view entails that one cannot do this.

Second, and less dramatically, it seems that recognizing that a narrator is (sometimes) unreliable implies that one recognises as "true" in the fiction some proposition which was not explicitly said. Indeed, my recognition of Humbert's unreliability is not my withholding my judgement about the expressed proposition that Lolita gave her consent but the plain recognition that she did not.

As for the undergeneration problem, D'Alessandro does not give any response. Here is his comment on Holmes's lacking a third nostril:

> The explanation for this, in my view, is not that implicitism is actually correct. On the contrary, I stand by the claim that "Holmes has two nostrils" is not true in the Holmes stories, in the proper sense of "true". (Of course it's not false in the stories either; it simply lacks a truth value.)[18]

I think this is beating around the bush. First, one may question what D'Alessandro means by "the proper sense of 'true'" here. If it means "explicitly true in the fiction",

[17] [Byrne 1993].
[18] [D'Alessandro 2016], §5.

then what he says is trivially true. If it denotes our intuitive conception of "truth in the fiction", then D'Alessandro is committed to the view that there is no contrast between, say, (1) and (2). This is a bad result for any theory of fictional "truth".

Curiously enough, D'Alessandro continues in the following manner:

> The correct explanation is rather that it's *reasonable for the reader to imagine*, or to *tentatively suppose*, that Holmes has two nostrils. The reader is entitled to imagine this because, for instance, that's the picture of Holmes that Conan Doyle presumably had in mind, and it's the one that Conan Doyle presumably wished for his readers to have in mind.[19]

This is precisely what implicitism is about: the reader needs to take into account *something else* than the set of fictional sentences to retrieve the set of fictional propositions, for instance they should consider the author's intentions. At this point, D'Alessandro's attachment to explicitism is purely verbal: what he calls "true in the fiction", everyone calls "explicit"; what he calls "reasonable to imagine", everyone calls "true in the fiction". What everyone calls the "problem of truth in fiction" corresponds to what D'Alessandro recognises as the problem of "what it is reasonable for the reader to imagine". D'Alessandro recognises here that explicitism is not a live option for the theorizing of reasonable imagining.

My description of the problem of truth in fiction makes it the case that the relevant data for the theorizing of fictional "truth" encompasses the reasonable inferences the reader draws in imagination. Hence, explicitism is not a live option.

Note on Retcons: D'Alessandro says he has a general argument against implicitism which leads him to reconsider explicitism. His argument relies on the possibility of retroactive continuity (retcon) in fiction. Retcon "is a literary device in which established facts in a fictional work are adjusted, ignored, or contradicted by a subsequently published work which breaks continuity with the former."[20] It so happens that a "fictional truth" in one fiction can be given *in another* fiction. D'Alessandro tentatively supposes that for any fictional "truth", one can imagine a relevant retcon which makes it the case that it is, in fact, not "true in the fiction". He further supposes that fictional "truth" is unalterable.

But his argument, as he remarks himself at the end of his article, is too powerful since it denies *both* explicitism and implicitism. It follows from his argument that there are *no* "truths" in any fiction. In which case, the above contrast is lost. It is not difficult to deny D'Alessandro's suppositions, though.

[19] [D'Alessandro 2016], §5.
[20] Wikepedia: Retroactive Continuity.

I will something to say about retcons and fictional "truth" at the end of this part, in §8.3, for I think they question the link between fictional "truth" and the reader's activity which is an interesting side issue about fictional "truth". I think this literary phenomenon points to a substantial qualification of the "authorial prerogative": this is the subject of the next subsection.

1.3.2 Intentionalisms

It is clear that the set of sentences explicitly written by the author are not sufficient to yield an adequate theory of fictional "truth". But it also seem clear that this set is necessary to get at what is "true in the fiction". Fictional "truth" has to depend, in some sense, on what the author actually wrote. Let us call this intuition the *authorial prerogative*. It states the obvious fact that the author has a crucial bearing on what is intuitively "true in the fiction", since they actually produced the explicit fictional propositions. The pre-theoretical intuition behind the authorial prerogative says that an author can make a proposition fictional by writing sentences expressing it in the story.[21] In this sense, it seems that fictional "truths" are a kind of truth by stipulation.

Explicitism can be seen as an untenable reduction of the authorial prerogative to the mere output of the writer's work, namely the set of fictional sentences. But, as D'Alessandro himself remarked, it is very likely that the author's intentions should matter. *Intentionalism* about fictional "truth" is the view that one should consider both what is explicitly said *and* the intentions of the author to determine what is "true in fiction". In other words, the author's intention is one of the guides to fictional "truth". Intentionalism thus counts as a kind of implicitism. Indeed, the author's intentions need not be explicit.

There is a wide variety of intentionalism, depending on one's theory of intentions, in which one should define what is characteristic of an authorial intention. There are also debates about what should count as an author in this view, whose intentions are relevant for fictional "truth". It is thus more accurate to talk about intentionalisms.

The debates over what is the best of intentionalisms are clearly open in the philosophical literature. Stock recently published a book in defense of what she calls "extreme intentionalism" ([Stock 2017]). Here is a glimpse of the debates when she describes her view against other forms of intentionalism:

> Extreme intentionalism is a variety of what is often called "actual author intentionalism". This characterises fictional content[22] in terms

[21] This is what is eloquently termed the "Say-so principle" in [Woods 1974].
[22] This corresponds to what we have called "fictional propositions".

> of what the actual author intended, in producing that fiction. Extreme (or actual author) intentionalism says that the authorial intention of a certain sort is both necessary (*has to be* present) for fictional content of a given kind *and* sufficient (its presence is *enough*) for fictional content of that kind.
>
> Despite near-universal opposition to it, at first sight, at least, extreme intentionalism has several things going for it. It is more streamlined and less ad hoc-looking than "modest" versions of intentionalism. On these latter sorts of view, an author's intentions are determinative of fictional content *only in certain circumstances*, but not always. For instance, they are determinative where they do not violate conventional sentence meaning, or where they do not otherwise fail. Where they do these things, some other feature determines meaning. Extreme intentionalism eschews such inelegant caveats.[23]

Stock then proceeds to define her preferred notion of intention which is adapted from Grice's notion of a reflexive intention. She gives a book-length argument against rival forms of intentionalism.

Although these debates are perfectly legitimate, I think discussing them in detail would lead us too far from the problem of fictional "truth" as defined above. I am interested in the inferential mechanism which yields the above contrast, given the set of fictional sentences expressing fictional propositions. As such, one should accept as true a form of intentionalism. But debates about the different versions of intentionalism have little bearing on the inferential mechanism proper. In other words, any form of intentionalism starts from the idea that we (readers of fiction) get contrasts like the above contrast and then proceeds in asking what is the underlying intentional structure relating us to the author in such a way that it makes us good and reliable at getting contrasts like the above one. By contrast, what this part is about are the specifics of such automatic inferences that we draw when plunged into the appropriate intentional structure. I will therefore remain neutral on the debates about intentionalisms.

In the philosophical literature, one can find two competing theories focusing on such a drawing of fictional inferences. Although many philosophers contributed to the discussion, it is fair to say that the two champions of the rival views are David Lewis and Kendall Walton.

[23] [Stock 2017], §1.2, p. 14.

1.4 Two incompatible solutions on the market

In this section, I will present the two philosophical solutions to the problem of fictional "truth" defined above. I will show that the most important distinction between the two theories is that one (the modal account) has it that fictional "truth" is a kind of truth, whereas for the other (the functional account) it is not the case. Consequently, from a metaphilosophical perspective, one claims that the problem of fictional "truth" is a problem for a general theory of truth whereas the other claims that it is irrelevant to such a general theory.

1.4.1 The modal account

The first theory has emerged from a general theory of truth, which was to account for our intuitions about modal sentences. The problem of truth in fiction was thus cashed out in the problem of giving adequate truth conditions for the sentences of the form "In F, ϕ", for F a given fiction, and ϕ a given sentence. So, according to the modal account, the contrast in truth-value between (1) and (2) corresponds to the contrast between:

(3) In Shakespeare's play, Hamlet is a human being.

(4) In Shakespeare's play, Hamlet is a crocodile.

[Lewis 1978] has been deeply influential in the development of this account. He suggested that the fictional operator "In F, ..." should be interpreted as a modal operator which selects the fictional worlds in which (1) is true and (2) is false.[24] In other words, "truth in fiction" is nothing more than *truth in some fictional world*. In his seminal paper, Lewis famously argues that we should analyse fictional "truths" within possible-world semantics.

Here is a paragraph from [Woodward 2011] in which he describes very clearly the theoretical intuition and the problems these theorists face:

> The thought is that when an author creates a representational artwork, they are *projecting* or *indicating* a fictional world, which we subsequently explore in our imagination when we engage with their work. To be fictionally true is to be true "at" the relevant fictional world.
>
> Whilst this imagery is attractive, simply appealing to fictional worlds leaves important questions unanswered. To illustrate this point, consider

[24] For the time being, I will not say anything about *how many* fictional world(s) are thus selected. This question will emerge in due time when we consider the "incompleteness problem", in §3.2.3.

> that merely appealing to fictional worlds neither tells us *what fictional worlds are* nor *how a work of fiction "indicates" a fictional world*. Just appealing to fictional worlds doesn't answer these substantial questions, which shows that the notion of "the world of the fiction" is opaque and explanatorily impotent. In order to explain truth-in-fiction in terms of fictional worlds, we need to know what fictional worlds are, and we need to know how a fictional world gets associated with a particular fiction.[25]

As Woodward aptly indicates, the basic idea of the modal account is both very appealing and problematic. In particular, he distinguishes between two important questions for the modal theorist: the ontological question ("What are fictional worlds?") and the epistemological question ("How does a work of fiction 'indicate' a fictional world?"). In chapter 4 and 5 I will show that the modal theorist cannot give a satisfactory answer to the ontological question, following two different argumentative strategies. As for the epistemological question, I will show that the modal theorists does not really address the question, for they do not need to care about interpretation when theorizing about fictional "truth". By contrast, the functional account is a theory of interpretation and thus provides a very detailed answer to the epistemological question.

1.4.2 The functional account

Some philosophers have criticised this idea that truth in fiction is yet another case of truth in a modal context. Walton famously argued in a series of articles and in his *opus magnum* [Walton 1990], that the "what-question" is fundamentally flawed. "What is true in F?" is not a good question, because there is no "truth" strictly speaking in a fictional context. When confronted with a contrast such as the above one, the theorist should refrain from interpreting the contrast as a difference in the *truth-value* of the two sentences, and thus transform (1) and (2) into (3) and (4). Rather, the good question is: "How does (1) get fictionalised (as opposed to (2)) in Shakespeare's play?" Walton gives an answer to this question by developing a general theory of fictionality based on the analysis of children's games of make-believe on the one hand, and *props*, that is the objects we use to represent such and such to be fictionally the case within the scope of games of make-believe, on the other hand. This account is called "functional" because, according to the theory, it is the *function* of novels, films, paintings (props in general) to generate fictional propositions, which we loosely call fictional "truths".

[25] [Woodward 2011], p. 158.

In [Friend 2017b], she nicely distinguishes between the two accounts in the following manner:

> Intuitively, what is fictionally true is whatever obtains in the "world of the story". Some philosophers take the intuition literally, spelling out fictional truth as truth at a set of possible worlds (most famously, [Lewis 1978]). This approach faces numerous challenges. In particular, by contrast with possible worlds, fictional worlds are typically incomplete – leaving many features indeterminate – and often impossible [Woodward 2011]. Perhaps these worries can be addressed. However, I prefer a functional account along the lines proposed by Kendall Walton ([Walton 1990]), for whom what is fictionally true, or simply *fictional*, is what a work of fiction prescribes that the readers imagine. This analysis is silent as to whether the content to be imagined should be specified by possible worlds or in some other way, focusing instead on the role of fictional truth in our engagement with stories.[26]

1.4.3 The two solutions are incompatible

In the above paragraph, Friend suggests that the two accounts coexist and that it is a matter of theoretical preference to go for one or for the other. However, I think we can safely argue that the two accounts are in fact incompatible so that only one is correct (if any).

Distinguishing between fictional "Truth" and truth

According to the modal theorist, (1) and (2) actually differ in their *truth-value*. According to the functionalist, (1) and (2) differ in *fictionality*. *Fictionality* is distinct from *truth*. Therefore, the two accounts make contradictory predictions about the truth-value of (1) and (2), hence about their semantic analysis. The important point is that they explain the contrast between (1) and (2) in incompatible ways.

Walton made it absolutely clear that *fictionality* is analogous but conceptually distinct from truth:

> Fictionality has turned out to be analogous to truth in some ways; the relation between fictionality and imagining parallels that between truth and belief. *Imagining aims at the fictional as belief aims at the true. What is true is to be believed; what is fictional is to be imagined.*

[26] [Friend 2017b], p. 2.

> [...]
> What we call truth in a fictional world is not a kind of truth. The phrase "In the world of the Unicorn Tapestries," preceding "a unicorn was captured," does not indicate in what manner or where or in what realm it is true that a unicorn was captured, or anything of the sort. *This is not true, period.* "It is believed (desired, claimed, denied) that p" is used not to assert that p is true but to attribute a different property to it, to assert that this proposition is believed, or that someone desires or claims or denies it to be true. Likewise, "It is fictional that p" and its colloquial variants attribute not truth but fictionality to p.[27]

Consequently, the problem of fictional "truth" is one of the problems for a general theory of truth according to the modal theorist. But for the functionalist, the "problem of fictionality" is not within the scope of a general theory of truth. In other words, according to the functionalist, fictional "truth" is a problem for the philosopher of fiction only.

Paraphrase and no-paraphrase

The other point of dissent between modal theorists and functionalists concerns the paraphrasing of (1) and (2) into (3) and (4). According to the modal theorist the paraphrase is merely a making explicit of the fictional sentences. The modal theorist is committed to the claim that (1) is "true in the fiction" iff (3) is true *simpliciter*. In other words, both the fictional sentence and its paraphrase express *the same* semantic content.

The functionalist denies that (1) and (3) have the same semantic content, hence that the latter is a proper paraphrase of the former. Indeed, whether (3) is true *simpliciter* is not obvious to the functionalist and needs to be explained (some have argued that it is not)[28]: the fictional operator "In F, ..." not being a straightforward modal operator for the functionalist. For the functionalist, when a sentence like (3) is uttered in an assertive tone of voice, it is true only insofar as (1) is fictional in the relevant fiction. This runs counter to the modal theorist's explanation following which (1) is fictionally "true" *because* (3) is true *simpliciter*.

Moreover, as indicated Friend quotation, the semantic content, for the functionalist, is a proposition *to be imagined*. By contrast, if (3) is true, then it is an assertion. But the semantic content of an assertion is *to be believed*. So even under the supposi-

[27] [Walton 1990], pp. 41-2.
[28] I present such a view in chapter 16.

tion that (1) and (3) share the same semantic content, our typical attitudes towards this content are different. Hence, (3) is not a good paraphrase of (1).

The two accounts are thus in contradiction about the alleged possibility to paraphrase fictional sentences using an overt fictional operator. As will appear in the discussion, the disagreement lies in the analysis of the fictional operator. The modal account claims that this operator should be on a par with other modal operators whereas the functionalist denies this.

1.5 Plan of action

I have defined the problem of fictional "truth" as that of explaining the contrast between (1) and (2). This is the central linguistic data which awaits semantic analysis. The debate between modal theorists and functionalists concerns this semantic analysis.

The remainder of this part is organised into two main sub-parts, each containing three chapters. The first sub-part is a *pars destruens*, in which I argue against the modal account; the second sub-part is a *pars construens*, in which I develop a version of functionalism which I call "pretence semantics".

Chapters 3, 4 and 5 together are thus a long argument against the modal account. In chapter 3, I present the modal account in more detail, presenting Lewis's account as the paradigmatic modal account against which I will argue. In chapter 4, I discuss what I call two "indirect arguments" against the modal account. They are "indirect" in the sense that they have the general form of a *reductio* argument, showing that if one assumes the central tenet of the modal account, one is led to run into troubles down the road. In chapter 5, I discuss what I call two "direct arguments" against the modal account. They are "direct" because they deny the suppositions of the modal account. By the end of chapter 5, we will have good reasons to consider the functional account.

Chapters 6, 7 and 8 together form a proposal to update pretence semantics as a leading functionalist account. In chapter 6, I present the Waltonian origin of pretence semantics, rehearsing the basic definitions and distinctions on which the formal system of pretence semantics is grounded. In chapter 7, I present the general framework which comes from Everett's recent work and discuss technicalities. In chapter 8, I apply the general framework to fictional discourses and show that the logical structure involved in fictional discourse is quite complex. I show the advantage of pretence semantics in that not only does it gives an adequate solution to the problem of fiction, but it also applies to other linguistic and narratological problems which are worth considering.

In chapter 9, I consider a side problem which concerns the implication of the reader as an essential piece in the problem of fictional "truth". I show how pretence semantics can avoid any commitment to the existence of a real author. I then show that it provides a general rational framework for devising arguments about what is "true" in the fiction. It appears, looking at legal issues surrounding fan-fictions, that such a rational framework could be useful in practice. This also paves the way for the following part in which I will discuss a case study of rational disagreement about fictional "truth".

Chapter 2

On modally minded theorists

2.1 Lewis on truth in fiction

2.1.1 Lewis's theory of counterfactuals

Lewis's central and original idea in [Lewis 1978] is that fictional sentences like (1) and (2) are of the same nature as a counterfactual statement like "Cleopatra's nose, had it been shorter, the whole face of the world would have been changed."[1] Intuitively, indeed, understanding a counterfactual statement consists, at least, in imagining a situation in which the antecedent is actually the case so that we can check whether the consequent is true. To illustrate this intuition, here is Quine's way of describing how we understand subjunctive conditionals in *Word and object*:

> We feign belief in the antecedent and see how convincing we then find the consequent.[2]

But imagining a situation in which the antecedent of a counterfactual is actually the case implies that we imagine *another world*, since the antecedent of a counterfactual, by definition, does not describe the actual world. Now, the definition of fiction reading also involves the imagining of a situation in which the fictional statements are true. And, by definition, a fiction may invite us to imagine *another world*. So it seems that fiction reading is an activity akin to the first part of the understanding of a counterfactual. Lewis's bringing the two phenomena of fiction reading and

[1] Translation from Pascal 1669 *Pensées*:

> Le nez de Cléopâtre, s'il eût été plus court, toute la face de la terre aurait changé.

[2] [Quine 1960], p. 222.

understanding of counterfactual sentences together is thus a very intuitive move. I take it that one should indeed explain the intuitive similarity between counterfactuals and fictional statements when theorizing about fictional "truth". Lewis's explanation is simple: they are the same.

The next step of Lewis's reasoning consists in applying his theory of counterfactuals to the semantic analysis of fictional sentences. Since he already had a theory which yields truth-conditions for counterfactual statements, applying it to fictional sentences should give an adequate theory of fictional "truth". Let us call this the *counterfactual analysis of fictional statements*. This next step is controversial and can be argued against, as we will see later on.

2.1.2 The counterfactual analysis of fictional statements

Lewis's analysis of counterfactuals is expressed in possible world semantics. A counterfactual statement ($\phi > \psi$: "If it were the case that ϕ, then it would be the case that ψ") has to be evaluated relative to the world of utterance of the counterfactual. So when Pascal produces a counterfactual statement about Cleopatra, his statement should be evaluated against the real world. Indeed, in another world where Cleopatra was not so central to the history of Western civilisation, for instance, the intuitive truth value of Pascal's counterfactual statement is not the same as in our world.

Informal truth conditions for counterfactual statements are the following:

- $\phi > \psi$ is true at w iff all the ϕ-worlds most similar to w are also ψ-worlds.

Lewis defines a *selection function* which captures the notion of similarity in the definition. The selection function takes a proposition p and a world w and returns the p-worlds most similar to w. Lewis famously argued that the selection function should follow at least these three constraints:[3]

- (Success) $f(w,p) \subseteq p$
- (Strong centring) $f(w,p) = \{w\}$ if $w \in p$
- (Uniformity) $f(w,p) \subseteq q \wedge f(w,q) \subseteq p \rightarrow f(w,p) = f(w,q)$

These constraints originate in an informal interpretation of the notion of "similarity" between worlds. The notion of similarity is very much debated in the philosophical literature and consequently, many philosophers have criticised these constraints. We

[3]Propositions are defined as sets of possible worlds, as usual.

need not enter these arguments here. One can find a more rigorous and detailed attempt to give an analysis of the similarity relation in [Lewis 1979a].

Given such a selection function, here is the formal definition of the truth-conditions for counterfactual statements:

$$[\![\phi > \psi]\!]_w = \{w : f(w, [\![\phi]\!]_w) \subseteq [\![\psi]\!]_w\}$$

So much for counterfactuals.

Lewis claims that fictional statements like (1) and (2) are the same as counterfactual statements and should thus be analysed in the same manner. But there is one immediate problem for this claim: fictional statements and counterfactuals do not seem to have the same logical form. Indeed, a counterfactual statement $\phi > \psi$ is a complex formula which is easily cut into two articulated propositions which are traditionally called the *antecedent* and the *consequent* of the counterfactual. Fictional statements, on the other hand, appear in one piece.

To handle this problem, Lewis postulates a covert "fictional operator":

> Let us not take our descriptions of fictional characters at face value, but instead let us regard them as abbreviations for longer sentences beginning with an operator "In such-and-such fiction ...". Such a phrase is an intensional operator that may be prefixed to a sentence ϕ to form a new sentences. But the prefixed operator may be dropped by way of abbreviation, leaving us with what sounds like the original sentence ϕ but differ from it in sense.[4]

The detailed analysis of the fictional operator is the concern of Lewis's article and all its unforeseen developments.

In order to guide him in this definition, Lewis states the general idea very clearly which fastens the similarity between fictional and counterfactual statements. Here is Lewis's idea in his own words:

> ... we might proceed as follows: a prefixed sentence "In the fiction f, ϕ " is true (or, as we shall also say, ϕ is true in the fiction f) iff ϕ is true at every possible world in a certain set, this set being somehow determined by the fiction f.[5]

Hence, just as there is a selection function whose role in the semantic analysis of counterfactuals is to select the worlds at which the antecedent is true, there should

[4] [Lewis 1978], p. 37-38.
[5] [Lewis 1978], p. 39.

be a selection function whose role in the semantic analysis of fictional statements is to select the worlds at which all the "truths" in that fiction are truths *simpliciter*. Evaluating a fictional sentence amounts to checking whether that sentence is true *simpliciter* in the fiction-selected worlds, just as evaluating a counterfactual amounts to checking whether the consequent is true in the antecedent-selected worlds.

So Lewis should say something about how we can get the relevant selection function from the fictional text. At this point Lewis considers and dismisses the idea that it should be the simple conjunction of all the explicit fictional sentences constituting the story. This is, in essence, explicitism as discussed above and Lewis gives the traditional counterarguments to such a view. Lewis's own proposal to avoid the problems of explicitism is to appeal to the act of storytelling:

> The worlds we should consider, I suggest, are the worlds where the fiction is told, but as known fact rather than fiction. The act of storytelling occurs, just as it does at our world; but there it *is* what here it falsely purports to be: truth-telling about matters whereof the teller has knowledge.[6]

The remainder of the article is an attempt to formalise this informal solution to the problem of fictional "truth".

Following Lewis's idea, we now have a general recipe to rewrite any fictional statement in a counterfactual paraphrase. First, one makes overt the covert fictional operator so that (1) and (2) become (3) and (4). Then, one develops the fictional operator in the manner presented above so as to get:[7]

(5) If Shakespeare's *Hamlet* were told as known fact rather than fiction, Hamlet would be a human being.

(6) If Shakespeare's *Hamlet* were told as known fact rather than fiction, Hamlet would be a crocodile.

Both (5) and (6) have the logical form of a counterfactual. So one can give their truth-value, applying Lewis's theory of counterfactual. (5) is true because all the worlds in which *Hamlet* is told as known fact and Hamlet is a human being are closer to the worlds in which *Hamlet* is told as known fact and Hamlet is not a human being. (6) is false because there is a world in which *Hamlet* is told as known fact and Hamlet is not a crocodile which is closer to the worlds in which *Hamlet* is told as known fact and Hamlet is a crocodile, namely a world in which Hamlet

[6] [Lewis 1978], p. 40.
[7] This corresponds to Lewis's *Analysis 0*.

is a human being. Given any reasonable definition of similarity across worlds, such reasoning holds. Consequently, (1) is "true in *Hamlet*" and (2) is "false in *Hamlet*".

There are problems and debate about what "told as known fact" precisely means here, which are the object of Lewis's discussion of his own Analysis 1 and 2.[8] I need not go into the detail here, for the objections to the modal account apply to the counterfactual account of fictional sentences in general.

2.2 On the kinds of counterarguments to the modal account

2.2.1 Some terminology

A modal account of fiction is a theory which claims that fictional "truth" is a kind of modal truth; hence, that the semantic analysis of fictional statement is akin to the semantic analysis of modal statements. Given that it is very common to give the semantic analysis of modal statements within the possible-world framework, modal theorists usually try to give a semantics for fictional statements using possible worlds. Lewis's counterfactual account of fictional statements is a paradigmatic example of a modal account, since counterfactual statements are a particular kind of modal statements.

The arguments given below target all modal accounts, and especially criticise the idea of using possible worlds to give a solution to the problem of fictional "truth". So Lewis is only a paradigmatic target for such arguments.[9] It will be useful to have his theory in mind for the developments, though.

2.2.2 Indirect and direct arguments

The modal account faces two well-known objections in the literature. Both objections focus on the fact that possible-world semantics is too strong for the analysis of truth in fiction. The possible worlds of possible world semantics are necessarily *complete* and *consistent*. But fictions are generally thought of as often *incomplete*

[8] [Lewis 1978], p. 42; p. 45.

[9] At least, Lewis's theory as I have presented it. For the record, my rendering of Lewis's theory of "truth in fiction" is a simplified version which may not be very charitable to Lewis. Indeed, I crucially avoid all mention of pretence which is, arguably, at the core of Lewis's theory. Consequently one could argue that I discuss a shallow version of Lewis. If that is the case, I hope Lewis would have forgiven my thus shallowing him.

and sometimes *inconsistent*. Here are two examples of these classic objections. In [Currie 1990], p. 54, Currie writes:

> Possible worlds are *determinate* with respect to truth; [...] they are *consistent* [...] but fictional worlds are always indeterminate and sometimes inconsistent.

In [Walton 1990], p. 64, Walton writes:

> [...] fictional worlds are not possible worlds. Two differences, especially, have been discussed elsewhere: fictional worlds are sometimes impossible and usually incomplete, whereas possible worlds (as normally construed) are necessarily both possible and complete.

Therefore, so the argument goes, the standard possible world semantic framework is not fit for fictions.[10]

I call these objections "indirect arguments". Indeed, the structure of the arguments is that of a *reductio*: one assumes that the central claim of the modal account is correct so that we can use possible-world semantics, and then one shows that possible-world semantics is not fit for the job. As a result, one concludes that the modal account is not a good theory of fictional "truth". Such arguments have received an enormous amount of attention in the literature, since they have been replied to in various ways. Interestingly, to reverse Smart's saying, it appears that one's *modus tollens* is another's *modus ponens*. Indeed, some philosophers, in order to meet these indirect arguments, have taken on board the fact that fictional worlds are neither consistent nor complete by extending the standard possible-world framework into including incomplete, inconsistent worlds. These are called "impossible" or "non-normal" worlds. They are still modal theorists about fictional "truth", but they can be called non-normal modal theorists, since they accept non-normal worlds into the picture. In chapter 4, I will engage with non-normal modal theorists.

But one can also produce "direct arguments" against the modal account. Such arguments deny the central tenet of the modal account, namely that fictional statements are a kind of modal statements. These arguments are very rare in the philosophical literature. In chapter 5, I will argue that although it is true that fictional statements are similar to counterfactuals, we need to have a fiction-first perspective on this similarity. In other words, I think that it is true that counterfactuals and

[10] I will use the term "incompleteness" and not "indeterminacy" in this part. Nothing hinges on this choice of words. The reason why I distinguish between these terms is because I will use the term "indeterminacy" to denote another phenomenon discussed in part 2.

fictional sentences should share characteristic semantic features, but I think counterfactuals should be theorised after a proper theory of fictional "truth", not the other way around. Such direct arguments will motivate the alternative account of fictional "truth", namely the functional account.

2.3 Aside on the literary minded modal theorists

Before I argue against the modal account, I want to dispel a possible confusion about modally minded theorists. What I have said so far might suggest that the modal account is designed for logically minded philosophers and formal linguists since possible-world semantics has become a second nature to them. This is not true. Many literary theorists went for the modal account too.

In [Lavocat 2010b], Lavocat gives a detailed overview of the main contributions and interests of literary theorists when it comes to the notion of possible worlds. Three major figures of literary theories stand out by using a version of the modal account: Thomas Pavel especially in [Pavel 1975] and [Pavel 1986], Umberto Eco in [Eco 1979] and Marie-Laure Ryan in [Ryan 1991]. All used a theory of fictional "truth" using possible worlds and went on theorizing about narratological problems using this new theoretical framework. Following these pioneers, many other literary theorists followed in adopting a version of the modal account. The recent collaborative work [Lavocat 2010a] shows that the modal account is still living in the literary studies.

Pavel's reasons for adopting a modal account of fictional "truth" are interesting to spell out, for he was deliberately trying to apply the logical framework to literary problems, thus putting forward the idea of literary criticism as applied philosophy. One of his original motivation was constructive. He thought structuralist approaches to narratology became too stiff in the seventies and took possible-world semantics to be a good candidate for giving a semantics to the syntactically minded approach of structuralist narratology. Another motivation is polemical. He wanted to use possible-world semantics to argue against historicism whose main contender is that the interpretation of fictional texts was first and foremost given by the knowledge of the author's historical background. Possible worlds, indeed, do not depend on historical features: they represent some abstract semantic content which is independent of the circumstances of utterance of modal statements.

Apart from literary theorists, some novelists also were interested in the modal account. Milan Kundera, for instance, is a very eloquent advocate of the modal account. In several places, both in his fiction and non-fiction, Kundera argued that *fictionalia are possibilia*, that is the idea that one should think of the fictional as a

subset of the possible. This claim is, in essence, the modal account. Here is, for instance, a excerpt from *The unbearable lightness of being*, part 5, ch. 15:

> And once more I see him the way he appeared to me at the very beginning of the novel: standing at the window and staring across the courtyard at the walls opposite.
>
> This is the image from which he was born. As I have pointed out before, characters are not born like people, of woman; they are born of a situation, a sentence, a metaphor containing in a nutshell a basic human possibility that the author thinks no one else has discovered or said something essential about.
>
> But isn't it true that an author can write only about himself?
>
> Staring impotently across a courtyard, at a loss for what to do; hearing the pertinacious rumbling of one's own stomach during a moment of love; betraying, yet lacking the will to abandon the glamorous path of betrayal; raising one's fist with the crowds in the Grand March; displaying one's wit before hidden microphones – I have known all these situations, I have experienced them myself, yet none of them has given rise to the person my curriculum vitae and I represent. The characters in my novels are my own unrealised possibilities. That is why I am equally fond of them all and equally horrified by them. Each one has crossed a border that I myself have circumvented. It is that crossed border (the border beyond which my own "I" ends) which attracts me most. For beyond that border begins the secret the novel asks about. The novel is not the author's confession; it is an investigation of human life in the trap the world has become. But enough. Let us return to Tomas.

Therefore, one should not think that the modal account is a mere logistic frenzy. On the contrary, it should be thought of as intuitive and appealing. However, I shall now argue that it does not give an adequate solution to the problem of fictional "truth".

Chapter 3

Indirect arguments against the modal account

3.1 On the two kinds of modal theorists in the literature

The two arguments start from the intuitive observation that the set of fictional worlds, if there is such a thing, cannot be a subset of the possible worlds as defined in possible world semantics. Indeed, it seems that fictional worlds do not behave and they cannot be regimented as logicians have done with possible worlds. In other words, the modal account as defined above entails that the set of fictional worlds is a subset of the set of possible worlds. But there are counterexamples: some fictional worlds are not possible worlds.

Faced with such an observation, there are two attitudes a modal theorist can take. The *hot-head* will answer: "No problem!" Just admit as possible worlds these worlds which are incomplete and inconsistent, and call them non-normal worlds. So fictional worlds are a subset of the union of normal and non-normal worlds. By contrast, the *cool-head* will first and foremost try and construct the problematic fictional worlds using set-theoretical constructions of possible worlds. If such construction is possible, one can see how fictional worlds reduce to possible worlds.

The hot-head thus holds the thesis that the set of fictional worlds is a subset of the set of the union of normal and non-normal worlds. The cool-head, by contrast, holds that the set of fictional worlds is a subset of the powerset of the set of possible worlds. In the philosophical literature, it is usually understood that the hot-head carry the day by putting extremely inconsistent fictions on the table and setting up

the stage so that the only way out is buying the whole lot of impossible or non-normal worlds.[1] I will show in this chapter that the cool-head is actually in a better position than usually acknowledged in the literature. The take-home message of this chapter will be that one can go further than expected solving the incompleteness and inconsistency problem using possible worlds only.

First, I will present the two problems and show that they are two sides of the same formal coin. Then I will show that the cool-head strategy can provide dual solutions using the logical machinery of supervaluations and subvaluations. I will note that this nicely shows how inconsistent fictions can be thought of as a double-bind phenomenon. Finally, I will discuss die-hard inconsistent fictions which force one to adopt the hot-head strategy in response. I will show that these die-hard cases also prompt a general argument against non-normal worlds accounts of fictional "truth". This will lead me to the conclusion that the non-normal account of fictional "truth" is a hopeless response to the two indirect arguments.

3.2 Dual solutions for dual problems

3.2.1 Incompleteness

Stories are intuitively leaving many things unspecified. For instance, although Hamlet has blood in his veins, we do not know what Hamlet's blood group is.[2] It also happens that we could not know Hamlet's blood group, since such a piece of information does not follow from any fictional proposition in *Hamlet*. One straightforward rendering of this intuition is to say that neither of the following fictional sentences are "true in *Hamlet*":

(7) Hamlet is of blood group O.

(8) Hamlet is not of blood group O.

But in possible world semantics, it is true by definition that in each world, every

[1] See [Priest 1997] for a paroxysmal example of such a strategy.

[2] In Act V, scene ii, Hamlet is wounded in a duel and bleeds. So that Hamlet has blood in his vein is an explicit fictional proposition. Some people want to deny that Hamlet has a blood group, although he has blood in his veins, because blood groups were not known around 1600 when Shakespeare wrote the play. Latour gave a similar argument leading to the conclusion that the Pharoah Ramses II cannot have died from tuberculosis, despite the biological evidence found in the examination of his mummy, because the *bacillus virus* was not known back in the days of Ancient Egypt. If you have Latourian intuitions, please change the example to something intuitive for you.

proposition is either true or false.[3] Formally, given a Kripke model $K =< W, R, v >$, for all proposition p:[4]

(Completeness) $\forall w \in W,\ v_w(p) = 1 \vee v_w(\neg p) = 1$

Because of (Completeness), if fictional worlds are a subset of possible worlds, one cannot have it that neither (7) nor (8) is true in *Hamlet*.[5]

3.2.2 Inconsistency

Some stories contain inconsistencies. For instance, in *The Adventures of Sherlock Holmes*, Conan Doyle "placed" Watson's war wound on his shoulder at one point of the story and on his knee at another, while it is "true" in the story that Watson has only one war wound. (This is an accidental inconsistency, so to speak, but some stories are essentially inconsistent. I will discuss this distinction later on.) We can rephrase the situation to make the inconsistency blatant: it is explicitly "true in *the Adventure of Sherlock Holmes*" that Watson has a wound on his shoulder and has not a war wound on his shoulder. If this is the correct way of describing the situation, then both the following fictional sentences are "true in *the Adventures of Sherlock Holmes*":

(9) Watson has a wound on his shoulder.

(10) Watson has no wound on his shoulder.

But in possible-world semantics, it is true by definition that in each world, a proposition and its negation cannot be true together. Formally, given a Kripke model $K =< W, R, v >$, for all proposition:

(Consistency) $\forall w \in W,\ v_w(p) = 1 \leftrightarrow v_w(\neg p) = 0$

[3] Interestingly, this is not the case in *situation semantics* as presented in [Barwise and Perry 1983], for situations are *parts* of possible worlds and thus crucially not complete. Thanks to François Récanati for pointing this out to me. I think it is akin to Lewis's original analysis in terms of sets of possible worlds: both miss the duality of the problems, which is, I argue here, remarkable.

[4] With W the set of all possible worlds, R the accessibility relation such that $R \subset W \times W$ and v an interpretation function for well-formed formulae. The notation is standard for possible world semantics and can be found, for instance, in [Priest 2008a].

[5] Here is how it is very eloquently put in [Lorand 2001], p. 432:

> The "gap" metaphor complements the "world" metaphor, since gaps are determined by the implied comparison between the "incomplete" report and the complete world.

Because of (Consistency), that such inconsistent fictions cannot be modeled by a possible world.

3.2.3 Dual problems

(Completeness) and (Consistency) are actually two sides of the same coin. Indeed, given bivalence, they are formally equivalent.

(Bivalence) $v(p) = 1 - v(\neg p)$

Moreover (Bivalence) holds at each world, since Kripke models are an extension of classical logic.

Here is the proof of the equivalence:

(Consistency)

$\equiv \forall w \in W,\ (v_w(p) = 1 \wedge v_w(\neg p) = 0) \vee (v_w(p) \neq 1 \wedge v_w(\neg p) \neq 0)$

$\equiv \forall w \in W,\ v_w(p) = 1 \vee v_w(p) = 0$ (using (Bivalence))

$\equiv \forall w \in W,\ v_w(p) = 1 \vee v_w(\neg p) = 1$ (using (Bivalence))

$\equiv$ (Completeness)

I think this duality is quite remarkable and calls for dual solutions.

Modal theorists, by contrast, do not provide dual solutions. Indeed, in Lewis's original account, there is technically no problem about incompleteness: the solution is hardwired in his semantics, so to speak. According to Lewis's analysis, a proposition is fictional (or "true in fiction") iff *all* the possible worlds in which the fiction is told as known fact and where the proposition holds *are* closer to the actual world than all the possible worlds in which the fiction is told as known fact and where the proposition does not hold. So, by Lewis's standard, a fiction selects a *set* of possible worlds.[6] So when a fictional sentence like (7) is such that it is not true in *all* the selected worlds, Lewis's prediction is that such a proposition is not "true in the fiction"; the same reasoning applies to (8).

The only problem for Lewis, then, is the inconsistency problem. He provided the following fixes for inconsistent fictions:

[6]This comes from Lewis's rejection of the "Uniqueness Assumption" in [Lewis 1973] when arguing against the principle of conditional excluded middle, according to which it ought to be the case that $(\phi > \psi) \vee (\phi > \neg\psi)$. See [Stalnaker 1980] for a defence of the uniqueness assumption.

> Perhaps we should take the maximal consistent fragments, obtained by deleting the bare minimum that will give us consistency. [...] what do we do with our several consistent fragments (or corrections) when we have them? [...]
>
> I suggested this *method of intersection*: ϕ is true in the original fiction iff ϕ is true in every fragment. Now I would favor instead this *method of union*: ϕ is true in the original fiction iff ϕ is true in some fragment. (Not that we need choose once and for all—we can have both methods, distinguishing two senses of truth in inconsistent fiction.)[7]

These fixes are, in essence, what should be done for *both* problems, as I will show below. Singling out the inconsistent problem for special fixes is, I take it, very inelegant in light of the shown duality of problems. As Phillipps put it:

> A better account would be a unified theory which supplied truth conditions for every fiction, rather than singling out inconsistent fictions for special treatment.[8]

This is what I will now do.

3.3 Doing the impossible with possible worlds

Let us start with the very plausible idea that a story selects a set of possible worlds. The story, because it is "told as known fact", can be seen as a report of the facts that belong to a certain world. But such a report is crucially partial. So the very same facts can be found in many different possible worlds. These are the possible worlds under discussion.[9]

One can thus see each possible world in this set as a complete specification of the world of the story. For instance, there are two ways of specifying the world of *Hamlet* when it comes to the question whether Hamlet is of blood group O or not. So in the set of worlds in which *Hamlet* is told as known fact, there is a world in which Hamlet is of blood group O and another in which he is not of blood group O.

But inconsistencies prevent from thus selecting possible worlds. Indeed, there is no possible world in which Watson both *has* and *has not* a wound on his shoulder. Following Lewis's advice, one should "delete" the contradiction. Here is the deletion procedure: Thanks to Christopher Badura from pointing this out to me.

[7] [Lewis 1983], p. 276.
[8] [Phillips 1999], p. 281.
[9] See [Currie 1990] for a similar argument.

- Extract a contradiction, which is of the form $p \wedge \neg p$.
- Run one specification of the story with p.
- Run another specification of the story with $\neg p$.
- Store p in D, the set of deleted propositions.
- Apply this procedure for each contradiction.

I suppose that one can extract a contradiction of the form $p \wedge \neg p$ from what is read. In the case of Watson's war wound, one need to *infer* the contradiction from the fact that Watson is said to have one single wound and that it is said to be in two different places. This inference is indeed presupposed.[10] In practice, though, one should be able to "spot" the contradiction otherwise, one is simply failing to recognise the story as inconsistent. Theorizing about this "spotting" of inconsistencies in sets of propositions is a somewhat different issue, I take it. The only thing I could say, when it comes to inconsistent fictions, it is sometimes difficult to extract a contradiction from an inconsistent story and this is surely part of the fun.[11]

One can thus define in full generality the set of fictional worlds targeted by a story. Let us define:

$$F = \{w | @ R_f w\}$$

with @ designating the actual world, and R_f the accessibility relation on the possible worlds given the fiction f (w is "accessible given f" whenever it is targeted by the fiction f).

In this set of worlds, one can define a supervaluation and a subvaluation function. Supervaluation is a way of modeling incompleteness[12], while subvaluation is a way of modeling inconsistencies.[13] They are dual logical techniques.[14] As I said above, this is a way of fleshing out Lewis's methods of union and intersection.

Oddly enough, Lewis indicates these methods to deal with the inconsistency problem only, thus failing to remark the duality of the two problems made explicit above. Hanley, one of Lewis's followers, did spot the duality in question though:

[10] Thanks to Christopher Badura for pointing this out to me.

[11] For instance, in the trilogy *Matrix*, there are several sophisticated inconsistencies which require some work to be extracted: take for instance Smith's way of "hacking" into reality to try to kill Neo trough Bane in *The Matrix Revolutions*.

[12] See especially [Fine 1975] and [Kamp 1981], among many others.

[13] See especially [Hyde 1997] and [Varzi 2001], among many others.

[14] See [Ripley 2013] and [Cobreros et al. 2012] for philosophical perspectives on this duality and its connection to a good definition of logical consequence.

> ... the shift to the method of union actually does shed some light on the nature of fictional truth, by showing the relationship between indeterminacy and inconsistency. Truth in fiction is not truth-at-a-world, but rather truth at a *set* of worlds.[15]

In the following, I present an extension of Lewis's theory which is essentially equivalent to Hanley's. I add two minor points, though: first, talking about subvaluation points towards the crucial idea of a fictional contradiction as a double-bind phenomenon, as made explicit below; second, I give a simple algorithm to fine-tune the interaction between the two strategies into one single notion of fictional "truth".

3.3.1 Supervaluation

Let us enrich Kripke models with a supervaluation function s^+: $K^+ =< W, R, v, s^+ >$. One can define "supertruth" over a set of possible worlds F in the following manner:

- $s^+_F(p) = 1$ iff $\forall w \in F$, $v_w(p) = 1$
- $s^+_F(p) = 0$ iff $\forall w \in F$, $v_w(p) = 0$
- p is *indeterminate* otherwise.

In such a model, it is easy to model that neither (7) nor (8) is "true in *Hamlet*".

Indeed, in the set of possible worlds where *Hamlet* is told as known fact F_{Hamlet}, there is at least one world in which Hamlet is of blood group O and at least one world in which Hamlet is not of blood group O. So by the above definitions, we get:

$s^+_{F_{Hamlet}}$(7) is *indeterminate*.

$s^+_{F_{Hamlet}}$(8) is *indeterminate*.

3.3.2 Subvaluation

Let us enrich our Kripke model with a subvaluation function s^-: $K^- =< W, R, v, s^- >$ One can define "subtruth" over a set of possible world F in the following manner:

[15] [Hanley 2004], p. 117. In the previous paragraph, he also makes explicit the connection between these problems about fictional "truth" and the Stalnaker-Lewis debate about Conditional Excluded Middle as made explicit above. He is, as far as I know, the only philosopher to have made this explicit in publication. Maybe it was too obvious for the community back then to stress this point.

- $s^-_F(p) = 1$ iff $\exists w \in F$, $v_w(p) = 1$
- $s^-_F(p) = 0$ iff $\exists w \in F$, $v_w(p) = 0$

In such a model, it is easy to model that both (9) and (10) are "true in *the Adventures of Sherlock Holmes*".

Indeed, in the set of possible worlds where *the Adventures of Sherlock Holmes* are told as known fact $F_{Sherlock}$, there is at least one world in which Watson has a wound on his shoulder and at least one world in which he has no wound on his shoulder. So by the above definitions, we get:

$$s^-_{F_{Sherlock}}(9) = 1.$$

$$s^-_{F_{Sherlock}}(10) = 1.$$

However, note that:

$$s^-_{F_{Sherlock}}((9) \wedge (10)) = 0$$

3.3.3 Combining the two strategies

We now have two ways of evaluating a sentence. As a general rule, I propose that we use the supervaluation function whenever no "deletion" was recorded. When evaluating a deleted proposition, one should favour the subvaluation function. This presupposes we have to keep track of the deletions, as made explicit in the deletion procedure above.

One can thus define a general notion of fictional "truth" $FT(p)$ in $K^{\pm} =< W, R, v, s^+, s^- >$ which avoids both the incompleteness and the inconsistency problem.

$$FT(p) = \begin{cases} s^+_F(p), & \text{if } p \notin D \\ s^-_F(p), & \text{otherwise.} \end{cases}$$

3.3.4 The revenge of inconsistent fictions

Essentially vs accidentally inconsistent fictions

The above definition of fictional "truth" arguably runs into trouble when it comes to *essentially* inconsistent fictions. Indeed, it is useful to distinguish between accidentally inconsistent fictions and essentially inconsistent fictions. *The Adventures*

of Sherlock Holmes are *accidentally* inconsistent, in the sense that Conan Doyle obviously forgot about Watson's war wound when he wrote about it in two places. In some other fictions, authors write inconsistencies *on purpose*. These essentially inconsistent fictions are usually taken to force one to accept the hot-head strategy.

Priest's reading test

[Priest 1997] is the classic example of such a fiction.[16] Graham Priest tells the story of his visiting Richard Routley-Sylvan's house soon after he died in 1996 with Nick Griffin who was Routley-Sylvan's literary executor. In putting some order in the house, Priest finds "Sylvan's box" which is an impossible object. Here is the first description of the box:

> At first, I thought that it must be a trick of the light, but more careful inspection certified that it was no illusion. The box was absolutely empty, but also had something in it.[17]

Later on Priest adds:

> Looking in the box was something like that: the experience was one of occupied emptiness. [...] The box was really empty and occupied at the same time. The sense of touch confirmed this.[18]

Priest then claims that the modeling of such a story requires that we add impossible worlds to the picture. See his fifth moral:

> There are, in some undeniable sense, logically impossible situations or worlds. The story describes (or at least, partially describes) one such.[19]

In my opinion, Priest's conclusion is a little hasty. Indeed, it seems that we can handle Sylvan's box with subvaluation. Take Priest's own rendering of what is "true" in the story, before he draws any conclusion:

> [...] let us take an old-fashioned comprehension test on the story.
>
> *Question 1*: In which country did the meeting take place? *Answer*: Australia.

[16] There are many other examples. In §12.2, I give many other examples created by real authors, not by a hoot-headed philosopher.
[17] [Priest 1997], p. 575.
[18] [Priest 1997], p. 576.
[19] [Priest 1997], p. 580.

> *Question 2*: Was Richard at the farmhouse? *Answer*: No.
>
> *Question 3*: Was the box empty? *Answer*: Yes and no.
>
> *Question 4*: How many times did Nick leave the property? *Answer*: Once.
>
> *Question 5*: Was the box shot off to the moon at the end of the story? *Answer*: No.
>
> Other answers are wrong, or in the case of Question 3, at least incomplete.[20]

Interestingly, *FT* gets all the answers right.

Despite this fact, Priest asserts in his fourth moral:

> Nor can we simply break the information up into (maybe maximally) consistent chunks and infer from each of these. If we could, we would have to infer that the character were astonished by the fact that the box was empty and/or by the fact that it had a figurine in it, which they most certainly were not.[21]

I simply deny this, for the following reasons:

- The characters were astonished (supertrue, hence fictionally true) because the box was full *and* empty (subtrue, hence fictionally true).
- The characters were astonished because the box was full (subtrue, but not fictionally true).
- The characters were astonished because the box was empty (subtrue, but not fictionally true).

Inconsistent fictions as double-binds

I think the disagreement with Priest points towards a very interesting phenomenon which I will try to explain.

One very intuitive interpretation of what is happening with *FT* is that it models inconsistent stories as double-bind phenomena. This view is, I think, very appealing.[22] It says that whenever a contradiction is fictionally "true", one can break up

[20] [Priest 1997], p. 579.

[21] p. 580.

[22] Note that double binds happen in many places: in legal systems as well as in computer programs. In such cases, one might also refrain from accepting true contradictions into the picture without trying something else.

the contradiction into two prescriptions to imagine. It fits nicely with the functional account to be discussed for which a proposition p is fictionally "true" iff there is a work of fiction which prescribes to imagine p. As for impossibilities, as Walton remarks:

> Can one imagine impossibilities? Not, presumably, if imaginability is a good test for possibility. But then can contradictory or metaphysically impossible propositions be fictional, on our account? I am inclined to think that even contradictions can be imagined in the relevant sense. But our understanding of fictionality is safe even if they cannot be. There can be prescriptions to imagine a contradiction even if doing so is not possible. (A badly drafted law might require one to do something and also to refrain from doing it.) There may also be separate prescriptions to imagine p and to imagine not-p, without a prescription to imagine their conjunction. The set of propositions fictional in a given world might be inconsistent even if no contradiction is fictional.[23]

The cool-head can thus use this idea to defend that, in practice, all inconsistent stories imply *distinct* contradictory injunctions.

This idea is naturally formalised so as to yield a non-adjunctive logic. A non-adjunctive logic is a logic in which the following rule fails:

$$A, B \vdash A \wedge B$$

FT, because it is has a subvaluationist feature, is non-adjunctive.[24]

Therefore, in the case of Watson's wound, indeed, FT predicts that both (9) and (10) are fictionally "true" without (9 $\wedge$ 10) being true. In the case of Sylvan's box, FT has it that the box is full and it is empty, but it is not fictionally the case that the box is full *and* empty.

Sylvan's *fempty* box

I think Priest is precisely denying that fictional inconcistencies should be modeled as a double-bind phenomenon. But his reading test does not do the trick.

It is arguably possible to think of a fictional inconsistency which resists being deleted and cut into consistent fragments. What if Sylvan's box was full-and-empty as a single requirement? Using an elegant neologism from [Badura and Berto 2019],

[23] [Walton 1990], pp. 65-6.

[24] A logical consequence relation based on subvaluations is non-adjunctive. See [Hyde 1997] on this.

let us call this "being fempty". One would argue that *Sylvan's box* prescribes imagining that Sylvan's box is fempty.

Fempty is a *simple*, *contradictory* predicate. As such, it cannot be deleted. Indeed, there is no way of extracting a contradiction of the form $p \wedge \neg p$ from "Fempty(b)". Moreover, there is no possible world in which "Fempty(b)" is true. Therefore, truth in this story cannot be modeled by FT.

The argument is indeed valid, although I want to question the truth of the premise which says that the story *Sylvan's box* requires the reader to imagine that Sylvan's box is fempty. If that is the case, then *Sylvan's box* is even stranger than it seems. Indeed, it introduces a new predicate, namely "fempty", which means the same as "both full and empty" without being a complex predicate.[25] My intuition is that nothing like this is going on in reading *Sylvan's box*. This requirement is indeed barely comprehensible, whereas *Sylvan's box* is a perfectly comprehensible story, albeit inconsistent. I suggest we should rather think of the inconsistency as a double-bind phenomenon as explained above.

At this point, I think we can appreciate Lewis's remark that:

> [...] where we have an inconsistent fiction, there also we have several consistent fictions that may be extracted from it. (Perhaps not in the very hardest cases—but I think those cases are meant to defy our efforts to figure out what's true in the story.)[26]

This shows that one can go much further than expected using the cool-headed attitude towards fictional inconsistencies.

3.4 Impossible worlds and their limitations

Although I doubt that die-hard inconsistent fictions *actually* exist, let us suppose, for the sake of the argument, that some do. Suppose, for instance, that *Sylvan's box* makes it fictional that Sylvan's box is fempty. Then one is led to accept that there are impossible, or non-normal worlds.

[25] Now it strangely sounds like Goodman's introduction of "grue" and "bleen" which he would take to be logical simples out of which one can define "blue" and "green". Although no formal contradiction in involved in Goodman's thought experiment.

[26] [Lewis 1983], p. 276.

3.4.1 The unrestricted non-normal world view

There are several ways of introducing impossible worlds.[27] I will follow Berto here, for he defends a version of impossible-world semantics which is designed to model imaginative attitudes in general (and thus in particular fictional imagination) in two recent papers: [Berto 2017] and [Badura and Berto 2019].[28]

The basic idea is to extend Kripke models so that there are two sets of worlds, the normal and non-normal ones: $K =< P, I, R, v >$, with $W = P \cup I$ (P represents the set of normal worlds, and I the set of non-normal worlds, W being the set of all worlds). In the normal worlds, formulae behave as usual and one can give recursive truth-conditions for complex formulae as usual. In the non-normal worlds, complex formulae behave like atomic formulae so that they can be attributed truth-values arbitrarily.

Consequently, one can arbitrarily violate any truth of logic one wants. Indeed, there is a non-normal world in which the complex formula $A \wedge B$ is true, but in which neither A nor B is true. In Priest's story, ex hypothesis, the introduction of "fempty" does exactly this. In the story, the box is fempty but it is neither full nor empty.[29] As for indeterminate "truths" in fiction, one should go in non-normal worlds in which bivalence fails.

Of course, one can group non-normal worlds as one wants, depending on how much they violate known laws of logic, that is depending on how much logical structure they retain. For instance, Berto distinguishes between "intentional non-normal worlds" in which only modal formulae get treated like atomic formulae; and "extensional non-normal worlds" in which all complex formulae get treated like atomic formulae.

Berto's contribution consists in defining an "imagination operator" $[A]B$ which reads: "It is imagined in act A that B; or, less briefly and more accurately: It is imagined in the act whose explicit content is A, that B". The idea is to think of $[A]$ as a modal operator indexed by the formula A. In other words, in this language, each formula comes with its own accessibility relation. Such an accessibility relation, when it comes to inconsistent stories, will eventually select non-normal worlds. So in general, without surprise, the truth conditions for $[A]B$ consists in checking whether B holds in all the worlds related to the world of utterance given A. Formally, we

[27] See [Nolan 2013] for a review and also [Berto and Jago 2019] for a more recent comprehensive presentation.

[28] See also an earlier presentation of the same material in https://cast.itunes.uni-muenchen.de/vod/clips/q9Q2wCjlNS/quicktime.mp4.

[29] If it was, then the subvaluation strategy would be available. But this is ruled out *ex hypothesis* here.

need to define the story based on A: $S_A^{w_0} = \{w | w_0 R_A w\}$. This is a generalisation of Lewis's idea that a story targets the set of worlds in which the story is told as known fact. Formally we have the following truth conditions:

$v_{w_0}([A]B) = 1$ iff $\forall w \in S_A^{w_0}$, $v_w(B) = 1$.

At this point, we are back to square one. We have a formal apparatus which can express that a formula B is "true according to some imaginative scenario A". This formal apparatus is unconstrained, for we allow any kind of non-normal world which can model any kind of inconsistencies. But this does not tell us *which* are the selected worlds. And if we target very anarchic worlds in which every formula gets treated like an atomic formula, then there is no *inference* properly speaking. We just record what is explicitly said in the fictional scenario, but we have no means of getting the implicit fictional propositions from the explicit ones. If possible-world semantics was too strong, impossible-world semantics (in its unrestricted version) is too weak.

3.4.2 Berto's thesis

Berto meets the worry by specifying what he calls the "mereology of imagination". That is some basic constraints any imaginative act (*a fortiori* any fictional story) should meet. Here is the most basic constraint he proposes:

> It seems to me that, when one imagines that a conjunction is the case, one also imagines each conjunct: you cannot imagine that Sherlock Holmes is a bachelor and lives in London without imagining that Sherlock Holmes is a bachelor.[30]

Formally, it gives the following inference schema:

- If $v_{w_0}([A](B \wedge C)) = 1$, then $v_{w_0}([A]B) = v_{w_0}([A]C) = 1$.

But Sylvan's box being fempty is precisely a counter-example to Berto's thesis. In order to accept the introduction of non-normal worlds, we had to accept that it is possible to imagine the box being both full *and* empty, without being full.

For his defence, Berto calls in some subtle distinctions from philosophy of mind. He says:

[30] [Berto 2017], p. 1286.

> Recall that we are not modelling the mere assumption or supposition of some content p, but more substantive (in Chalmers' jargon) positive conceivability – someone's bringing to one's mind a mental scenario: a state of affairs, or a configuration of objects and properties, which verifies p. One can assume or suppose that p in a proof, without perforce representing in the mind a state of affairs verifying p. One may then suppose (in this sense) a conjunction without supposing the conjuncts separately, or vice versa.
>
> But configurations of objects and properties, or states of affairs, seem to allow for constituent parts [...] Next, it seems to me, when one positively imagines a whole scenario or state of affairs, it appears that one automatically imagines its constitutive parts.[31]

Berto's justification either implies that fictional imagination is *not* positive conceivability. But he takes his crucial example from *The Adventures of Sherlock Holmes*, so I tend to believe that he would not say this. Or Berto would agree with the cool-head from the previous paragraph that such he would deny that die-hard inconsistent fictions actually exist. This leaves us with the same problem of defining a consequence relation for fiction in an unrestricted non-normal world framework.

3.4.3 A diagonal argument against the unrestricted non-normal world account

The argument

I think the discussion around "fempty" can be generalised into what may be called a "diagonal argument". The basic idea is the following: each time one restricts the non-normal worlds so as to have enough structure to define a consequence relation, one can devise a fiction in which the minimal logical structure is violated. As a result, there is no way of defining a consequence relation of *all* stories within the unrestricted non-normal framework.

Interestingly, this argument can be found in two places in the philosophical literature. First Routley:

> Given that the logic of a fictional world may be any logic, it follows that *there is no general uniform logic of fiction*. For the intersections of all logics is a null logic, no logic, as each purported logical principle is

[31] [Berto 2017], pp. 1287-8.

> cancelled out by a logic where it does not hold good. Consider, to illustrate, one of the more promising principles for a logic of fiction, formed by introducing a fictional functor O (Woods' *olim* operator) read, say, "it holds in fiction that", namely the principle $O(A \wedge B) \rightarrow OA$. Spelled out semantically the principle has it that if $A \wedge B$ holds in the world of an arbitrary work N then so does A. But consider now a novel where the principles of connexive logic govern, and where hence $A \wedge \neg A$ may hold though A does not. The world of such a novel repudiates $O(A \wedge B) \rightarrow A$. In claiming that there is no uniform logic of fiction, it is not implied that fiction has no logic, far less that it is illogical. In general, *each work will have its own internal logic*: it is simply that the emerging set of *common* logical principles will be zero. The semantical structure will reflect this situation.[32]

Proudfoot made the same argument quite independently. Her formulation of it is also quite nice, since she explicitly compares the possible-world attempt with the impossible-world attempt:

> Both the possible world semanticist and the impossible world semanticist are caught by the diversity of fiction. Just as there are more fictions than the possible, so the class of "impossible fictions" includes more cases than the impossible worlds semanticist can deal with. [...] As a striking and simple example, consider a logician's in which both A and its negation as defined by the familiar truth-tables are true (and true only). Even impossible worlds semanticists agree that there are no worlds in which A and its truth-table negation are true *simpliciter* (for by the truth-table if A is true then its negation is false).[33]

The unrestricted non-normal framework is thus inert. It is arguably possible to *express* all "possible" fictional truths in this framework, even the die-hard inconsistencies. But it is impossible to define a consequence relation to model fictional inferences in general. So the distinction between explicit and implicit fictional truths which was at the core of the problem is lost: there is no *solution* to the problem of fictional "truth".

[32] [Routley 1979], p. 10-11.
[33] [Proudfoot 2006], p. 31.

The end of a dialogue

I think this argument settles the matter, although it is not, strictly speaking, a knock-down argument. I think we simply reach an impasse.

In order explain what I mean, consider the following dialogue. On part is the functional theorist (F) who has been raising indirect arguments against the modal account all along. The other is the heroic saviour of the modal account (M) who has been led to defend the unrestricted non-normal account in the face of die-hard inconsistent stories.

F: All right, let's take non-normal worlds on board. The whole lot of them, even the wild ones. But the natural question is: *how* do we get the fictional "truths"? And the natural answer is: by reading the text.

M: True. But my job was not to give a theory of *interpretation*. I gave you what you asked for, namely a theory of *truth* in fiction. Interpretation, I take it, is a completely different problem.

The last move of the modal theorist thus consists in distinguishing between *truth* and *interpretation*, when it comes to fiction. I think this is a move for the hopeless.

On the contrary, we will see that the functional account of fictional "truth" articulates interpretation and "truth", explaining the latter by the former. As I suggested in the cool-head last defence, I think this basic tenet of functionalism can be combined with the modeling of fictional worlds as sets of possible worlds. This may suggest that, if functionalism is correct, one could use its notion of (fictional) interpretation so as to revitalise non-normal world semantics by providing the missing fiction-relative notion of consequence that it lacks.[34]

3.5 Conclusions

This section focused on two indirect arguments against the modal account: the argument from incompleteness and inconsistency. They consist in denying that the set of fictional worlds is a subset of the possible worlds of possible-world semantics. The observation that fictional worlds, if there are such things, are often incomplete and sometimes inconsistent shows that the possible-world framework is too strong for modelling fictional worlds. Hence, it is not adequate for modelling fictional "truth".

[34] I imagine that this is what Berto is trying to do, since he calls for our intuitions about fiction at crucial moments of his exposition. Perhaps one can see the functionalist account developed below as a way of fleshing out these important intuitions so that they can, eventually, be built into impossible world semantics.

There are two ways of responding to these indirect arguments. The first way (that of the hot-head) is to extend the set of possible worlds so that some worlds are incomplete and inconsistent. Consequently, the set of fictional worlds is indeed a subset of the worlds of the extended semantic apparatus which contains both possible (or normal) and impossible (or non-normal) worlds. The second way (that of the cool-head) accepts the claim the set of fictional worlds is not a subset of the set of possible worlds, but makes a counterproposal according to which the set of fictional worlds is a subset *of the power set* of the set of possible worlds. In other words, according to the cool-head, a fictional world is thought of as a set of possible world.

As can be seen, both ways involve a substantial revision of the central tenet of the modal account, according to which fictional worlds are possible worlds. But they are still *modal* accounts of fictional "truth", for they both claim that fictional sentences are one kind of modal sentences among others and so they give a theory of fictional "truth" in a general semantic framework for all modal sentences. In this sense, against the functional account, they treat fictional "truth" as a kind of truth.

I showed that the cool-head strategy was interesting to investigate for at least two philosophical reasons. First, it recognises that the two problems are actually two sides of the same coin and tries to provide dual solutions to dual problems. It is thus elegant. Second, the subvaluation strategy, which is one side of the coin, induces a very natural interpretation of what a fictional inconsistency is: it is a double-bind phenomenon. I gave an argument that shows that this idea was more robust than it first seems, based on a distinction between contradictions which are logically complex and contradictions which are logical simples. I explained why I am sceptic about the existence of the second kind of contradiction when it comes to fiction, for it seems to me that the alleged examples of fictions which require to imagine a logically simple contradiction do not keep their promises, the paradigmatic example being Priest's short story *Sylvan's box*. But I acknowledge that if contradictions can be logical simples and if these can appear in fiction, then the cool-head strategy should give way to the hot-head strategy.

Impossible possible worlds are strange entities which one might want to avoid if possible. If accepting non-normal world is like climbing a mountain up to the summit, then I think one can see the cool-head strategy as a base camp one should stay in for a few days in order to get used to the lack of oxygen up there. What I tried to show is that this base camp was higher than one might have guessed from the normal ground.

The last part of this section was thus concerned with the non-normal world account of fictional "truth", in order to account for die-hard inconsistent stories, if there are such things. The paradigmatic case being *Sylvan's box* rewritten so as to

involve a *fempty* box. I presented Berto's non-normal framework for two reasons. First, he presents the most unrestricted framework, which contains the worlds in which a box can be fempty, without being either full or empty. Second, he claims that this framework is adequate for the modelling of all imaginative acts, so *a fortiori* he should be able to model acts of fictional imagination.

I showed that this unrestricted non-normal framework meets a diagonal argument, when it comes to defining a notion of logical consequence. This argument was already present in the literature, in the independent works of Routley and Proudfoot. It says that any way of defining a general consequence relation for fictional imagination will meet *ad hoc* counter-examples in the form of a fiction in which the definition fails. *Sylvan's box* interpreted as involving a fempty box is, for instance, a counterexample to the rule of adjunction, which Berto takes as a good candidate for defining a minimal consequence relation for all imaginative acts. Consequently, the argument shows that non-normal-world semantics is bound to fail to model fictional *inferences* if it can model *all* fictional "truths".

I suggested that, quite surprisingly, the functional account could come to the rescue here by providing systematic reasons to put constraints of the acts of fictional imagination *via* a proper analysis of what counts as an *inference* in a fictional context. This bridge between functionalism about fictional "truth" and impossible-world semantics is rather tentative and cannot be thoroughly discussed until I presented the functional account in detail. However, as will be seen when I discuss the functional account, the formalism is far less mature than that of non-normal world semantics. So the bridging may be technically difficult if one ever wants to do it. Time will tell.

I think the take-home message of this chapter is that the two indirect arguments are interesting in that they force the possible-world framework to adapt and develop general formal tools which are used for many other purposes than modeling fictional "truths". Such adaptation, I tried to show, are also interesting conceptually. I think indirect arguments cast serious doubts on the modal account but they are not knock-down arguments. As such, they should pique the wise philosopher of fiction's curiosity into considering rival accounts. Direct arguments against the modal account are following and they will, hopefully, end up convincing even the obtuse modally-minded philosopher of fiction.

Chapter 4

Direct arguments against the modal account

> It is helpful to bear in mind that, while the *formal* notion of world is crystal clear in modal logic, it is of absolutely no interest for fiction.
>
> [Woods 2018], p. 86.

4.1 Plan of action

Direct arguments against the modal account, as opposed to indirect arguments, are arguments which deny that the initial drive of the modal account is true. It consists in denying that fictional statements should be construed as *modal* statements. Hence, if the arguments are successful, the modal account of fictional "truth" should not have any *raison d'être* anymore.

In this chapter, I will give two such arguments. The first argument is metaphysical in spirit. It consists in denying that fiction is just like the other modalities by turning the modal intuition upside-down. In fact, so the argument goes, fiction is the mother of all modalities so modal statements should be analysed after fictional statements and not the other way around. In order to do argue in this manner, I will have to show that the notion of *world* can be constructed out of the notion of *fiction*, the notion of world thus construed can later be used to give an analysis of modalities.

The second argument is semantic in spirit and comes from Evans's reply to Lewis. It consists in showing that Lewis's counterfactual analysis of fictional statements is flawed because the antecedents of these counterfactuals are not only contrary to facts

but they are conceptually impossible. So Lewis's analysis misfires. The argument rests on a plausible causal view of semantic information. To put it simply, the argument shows that one cannot hold *both* Lewis's view on fiction *and* a causal theory of semantic information. Since the causal theory of semantic information is independently plausible, this puts Lewis's counterfactual analysis of fictional statements under a lot of pressure.[1] This argument, I will show, also suggest a fiction-first perspective on counterfactuals and modal statements.

4.1.1 The modal intuition

Let us go, then, back the modal theorist's initial drive. It is the observation that fictional statements and counterfactual statements *are* similar in some relevant sense. The idea is that both kinds of statements talk about non-actual events and both require that the listener determine what (plausibly) follows from an explicit imaginative scenario. Failing to meet these requirements, in both cases, is tantamount to failing to *understand* these statements. This similarity is, I take it, undeniable.

The modal account consists in taking a modality-first perspective on this similarity. A functionalist can typically resist this step and provide a fiction-first perspective on this similarity. Historically, the logical analysis of modality was first. Conceptually, though, there is no reason to follow history. Restoring the fiction-first perspective is the task of the functionalist. This is what this chapter is all about.

4.1.2 Two suppositions of the modal account

The modal account supposes two interconnected theses. The first is that fictional statements are a kind of modal statements. There are different kinds of modalities: alethic, temporal, deontic, epistemic, doxastic, etc. Alethic statements are statements of the form "It is necessary / possible / impossible that p"; temporal statements are of the form "It was / is / will be the case that p"; deontic statements are of the form "It is permissible / mandatory / it ought to be the case that p"; epistemic statements are of the form "It is known / ignored (by the subject A) that p"; doxastic statements are of the form "It is believed / doubted (by subject A) that p"; etc.

Each kind of modality has a proper object of study which comes with a certain logical structure which is enclosed in a modal operator. For instance, epistemic logic systematically studies the natural phenomenon we usually call *knowledge* while

[1] The work of Ruth Millikan presents several independent arguments in favour of a causal theory of semantic information. See in particular [Millikan 1998].

doxastic logic is concerned with *belief*. Evidently, knowledge and belief are governed by different logical rules, for it is clear a false belief is a belief while false knowledge is not knowledge. The study of modal statements is a chapter of a general theory of truth because in natural language modal statements are expressible and are truth-evaluable. Indeed, statements of the form "A knows that p" or "A believes that p" are common and can be either true or false, depending on how the world is (from A's perspective). The modal theorist supposes that the fictional operator "In the fiction F, p" is to be added to the list of modalities.

The second supposition is linked to the counterfactual account of fictional statement. It says that possible worlds are conceptually more primitive than fictional worlds. Counterfactual statements are complex modal statements. There are different kinds of counterfactuals depending on the kind of modality they call in. For instance, the distinction between indicative and subjunctive conditionals is usually interpreted as involving different kinds of modalities: indicative conditionals breed on epistemic modality whereas subjective conditionals rather call for alethic modality. Counterfactuals are a complex linguistic phenomenon, though. Indeed, it is not easy to reduce counterfactuals to any other modal logics so that some even talk of a counterfactual modality in and of itself.[2] What is uncontroversial is that counterfactual statements are modal statements. The most popular semantic analysis of counterfactuals is, by far, done in the possible world semantics framework. As a result, the counterfactual analysis of fictional statement implies that the intuitive notion of "fictional worlds" should be cashed out in possible world semantics.

In theory, the two suppositions are independent, for it is not necessary that all modalities should be analysed in the possible world semantic framework. It could be that fiction is a modality, but not the sort of modality that requires the notion of possible world. In practice, though, possible world semantics (and its extensions) is so dominant a view that it usually functions as a benchmark for providing a semantic analysis to all kinds of modal statements. Keeping things separate is not easy when it comes to detail. However, I will try to keep the two suppositions as separate as I can in this section.

4.1.3 The plot

Direct arguments are arguments which lead to the denial of these suppositions. Rarely are they to be found in the philosophical literature, and they are never ex-

[2]See for instance Starr, William, "Counterfactuals", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2019/entries/counterfactuals/.

panded into full length argument against the modal account.

I will start with an argument against the first tenet. The idea, again, is to switch perspective and offer a fiction-first perspective on modalities in general. In particular, since fiction intuitively comes with the notion of a fictional *world*, I will argue that one can resist the temptation of reducing fictional worlds to possible worlds by defining a notion of fictional worlds which is independent of that of a possible worlds. This will naturally lead me to the reversal of the second tenet, namely that the set of possible worlds is actually a subset of the set of fictional worlds, as previously defined. This completes the first direct argument.

The second direct argument will be targeting the second supposition directly. This argument comes from an idea of Evans in [Evans 1982]. Evans, unfortunately, did not have time to develop his idea into a full-fledged argument. I think such an argument can be given. The detail of this argument, I think, is one of the main motivations of Walton's functional account so it will provide a natural transition for the second half of this part (chapters 5, 6 and 7) in which I develop the functional account.

I do not claim these to exhaust all possible direct arguments against the modal account.

4.2 Fiction is not just like the other modalities

4.2.1 Structure of the argument

I will give the general structure of the first argument here. It contains two premises which will be argued for independently in the following sections.

Possible world semantics was originally designed to account for alethic logic, that is the logic of necessity and possibility. Kripke formalised an idea which traces back to the work of Leibniz.[3] But it quickly was used to formalise other kinds of modality, notably through the works of Hintikka, Montague and Kanger.[4] Let us call these modalities "P-modalities".[5] If there ever was a general framework in which all modalities could be expressed and theorised, possible world semantics is the best candidate one could think of.

When it comes to fictional sentences, as we have seen in the previous section, one can think that the possible-world framework is too strong. Interestingly, one can argue that there are other modal sentences which behave mush like fictional

[3]See in particular [Kripke 1959] and [Kripke 1963].
[4]See for instance [Hintikka 1962b], [Montague 1970] and [Kanger 1957].
[5]"P" for possible.

sentences in that the possible-world framework is also too strong for them. Let us call them "non-P-modalities". I will show that there are non-P-modalities in the following section.

From this, one can conclude that possible-world semantics is not a good framework for theorizing about *all* modalities.[6] This goes against the first tenet of the modal account of fictional "truth".

At this point, there are two ways to go. Either one tries to devise a mechanism which transforms non-P-modalities into P-modalities, or one tries to come up with a notion of "world" which can naturally subdivide into P-worlds and non-P-worlds. If one is endowed with such a notion of "world", one can have a fiction-first perspective on modalities. Indeed, this notion of "world" unifies all the modalities and grounds the fundamental distinction between P- and non-P-modalities. In order to run this argument, I will provide such a notion of "world" which stems from an analysis of what a *fictional* world is.

Given the two premises which remain to be argued for, one can conclude to a complete reversal of the modal account. Indeed, the argument concludes that the set of possible worlds (the P-worlds) is a subset of the fictional worlds (the non-P-worlds).

4.2.2 Non-P-modalities

I take it for granted that alethic modality is a P-modality.[7] Therefore, some modalities are P-modalities. I need to show that some modalities are non-P-modalities. Deontic modality, I shall argue, is non-P-modal.

Deontic modals are the formalisation of the dual notions expressed in natural languages by *obligations* and *permissions*. Deontic logic thus aims at formalizing what ought to be the case. The modal vocabulary of deontic logic is concerned with a formalisation of natural language sentences of the form: "it is obligatory that ϕ" (or "it ought to be the case that ϕ") and "it is permitted that ϕ" (or "it is permissible that ϕ"); as well as any other idioms expressing these ideas (e.g. "It is

[6]This is where the proponents of non-normal worlds come in. Note that it is arguably more interesting for them to rely on non-P-modalities than on fiction. Indeed, if fiction turns out not to be a modality after all, then they would not be able to claim that non-normal world semantics provides *the* general framework for theorizing modalities.

[7]Some have doubted this. For instance [Peacocke 1978] argues that the traditional possible-world semantics is not best suited for the semantic analysis of the necessity operator. See also the more recent [Fine 2012]. As for epistemic modality, although the possible world analysis of epistemic modals is usually taken for granted, the hard problem of logical omniscience casts doubts on the idea that epistemic modality is a P-modality.

forbidden / prohibited that", "It is required / one's duty to", "One must / can do that", etc.). The well-formed formulas of deontic logic are thus all the sentences of predicate logic, plus simple deontic formulae, plus the complex sentences resulting from any well-formed combination of the logical connectives and the deontic symbol. For instance, here is a complex well-formed formula of deontic logic: "If it is not permissible that ϕ, then it ought to be the case that $\neg\phi$".

There are real-world examples of sets of obligations, here are two: the Ten Commandments of the Bible and the Highway Code. According to these codes, one can produce deontic sentences which can have truth conditions. The truth-value of a sentence will, of course, depend on the interpretation of the codes, which can be a delicate matter (especially for the Bible). So there will be borderline cases, but on the other hand there are clear-cut cases of true and false deontic sentences in the context of these codes. For example the sentence: "It is permissible that a man have sex with any woman they want" is false according the Ten Commandments because it is explicitly true according the Ten Commandments that "You shall not covet your neighbour's wife". For example the sentence: "If the light is red, then it is not permissible to pass crossroads" is true according to the Highway Code because it corresponds to the basic rules about traffic lights set up in the Highway Code.

But such codes do not define anything like a P-world. Indeed, codes need not be complete and can be inconsistent.

There are many deontic sentences which are neither true nor false according to the Ten Commandments and according to the Highway Code. Take the following sentence: "It ought to be the case that one scratches one's head whenever one feels like it". It seems clear that the Ten Commandments and the Highway Code are silent about this sentence, so to speak.

It is also possible that a set of obligations be inconsistent. I do not have steady intuitions about the Bible and the Highway code on this though. But here is an example taken from [Priest 2008a], §4.8.3:[8]

> Suppose that an (absent-minded) state legislator passes the following traffic laws. At an unmarked junction, the priority regulations are:
>
> (i) Any woman has priority over any man.
>
> (ii) Any older person has priority over any younger person.
>
> (We may suppose that clause 2 was meant to resolve the case where two men or two women arrive together, but the legislator forgot to make it

[8] In the context Priest is discussing the "paradox of strict implication" and in particular the explosion principle, so his dialectic is not exactly the same as here. But the general idea is equivalent: the need for non-P-world is pressing in conditional logic too.

> subordinate to clause 1.) The legislation will work perfectly happily in three out of four combinations of sex and age. But suppose that Ms X, of age 30, approaches the junction at the same time as Mr Y, of age 40. Ms X has priority (by 1), but has not got priority (by 2 and the meaning of "priority"). Hence, the situation is inconsistent.

This is an artificial example but such double-bind phenomena happen all the time in deontic systems: subsystems of law contradict each other, promises easily create double-binds, ethical systems are usually dilemma-ridden, many subtle bugs in computer programs come from contradictory instructions, etc. So it is possible for a deontic system to be inconsistent.[9]

Hence, deontic modality is non-P-modal. Any Kripke-style semantic for deontic sentences will thus run into the same difficulties as those we consider for the modal account of fictional "truth".[10] The same kinds of response will be available: either construct a complex notion of deontic (or "ideal") worlds out of P-worlds, or extend possible-world semantics to impossible world semantics.

This completes the first missing part of the argument given above.

4.2.3 On what makes a *fictional* world a fictional *world*

There is a deep-rooted intuition that stories tell us about worlds, or universes, that are not actual or real. This intuition can be fleshed out when we consider fiction so that we have a general definition of a *world*. Once we have such a general definition of a world, we can then define what should count as a P-world. In the following, I propose three possible definitions of a *world* which, suitably applying to fictional practices, are good candidates for the required notion.

A short disclaimer first. That fictional practices can ground a definition of a world may recall Nelson Goodman's notion of a "worldmaking activity", presented in detail in his [Goodman 1978]. Goodman's main argument in this book is that the *origin* of all worldmaking activities is to be found in the structural features of the *symbol systems* used when performing these activities. If fictional practices are worldmaking in Goodman's sense, then his argument should apply to fiction. I will not engage with Goodman's argument though. For one thing, I do not need to speculate about the origin of the creative powers from which worlds spring, so to speak, in order to define

[9] Of course, one can argue that such systems will be *inefficient*. First, this remains to be proved. Second, this is besides the point I am making about deontic modality being non-P-modal.

[10] And some other which are peculiar to deontics. See for instance McNamara, Paul, "Deontic Logic", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2019/entries/logic-deontic/.

rigorously what a fictional world is. For another, I do not have much to say about symbol systems, nor do I think that the notion can be applied without difficulty to fictional practices. I guess the best way of putting the somewhat narrower question under discussion here is: What is it that makes a *fictional* world a fictional *world*?

Fictional worlds are spatio-temporal systems

> Had we but world enough, and time,
> This coyness, Lady, were no crime
> We would sit down and think which way
> To walk and pass our long love's day.
>
> Opening of Andrew Marvell 1681 "To His Coy Mistress"

There is this idea that the notion of a *world* does not designate a *thing*, but the general background of all things and events. In this sense, the real world is the condition of possibility of all the real things and events.[11]

Indeed there are natural phenomena or events, those we experience or those predicted by natural sciences for instance. So natural events are possible. But the condition of possibility of such events is that there is a world *in which* these events occur. In particular, because certain events are predictable by the natural sciences, there must be an underlying structure which makes scientific predictions possible. So each time we recognise an *event*, it comes with a background which we call a "world".

In particular, this background is spatio-temporal. That is, events occur in space and time. This is tantamount to saying that there can be no event outside space and time. To simplify, I will identify a world with its corresponding spatio-temporal framework, setting aside the thorny issue of deciding what are the other non-spatial, non-temporal dimensions of worlds in this sense.[12]

This definition of a world as a spatio-temporal system can be generalised into talking of fictional worlds.[13] Indeed, we easily recognise *fictional* events such as Hamlet deliberating for quite a long time on the possibility of ending his life instead

[11]Note the Kantian undertone of this idea. I do not claim that the notion I describe here is the same as Kant's technical definition of a *world* in the Transcendental deduction, though.

[12]Note that, in old English, the shade of meaning for "world" as synonymous with "space" was available, as shown by Marvell's poem in epigraph of this subsection.

[13]This idea can be found also in the work of Enrico Terrone. I had several personal discussions with him on the subject. Terrone found this definition when reading the work of Strawson. It is possible that Strawson gets this idea from his reading of Kant. Again, I do not claim that Strawson's reading of Kant is accurate.

of avenging his father. By definition, such fictional events do not occur in the real world. Hence, we form the idea of a different world *in which* the fictional events take place. Fictional worlds are nothing more than the imagined background which makes it possible one imagines the fictional events.

Once we have this definition of a world, one can define what a P-world is. Given a spatio-temporal frame, one can define a notion of spatio-temporal distance. A P-world is a world containing all the things and the events at a distance from each other.[14] This definition of world also allows for non-P-world: isolated events ground the idea of an incomplete world. Impossible events require that we form the idea of an impossible world. Imagining the background of impossible events may not be phenomenologically easy but it is arguably conceptually possible. Indeed, it seems that we do so when, for instance, we read stories involving inconsistent time travel.

Fictional worlds are Stoic worlds

Fictional events seem to be *inescapable*. Indeed, Othello will kill Desdemona and there is no way around it.[15] No spectator can prevent the murder, no character will. There is a clear sense in which the fate of fictional characters is settled for good. If Othello were not to kill Desdemona, it would not be the same fiction.

On the other hand, it seems implausible that all fictional events are all equally inescapable. For instance, suppose in one version Othello blinks before killing Desdemona, and in another version he does not. Clearly, this does not support the intuition that there are two different fictions. So Othello's blinking before killing Desdemona does not seem to be part of his inescapable fate.

Therefore, the inescapability of fiction should rather be thought of at the level of the plot of the story. Changing the plot *is* changing the story. In this sense, the plot is the inescapable fate of the fictional characters.

The stoics had this idea that fate was the world's pneuma, the deep-down structure of the cosmos. Similarly, the character's fate can ground the idea of a world as a definite, inescapable, determinate sequence of events. The plot of a story is the backbone of the fictional world thus conceived. Interestingly, if one takes this general definition of a world, one can easily explain why it is so interesting and fruitful to investigate the systematic relations between narrative structures of a story and the causal structures of the associated fictional worlds.

Stoic worlds defined as determinate sequences of event can yield P-worlds. From the idea of fate, one should define a notion of causality. For instance, take that of the

[14]This is, in spirit, Lewis's famous definition of a world in the opening of [Lewis 1986].
[15]This example and the inescapability of fiction is discussed in [Walton 1978].

physical sciences. In this case, events are causally linked into a particular sequence, the sequence thus defined being inescapable. A P-world is the closure of the sequence of events under the relevant causal relation. It thus includes all the causes and all the effects of the sequence of events in question. Stoic worlds can also be non-P-worlds, since there is no constraint on the connection of the events into a sequence in the above definition. It will naturally be incomplete, and sometimes inconsistent given the plot.

Fictional worlds are clusters of fictional truths

Here is a definition of fictional worlds taken from [Walton 1978]:

> "Fictional worlds" can be understood as collections of fictional truths. The fact that "Robinson Crusoe exists in a given fictional world" amounts simply to the fact that this is fictional that Crusoe exists, and that this fictional truth belongs to a collection of fictional truths. The fictional world is not construed as a realm in which it is *true* that Crusoe exists.[16]

So given a definition of fictional "truth", one can define what a fictional world is by clustering them together. One can appreciate how Walton's proposal is a complete reversal of the modal account. First we get a "cluster of fictional truths", and then we call that a world. It is *because* we have a bunch of fictional propositions about Robinson Crusoe, that we say he lives in a fictional world. Providing such a theory of fictional "truth" is the task of the functionalist.[17]

One way to flesh out Walton's idea is the following. We have propositions, extracted from the fictional text. Fictional reading consists in "realizing" or "satisfying" these well-formed formulae by constructing a set-theoretical model. What we usually call "fictional imagination" is precisely this way of "realizing" or "satisfying" a set of propositions. Walton calls fictional *world* this set of interpreted propositions in the imagination.

A fictional world, understood in a full-fledged phenomenological sense, is usually imagined to quench our irresistible desire to reify propositions. But it is not necessary, really. In principle, we could compute fictional propositions without imagining a fictional world in any substantial phenomenological sense.[18] It is just the way we

[16] pp. 15-6.

[17] The technical detail are presented in chapter 6 and 7.

[18] Arguably, some abstract fictions force us to compute fictional propositions while preventing us from imagining a "world" in any phenomenological sense. See for instance the work of Norman McLaren, like Blinkity Blank, Rythmetic, or Canon. Perhaps non-narrative poems are akin to abstract fictions in this sense. Thanks to Marion Renauld for helping me thinking along these lines.

usually understand proposition which takes over.

Given the definition of a world as a cluster of fictional "truths", one can define what a P-world is. A P-world is a set of propositions which is closed under a classical entailment relation. This can be regimented, for instance, by forcing the set of fictional "truths" to be expressed in predicate logic endowed with a relevant semantics for the notion of entailment. Set-theoretic semantics is one way to go. But a cluster of fictional "truths" need not be closed under classical entailment. So that they can be just a set of propositions, thus defining a non-P-world.

This completes the second missing part in the first direct argument given above. A fiction-first perspective is now available and, if my argument above goes through, desirable.

4.3 Evans against Lewis

I will now turn to the second direct argument against the modal account. It consists in attacking Lewis's counterfactual analysis of fictional statements. In order to do so, I propose to follow Evans who argued that Lewis's paraphrases systematically yields counterfactuals with impossible antecedents.

4.3.1 Evans's dazzling paragraph

At the end of [Evans 1982], Evans encounters the problem of fictional reference. He adopts a functionalist view of fictional "truth". While presenting his version of functionalism, inspired by Walton, he formulates a devastating objection to Lewis's counterfactual analysis of fictional statements. Unfortunately, his objection is not developed and remained very sketchy, to say the least. I do not know of anyone who did the work of explicating this objection.[19] I will try to do this work in this section. Before I quote Evans, some contextual elements are needed for the understanding of the paragraph.

First, Evans talks of the "mud-pie game". It is a game of make-believe in which some children are supposed play with mud pats pretending these are pies. They play at cooking them, eating them, and so on.

Second, when Evans writes "rule of fiction" he means the rules which are needed for making inferences from explicit fictional "truths" to implicit ones. To get back to

[19]Chapter 10 of [Evans 1982] is one of the few chapters without appendix by McDowell (with chapter 1 and 2, as well as chapter 4). I think this is really surprising, given the fact that the chapter is manifestly a first draft (as opposed to chapter 1 and 2 for instance).

our original example a "rule of fiction" would be that "all princes are human beings" in *Hamlet*, so that one can infer from the explicit "truth" that Hamlet is a prince to the implicit one expressed in (1). In a previous paragraph, Evans formulated what he takes to be general "rules of fiction" in the explicit form of a counterfactual. He said that one gets premises like "all princes are human beings" in *Hamlet* by applying a general counterfactual of the form "if such and such holds in reality and is not contradicted by any explicit fictional statement, then it ought to be 'true' in the fiction".

Evans thus comments on his using counterfactuals to define the "rules of fiction":

> By formulating the rules of [fiction] explicitly in terms of counterfactuals, I do not intend to suggest that the capacity to understand the counterfactual idiom is a more primitive capacity than the capacity to engage in [fiction]. In fact I think that they are the *same* capacity. (Using counterfactuals is engaging in a purely cognitive pretence; though it might sometimes be better to speak of cognitive acts within the scope of a supposition rather than pretence.) This is obviously not the place to attempt the analysis of this capacity. However, I should say that this is important for my purposes that counterfactuals such as the one I formulated at the end of the last paragraph can be *true*, and this probably means a rejection of the popular "possible-worlds" analysis of counterfactuals. [Footnote: See, e.g., [Lewis 1973].] For almost all of these counterfactuals have *impossible* antecedents; there simply are no possible worlds in which these mud pats are pies. Equally, and more relevantly to our proper concerns, there is no possible world in which the illusory information which we receive is not illusory. The initial supposition generating the [fiction] within which we make conniving use of empty singular terms is that certain information, which is in fact deceptive, is information from actual objects and events. And I would argue that there are no possible worlds in which just that information is veridical; [indeed,] information is individuated by causal origin, which makes it impossible to understand how the same bit of information might have had a different origin. [Footnote: Lewis half-perceives this as a problem for his possible-worlds analysis of truth in fiction: see [Lewis 1978], p. 40.][20]

This paragraph has two parts.

In the first part, Evans clearly accepts Lewis's original motivation for developing his counterfactual account of fictional statements: understanding counterfactual and

[20] [Evans 1982], pp. 355-6.

fictional statements require "the *same* capacity". But, contrary to Lewis, he takes a fiction-first perspective on this similarity, or I should rather say a "pretence-first" perspective. From what Evans writes, one can guess that what he has in mind is the following picture: pretence is one kind of mental activity which subdivides into fictional imagination and counterfactual thinking. Fictions are social, recreational pretences; while counterfactual thinking is a "cognitive" pretence, which might better be termed a "supposition". Since games of make-believe (i.e. fictions) are the paradigmatic cases out of which the study of pretence has been systematically developed after Walton, I think it is fair to say that Evans's "pretence-first" perspective on the observed similarity between fictional statements and counterfactual statements is indeed a fiction-first perspective (as opposed to a modal-first perspective).

In the second part (from "However" on), Evans points towards a problem for Lewis's counterfactual analysis of fictional statements as expressed in the possible world semantics framework. The problem is that the counterfactual statements which we need to have access to premises like "princes are human beings" in fictional inferences, have "*impossible* antecedents". He gives an example, then he gestures at a possible argument ("I would argue that") and finally he suggests that Lewis only (or already) "half-perceives" the problem.

In order to elucidate Evans's argument against Lewis, I will proceed in the opposite direction. I will first explain Lewis's "half-perception", then I will flesh out Evans's argument and finally comment on the mud-pie example.

4.3.2 Explanation of Evans's argument

Lewis's storytelling

For Lewis, getting the fictional "truth" of a fictional statement implies that we first recognise that the statement is fictional, i.e. the result of an act of storytelling. Lewis then says:

> Storytelling is pretence. The storyteller purports to be telling the truth about matters whereof he has knowledge. He purports to be talking about characters who are known to him, and whom he refers to, typically, by means of their ordinary proper names. But if his story is fiction, he is not really doing these things. Usually his pretence has not the slightest tendency to deceive anyone, nor has he the slightest intent to deceive.[21]

This way of describing what storytelling is motivates Lewis's Analysis 0 which amounts to paraphrasing (3) and (4) into (5) and (6). The fictional operator "In

[21] [Lewis 1978], p. 40.

fiction F, ..." is thus developed into "Were F told as *known fact* rather than fiction". This paraphrase was explained in §2.1.2.

Consequently, the counterfactual analysis of fictional statements requires that one considers the closest possible worlds in which the act of storytelling that happens in our world is a truthful testimony. So the two speech acts (the real act of storytelling and the possible act of truthful testimony) differ *only* in their intention: one is pretended, the other is serious. In other words, the two speech acts must have *the same* semantic content, i.e. contain the same propositions, namely those we recognise as the fictional "truths" in the real world.

It is important that the second speech act (the possible act of truthful testimony) has a *serious* intention. Indeed, there can be another world in which the second speech act is also an act of storytelling (i.e. pretence), albeit distinct from the original one. Such a case is illustrated by Pierre Ménard in Borges's eponymous short story. Ménard is writing fiction in the early 20th century. By a very rare combination of factors, Ménard's fiction happens to match word for word Cervantes's *Don Quixote*. It is explicitly fictional that Ménard *independently* wrote a story which is identical to Cervantes's original story. However, the aesthetic value of the two works, so Borges argues, are completely different. Ménard is writing in a language and style which is very remote from the spoken Spanish of the early 20th century, while Cervantes uses his own everyday language. So Ménard, Borges concludes, produces a work whose originality is far greater than that of Cervantes, albeit the *same* work.

Crucially, Ménard intends to produce a *fiction*, i.e. he is pretending. So when reading Cervantes's *Don Quixote*, one should not pick up the possible world in which *Ménard* tells the story. Otherwise we should then go to the closest possible world in which Ménard's story is told as known fact, which is, by definition, a different possible world. Ménard does not inhabit the same world as Don Quixote, Sancho Panza and Rocinante. Instead, one should go directly to the world in which Cervantes's story is told as known fact.

Evans proceeds and denies that there is such a possible world in which a story is told as known fact *and* has the same semantic content. In other words, the second speech act *cannot* have the same semantic content as its corresponding act of storytelling.

Evans's argument

In order to show that the two speech acts (the real pretence and the possible truthful testimony) cannot have the same semantic content, let us, first, focus on names.

Names contribute to the semantic content of the speech act in which they occur

by providing a referent, or failing to do so. In other words, the semantic content of a name is either a referent, or not a referent. By definition, a fictional name has no referent. So a fictional name does not contribute to the semantic content of the speech act in which they occur by providing a referent.

But the name, say "Hamlet", is used both in Shakespeare's speech act and in the possible truthful testimony associated with Shakespeare's pretence. According to Lewis, it must be that the fictional name in the actual world should become a real proper name in the possible world in which the story is told as known fact. This entails that, in the targeted possible world, "Hamlet" has a referent. Hence, by definition, "Hamlet" contributes to the semantic content of the second speech act (the possible truthful testimony) by providing a referent.

Consequently, the first speech act and the second speech act do not have the same semantic content. Indeed, they differ in semantic content at least as far as the name "Hamlet" is concerned. So there are no possible worlds in which *Hamlet* can be told as known fact. In other words, the possible truthful testimony associated with *Hamlet* does not have the same semantic content as Shakespeare's actual story.

Pretence is powerful: it secures the fact that a fictional name is necessarily fictional. Since a fictional name is *necessarily* fictional, there are no possible worlds in which it is not fictional. This is, in essence, Evans's argument against Lewis.

Note that Lewis does not agree with direct reference about names, he rather holds the view that names have descriptive content. So Lewis's general picture is not affected by Evans's argument. Evans's argument should rather be thought of as an incompatibility result: it is simply impossible to combine direct reference with the counterfactual analysis of fictional statements.

Kaplan and Kripke's arguments: The same argument was made independently by Kaplan and Kripke, though it was not directed at Lewis's analysis of fictional statements.

In appendix XI of [Kaplan 1973], Kaplan warns us about a surreptitious change of language. Focusing on the worlds in which the myth of Pegasus is "truthfully told" (the "M worlds"), Kaplan remarks that the name has changed its meaning:

> But beware the confusion of our language with theirs! If w is an M world, then *their* name "Pegasus" will denote something with respect to w, and *our* description "the x such that x is called 'Pegasus'" will denote the same thing with respect to w, but *our* name "Pegasus" will still denote nothing with respect to w. Also, in different M worlds, different possible individuals may have been dubbed "Pegasus"; to put it another

> way, *our* description "the x such that x is called 'Pegasus'" may denote different possible individuals with respect to different M worlds.[22]

In [Kripke 1972], Kripke makes explicit that a metaphysical consequence of his causal theory of reference is the following:

> I hold the metaphysical view that, granted that there is no Sherlock Holmes, one cannot say of any possible person that he *would have been* Sherlock Holmes, had he existed.[23]

Lewis's "half-perception"

Lewis indeed "half-perceives the problem", namely that is there is no possible world at which the story is told as known fact *and* the two speech acts have the same content. He explicitly discuss the problem of names like "Sherlock Holmes" changing their semantic contribution. As he puts it:

> At those worlds where the same story is told as known fact rather then fiction, those names really are what they here purport to be: ordinary proper names of existing characters known to the storyteller.[24]

His way of responding to the worry is to suggest that the name "Sherlock Holmes" as used in our world is not an ordinary proper name. On the contrary he suggests the following picture:

> [...] the sense of "Sherlock Holmes" as we use it is such that, for any world w where the Holmes stories are told as known fact rather than fiction, the name denotes at w whichever inhabitant of w it is who there plays the role of Holmes. Part of that role, of course, is to bear the ordinary proper name "Sherlock Holmes". But that only goes to show that "Sherlock Holmes" is used at w as an ordinary proper name, not that it is so used here.[25]

As we have just seen, Kaplan, Kripke and Evans precisely deny that this can happen. For there is no world at which the name "Sherlock Holmes" is the same *and* picks up an individual. As Kaplan puts it, the only way one can do this is to change one's language.

[22] [Kaplan 1973], p. 507.
[23] [Kripke 1972], p. 158.
[24] [Lewis 1978], p. 40.
[25] [Lewis 1978], p. 41.

So far Lewis is alright, for he does not hold that fictional names are rigid designators. But in footnote 9 he considers the objection that one can use "Sherlock Holmes" without knowing the descriptive content of "Sherlock Holmes":

> Many of us have never read the stories, could not produce the descriptions that largely govern the non-rigid sense of "Sherlock Holmes", yet use this name in just the same sense as the most expert Baker Street Irregular. There is no problem here. Kripke's causal picture of the contagion of meaning, in "Naming and Necessity", will do as well for non-rigid senses, as for rigid ones.[26]

So Lewis fails to see the incompatibility between the causal theory of names and his counterfactual analysis of fictional statements. Lewis's failing to perceive this is difficult to understand.[27]

Generalisation of Evans's argument

What is true about fictional names actually generalises to all kinds of semantic information, as Evans "would have argued". He holds that information (and misinformation) in general (hence semantic information) is individuated by causal origin. That is, information enters the cognitive system from the outside world through perception and testimony seen as causal mechanisms.

Hence, only the information originating in the outside world and causally related to the agent can be interpreted by the cognitive agent. In the case of fiction, pretence makes it the case that the information received is not veridical. Pretence is a case of deception, albeit a case where no harm is done. As a result, what is represented in fiction does not hold in reality. Indeed Hamlet is not a human being in reality, although the information contained in the fictional proposition expressed by the fictional sentence "Hamlet is a human being" is definitely part of the fictional content of *Hamlet*. The real-world origin of this fictional information is clearly identified: it is Shakespeare's play *Hamlet*, this sophisticated real-world object with which we can causally relate by reading it (or watching it on stage).

Were the complex fictional information contained in *Hamlet* veridical, everything is changed. The origin of the information in this case would clearly not be the play

[26] [Lewis 1978], p. 41.

[27] Kripke's above quote comes from the addenda to the second edition of *Naming and Necessity*, published in 1980, though. And Lewis was working with the first edition of 1972. So either Lewis *really* misunderstood Kripke's lectures, or Kripke was definitely cryptic during his lectures. As for ignoring Kaplan's version of the argument published in 1973, I guess Lewis has no excuse...

(the real-world object, written by Shakespeare), but the events that it truthfully tells the story of. However, if such information could enter one's cognitive system, then it must be that one is causally related to it. There is no world in which we, ourselves, are causally connected to Hamlet. We, ourselves, could only possibly be causally related to individuals in the actual world (and, unfortunately, only very few of them).

So there is no sense in which the fictional information we can gather by reading fictions and going to the theater could be gathered otherwise, in a causal manner. The same incompatibility result generalises: one cannot hold both a causal theory of semantic information and Lewis's counterfactual analysis of fictional statements.[28]

Application on the mud-pie game

The mud-pie game example is structurally identical to *Hamlet*, but the causal relations are much simpler to describe. Some children play with mud, forming globs of mud, and pretending that *these* mud globes are cakes.

Manipulating the globs of mud in such and such a way in the real world would correspond to such and such "truths" about the corresponding cakes in the pretence. For instance, adding some gravel on top of one glob of mud is likely to suggest that "this is a cake with chocolate chips on top of it" is "true" in the pretence, and so on.

But these globs of mud they manipulate and pretend about could not possibly be cakes. If they were cakes, then they would not be globs of mud. The information the children gather about these globs of mud originates in the causal relations they have with them, especially through manipulation. It is just impossible that they could gather the same kind of information, were these globs of mud to be real cakes.

In other words, there is no possible world in which these globs of mud *are* cakes. Consequently, any counterfactual statement whose antecedent is "if these globs of mud *were* cakes" is bound to run into trouble.[29]

4.3.3 Conclusions drawn from Evans

> People are, in short and among other things, gatherers, transmitters and storers of information.
>
> [Evans 1982], §5.2, p. 122.

[28] Same *caveat* as before: Lewis does not hold a causal theory of semantic information, so the argument does not affect Lewis's position as a whole.

[29] Unless one does not hold that demonstrative directly refer, but instead descriptively refer.

> We human beings are natural objects, denizens of the causal order. [...] We are seekers of knowledge and seekers of information.
>
> [Woods 2018], §3.3, p. 59.

Evans's argument is the following: there are no possible worlds in which a story told in pretence in the actual world can become a truthful testimony without changing its meaning, i.e. crucially differing in semantic content. Its explicit premise is a causal theory of information in general. In particular, a causal theory of names is enough to produce this incompatibility result.

Different causal theories of reference have been developed from the seventies on at the instigation of [Kripke 1972]. Donnellan is an early contributor of such theories, Putnam extended its results, Kaplan and Evans also gave their version of it. Evans's argument is thus not original in this respect, and the same arguments (although not directly addressed to Lewis's account of fictional "truth") can be found in the work of Kripke and Kaplan as I showed above. Evans's original contribution lies in its direct reply to Lewis and to its generalisation from reference to information in general. To put it simply, what this argument shows is the following incompatibility result: one cannot hold *both* Lewis's view on fiction *and* a causal theory of semantic information.

One can see that this is not a knock-down argument against Lewis's position. Indeed, Lewis explicitly denies the causal theory of semantic information.[30] Consequently, Evans's argument is compelling only insofar as one has compelling reasons to hold a causal theory of semantic information.

Functionalism can thus use the causal theory of reference as ammunition in the debate against Lewis. Evans, I think, was doing exactly this at the end of his unfinished book. I think this also deeply influenced more recent functionalists about fiction like Everett, whose theory I will discuss in great detail in chapter 6 and 7.

Fictional names, on the other hand, have become an open problem for causal theorists of reference. Kripke devotes his second series of lectures to this problem, later published as [Kripke 1973/2013]. Providing an adequate theory of fictional names within causal theory of names is still an open area of research. Functionalism, I will argue, can help solve these problems by providing a general theory of fiction which explains our use of fictional names. Part 3 of this dissertation is devoted to this question.

Consequently, the union of functionalism and the causal theory of names is not

[30] For instance in [Lewis 1984].

merely an opportune alliance, it is clearly a virtuous circle. Evans, I think, saw this clearly although he did not have time to expatiate on it.

As for Walton himself, it is really not clear what he thinks about this. He liked and used Evans's idea as the last chapter of [Walton 1990] demonstrates. He says that Evans's ideas are "much like [his] in spirit and in some of its detail." But at the same time, he never used anything like Evans's argument while arguing against Lewis. Moreover, if there are some clues that he would go for a causal theory of semantic information, Walton does not pronounce on it. On this, Walton is difficult to interpret.[31]

[31]See [Schuppert 2019], §5.3, for textual evidence and a detailed discussion on how to interpret Walton's texts on this point.

Chapter 5

Walton's programmatic ideas

After arguing against the modal account, I hereby open a constructive phase in this chapter. My presentation and defence of the functional account will expand in three subsequent chapters. This chapter will trace back the functional account to Walton's original work, making salient his programmatic ideas for further development in the following chapters. It will consists in basic definitions and distinctions which ground the formal apparatus developed twenty years later by Everett.

I call Walton's account of fictional "truth" "functional", following [Friend 2017b]. The relevant paragraph was cited in §1.4.2. However, if we were to follow Walton's terminology, one would rather call it the "generative" account. For the master idea originating in Walton's work is that fictional propositions are *generated* by a fictional work as we will shortly see. Friend's term is, I think, better for it does not rely on a technical term and is also an adequate description of the account. Indeed, fictional texts and works, according to Walton's idea are to be regarded as *props*, that is objects whose *function* is to prompt imagining certain propositions in certain contexts.

In general, the functional account's motto is that p is fictionally "true" iff there is a work of fiction which *prescribes* imagining p. The problem of "truth in fiction" thus boils down to the following question: Which (and how do) propositions get "fictionalised", i.e. are to be imagined relative to a work of fiction? In particular, a proper account of this mechanism should explain how inferences like the one leading to the fictionality of (1) and the non-fictionality of (2) are to be construed.

In this chapter, I will cite Walton's work in order to have the basic definitions on the table. In the following chapter, I will present Everett's attempt to systematise Walton's programmatic ideas; I will discuss the features of Everett's formalism and contribute to its improvement. In chapter 7, I will apply Everett's formalism to

linguistic fictions. This will give us a solution to the problem of fictional "truth" and I will show how this solution gives us important insights on the logical structure of fictions in general.

5.1 Basic definitions

5.1.1 Props

The first basic definition is that of a *prop*. Props, in Walton's terminology, are objects which prompt imaginings in the context of a game of make-believe. As such, in the vast majority of cases, props are *artefacts* whose function is precisely to be used in games of make-believe as prompters of imaginings.[1] Typical examples of props in children's games of make-believe are dolls, toy trucks and the like. Typical examples of props in adult games of make-believe are actors and comedians, pictures and paintings, movies, fictional texts.

Qua artefacts, props have a function. Their function, Walton says, is to *generate* fictional propositions (or fictional "truths") in context, when they are actually used as props in a given game of make-believe. That is to say, the imaginings prompted by props in such and such a game of make-believe typically involve propositional imagination. A child typically imagines *that* this doll is a baby, or *that* this toy truck is driving slowly. A theater audience typically imagine *that* this actor is Hamlet and *that* he contemplates the idea of suicide. When put in such contexts, props do not prompt imaginings accidentally. The rules of the game of make-believe make it the case that any participant in the game is required to at least propositionally imagine that such and such is the case when interacting with the prop in the actual world. This explains how it is possible for participants to coordinate their cognitive systems into imagining *the same* propositions. The propositions thus imagined are the *fictional* (or pretend, or make-believed) propositions.[2]

As Walton puts it:

> Props are generators of fictional truths, things which, by virtue of their nature or existence, make propositions fictional. A snow fort is a prop. It is responsible for the fictionality of the proposition that there is a (real) fort with turrets and a moat. A doll makes it fictional in a child's game that there is a blond baby girl.

[1] According to Walton, some props are natural objects. Some authors disagree with such an extension of the notion though. See for instance [Currie 1990].

[2] What we usually denote by the term "fiction" is but one kind of games of make-believe in this framework.

> Representational works of art are props also.[3]

The idea that props have a *function*, i.e. to prompt imagining, is not very strong. Indeed:

> The prompting function of such things as stumps, snowmen and toy trucks as it is, should not be overemphasised. [...] The toy [truck] is hardly to be credited with prompting *these* imaginings, nor is its movements. [The child] made the toy move in certain ways because of how he wanted to imagine a truck moving, rather than vice versa.[4]

Let us make it a definition: a *prop* is a real-world object which is used in a game of make-believe in such a way that it *generates* fictional propositions. Props do not have general essential features, they are defined by their *function* in the context of a game of make-believe.

5.1.2 Principles of generation

In Walton's setting, fictional propositions are thus generated. This "generation" needs two ingredients. First, a prop which serves as a prompter. Second, the rules of the game in which the prop has a function. For instance there is a rule for games of make-believe which says that the doll *counts as* a baby girl. Consequently, for a participant to imagine the fictional propositions they need to gather information about the prop in the actual world *and* grasp the rules of the game of make-believe they are participating in.

The rules of a game of make-believe are called the *principles of generation*, since they are the "principles" from which one can derive the fictional propositions, given a (set of) props. Here is how Walton introduces them:

> Props generate fictional truths independently of what anyone does or does not imagine. But they do not do so entirely on their own, apart from any (actual or potential) imaginers. Props function only in a social, or at least human, setting. The stump in the thicket makes it fictional that a bear is there only because there is a certain convention, understanding, agreement in the game of make-believe, one to the effect that wherever there is a stump, fictionally there is a bear. I will call this a *principle of generation*. This principle was established by explicit stipulation: "Let's

[3] [Walton 1990], pp. 37-8.
[4] [Walton 1990], p. 24.

> say that stumps are bears." But not all principles are established thus. Some, including most works of art, are never explicitly agreed on or even formulated, and imaginers may be unaware of them, at least in the sense of being unable to spell them out. I do not assume that principles of generation are, in general or even normally, "conventional" or "arbitrary", nor that they must be learned. Nevertheless, what principles of generation there are depends on which ones people accept in various contexts. The principles that are in force are those that are understood, at least implicitly, to be in force.[5]

Let us make it a definition: a *principles of generation* is a rule of inference which makes explicit how one can combine real propositions about props and fictional propositions to yield further fictional propositions. Principles of generation can originate in convention or in arbitrary stipulation. They must accepted by all the participants in a game of make-believe in order for them to play *the same* game of make-believe. This acceptance need not be explicit, though. In this sense, principles of generation are like rules of games.

Once principles of generation have been accepted, they are in force. When they are in force, they should be followed by all the participant of the game. It is very important to highlight that what follows from the principles of generation and the real features of the props need not be *actually* imagined to be fictional. This points towards a first basic distinction.

5.2 Basic distinctions

5.2.1 Imagination and fictionality

From these definitions follows an important distinction between the propositions which are *de facto* imagined and those that are *de jure* fictional. The set of fictional propositions is not what is *de facto* imagined by the participant in the game of make-believe, but what is *de jure* imagined.

In other words, there is a precedence of the "rule" over the "moves" in a game of make-believe when it comes to determining what counts as fictional.[6] This distinction fastens the analogy between truth and fictional "truth". As Walton points out:

[5] [Walton 1990], p.38.

[6] This might create some problems when we consider the interesting practices of literary fanfictions which, arguably, put this basic distinction into question. I will discuss these in chapter 9.

> Being fictional and being imagined are characteristics that many propositions share. [...] But it would be a serious mistake simply to identify the fictional with what is imagined. What is fictional need not be imagined, and perhaps what is imagined need not be fictional.
>
> "Let's say that stumps are bears," Eric proposes. Gregory agrees, and a game of make-believe is begun, one in which stumps – all stumps, not just one or a specified few – "count as" bears. Corning upon a stump in the forest, Eric and Gregory imagine a bear. Part of what they imagine is that there is a bear at a certain spot – the spot actually occupied by the stump. "Hey, there's a bear over there!" Gregory yells to Eric. Susan, who is not in on the game but overhears, is alarmed. So Eric reassures her that it is only "in the game" that there is a bear at the place indicated. The proposition that there is a bear there is fictional in the game.
>
> Or so Eric and Gregory think. They approach the bear cautiously, but only to discover that the stump is not a stump at all but a moss-covered boulder. "False alarm. There isn't a bear there after all," Gregory observes with surprise and relief. And for the benefit of outsiders, "We were mistaken in thinking that, in the world of the game, there was a bear there." Eric and Gregory did imagine that a bear was there, but this did not make it fictional in their game. They do not say that fictionally there was a bear which evaporated when they approached, nor that it is no longer fictional that a bear was there at the earlier time. Gregory takes back his previous claim that fictionally a bear was in the place indicated, and he is right to do so.
>
> Meanwhile, however, unbeknownst to anyone, there is an actual stump buried in a thicket not twenty feet behind Eric. Fictionally a bear is lurking in the thicket, although neither Eric nor Gregory realises the danger. No one imagines a bear in the thicket; it is not fictional that a bear is there because somebody imagines that there is. But it is fictional. What makes it fictional? The stump. Thus does the stump generate a fictional truth. It is a prop. Props are generators of fictional truths, things which, by virtue of their nature or existence, make propositions fictional.[7]

So the distinction is the following: a proposition is *imagined* when someone imagines it; a proposition is *fictional* when someone should imagine it. As shown in the above quote, a proposition can be imagined without being fictional *and* it can be

[7] [Walton 1990], p. 37.

fictional without being imagined by any participant in the game of make-believe.

5.2.2 Fictionality and truth

This important distinction grounds the other main result of Walton's theory: fictionality, though analogous to truth, is is distinct from truth. The distinction was already made in §1.4.3 above, when showing that the functional and the modal account make incompatible predictions about the truth-value of (1) and (2).

What is very interesting is that fictional propositions "correspond", in some very precise sense, to real facts, namely the relevant features of the prop supplemented with the principles of generation in force when considering these features. But, typically, these fictional propositions are not true.

For instance, when Gregory says "Hey, there is a bear over there!" the fact "corresponding" to the proposition expressed is the real fact that *there is a stump over there* (let us suppose Greg is not mistaken). But, of course, there is no bear in reality. So the proposition expressed by Gregory's fictional sentence, although fictional, is false. On the other hand, suppose Gregory says "Hey, I am over there!". The proposition expressed by this sentence is both fictional and true. Indeed, it is clear from the description of the game that, in the game of make-believe, Gregory is Gregory.

So a fictional proposition can be either true or false, depending what the real situation is. *Fictionality* and *truth* are orthogonal concepts.

This distinction is very important, because it explains why fictional propositions are "objectively" fictional. In the following passage, one can see how Walton's basic definitions and distinction are nicely interconnected to ground this "objectivity" of the fictional:

> The role of props in generating fictional truths is enormously important. They give fictional worlds and their contents a kind of objectivity, an independence from cognisers and their experiences which contributes much to the excitement of our adventures with them. This objectivity constitutes another affinity between fictionality, insofar as it derives from props, and truth. The stump game shows that what is fictional, when props are involved, is detached not only from our imaginings but also from what people think and what they take to be fictional. We can be unaware of fictional truths or mistaken about them as easily as we can about those aspects of the real world on which they depend. Eric and Gregory are genuinely surprised to discover that fictionally a bear is lurking in the thicket. It is not thinking that makes it so; the prop does. Fictional

> worlds, like reality, are "out there," to be investigated and explored if we choose and to the extent that we are able. To dismiss them as "figment of people's imaginations" would be an insult and underestimate them.[8]

This notion of "objectivity" is crucial in that it allows to distinguish between *imaginings* and *misimaginings*. It corresponds to the idea that one can be *mistaken* in their fictional reading. Fictional "falsity", like (2), are *misimaginings*. In part 2, when I fine-tune the functional account, this distinction will play a central role and I will try to find the origin of this "objectivity". In this part, the functionalist framework consists in grounding such a distinction without appealing to any notion of modal truth.

5.3 Towards pretence semantics

5.3.1 Normativity and fictionality

The key feature of Walton's functional account is thus its normative component. A proposition is fictional *only if* it *is to be* imagined, in certain contexts involving props and principles of generation. This is how the functional account will provide a solution to the problem of fictional "truth", since one can expect to explain the *contrast* between (1) and (2) without appealing to truth-conditions.

Walton's functionalist account of fictional "truth" is usually summed up in the following definition:

> a proposition is fictional iff there is a prescription to the effect that it is to be imagined. More precisely, a proposition is fictional (n (the world of) a particular work, W, just in case appreciators of that work are to imagine it, just in case full appreciation of W requires imagining it.[9]

However, one might question whether this definition gives adequate necessary and sufficient conditions for fictionality. The sufficiency condition has been called into question several times.[10] Indeed, it seems that, sometimes, there is a prescription to imagine a proposition in a work of fiction but the proposition fails to be fictional in the work. For instance, in a work where a character's dream is represented, one is required to imagine the content of the dream. But the content of the dream fails to

[8] [Walton 1990], p. 42.
[9] [Walton 2013].
[10] Especially in [Friend 2008].

be fictional.[11] I will postpone this problem until I have given a detailed theory of the principles of generation which explains what the prescriptions are.

Second, it might be thought that there are more fine-grained distinctions to be made about the *kinds* of prescriptions one can find in fictions. Indeed, given what has been said above, it seems that some propositions are "more important" than others when it comes to the fictional content of a story for instance. In [Friend 2017b], Friend thus proposed useful subtle distinctions about the nature of the normative component underlying what counts as fictional which are worth having in mind:

> Although prescriptions to imagine are sometimes associated with mandates, we need not imagine everything that is fictional. If we want to understand a work, some kinds of imagining are required. One could not grasp the basic plot of *Gulliver's Travels* without imagining Gulliver travelling to Lilliput, Brobdingnag, and so forth. A fuller appreciation demands recognizing how mistaken Gulliver is about himself (something children often miss). Still, even a full appreciation does not require imagining that Gulliver has internal organs, though it is surely fictional that he does. It is helpful to distinguish these obligations. I will say that a work *mandates* imagining that P if failure to imagine that P would mean falling below a minimum threshold for comprehension. A work *prescribes* imagining that P if we should imagine that P to have a fuller appreciation of the story. Finally, a work *invites* imagining that P on the following condition: if the question arises and we must choose between imagining that P and imagining that not-P, we are required to imagine the former. [Footnote: This is Walton's most careful formulation. Note that where fictions prescribe inconsistent imaginings we should not choose between P and not-P; we are invited to imagine both (regardless of whether we are able to do so).] What is fictional in a work is what the work invites imagining. Although we need never imagine that Gulliver has internal organs, if the question came up it would be absurd to deny that he does.[12]

These fine-grained distinctions will play an important when I will fine-tune the functional analysis in part 2 of this dissertation.

[11] See [Walton 2013] for more counterexamples and attempts to fix the definition.
[12] [Friend 2017b], p.2.

5.3.2 Fictionality conditions

Pretence semantics is based on the existence of such a normative component of fictionality. It then develops the idea into what might be called "fictionality conditions", making the best of Walton's analogy between truth and fictionality. Pretence semantics is thus a "parallel" semantics.

As Everett nicely puts it:

> I should emphasise that, on this view, truth within a pretence is *not* really a species of genuine truth, rather it is a distinct linguistic norm which mimics real truth within the scope of the pretence.[13]

The difficulty of formalizing these ideas lies in the incredible lability of ordinary language when it comes to combining truth-conditions and fictionality-conditions. As will be seen in the following two chapters, the framework one should develop to account for fictionality has to be incredibly flexible. This, to some extend, is a challenge to systematisation.

Now we have all of Walton's building blocks, let us present Everett's pretence semantics.

[13] [Everett 2013], p. 29.

Chapter 6

Pretence semantics

6.1 Everett's functionalism

In [Everett 2013], Anthony Everett took up Walton's idea and introduced what can be termed "pretence semantics". As such, pretence semantics is a version of functionalism. It is thus an alternative to the modal account of fiction. [Everett 2013] can be seen as a specifying the detail of Walton's programmatic ideas as presented in the previous chapter so as to model (linguistic) fictions. Everett uses this formal apparatus to defend substantial philosophical theses, which are mainly in line with Walton's work, thus participating in the debates of the philosophy of fiction (about the ontological status of fictional characters, the metaphysics of emotions, the theory of fictional reference, the question of how we can learn from fictions, etc.). In this chapter, I will present Everett's general formalisation of what pretences or games of make-believe are. I will in particular discuss his way of interpreting the general principles of generation. Application of this general framework to linguistic fictions, so as to find the principle(s) of generation which are constitutive of (linguistic) fictions as opposed to other pretences or games of make-believe, will be the focus of the following chapter.

Pretence semantics is the systematic study of the principles of generations aiming at giving a general theory of "fictionality-conditions". The main concern of pretence semantics is thus to give recursive mechanisms which can predict contrasts between fictional sentences and non-fictional sentences. Consequently, the aim of pretence semantics is to model our intuitions about fictional "truth" and "falsity" which are reflected in the contrast between (1) and (2). It will provide a solution to the problem of fictional "truth".

In practice, it consists in providing a base language with an operator "in the

fiction" which systematically interacts with the known semantic rules for that language. Of course, this operator should not be interpreted as a traditional intensional operator which calls for a modal interpretation. Rather, one should think of it as a context or a background in or against which fictional sentences are to be understood and interpreted. Roughly, the cognitive mechanism underlying this operator is that of pretence or simulation. Pretence semantics is thus the theory of the simulation of ordinary semantics. It is an extension of natural language semantics.

In principle, such an operator is not purely *linguistic* and should work for any other medium in which fiction is possible. However, for reasons of space and competence, I will limit myself to linguistic fictions only. Following Everett, I will also simplify matters a lot and take predicate logic and its traditional set-theoretical semantics as a toy-model for natural language.

Given a base language, pretence semantics adds a fictional or pretence operator $[^f ...]^f$ yielding fictionality-conditions.[1] Any well-formed formula can be put within the brackets and yield a fictional sentence. $[^f \phi]^f$ thus reads "ϕ is fictional" or "In the pretence or fiction f, ϕ".[2] Principles of generation should tell us in which condition $[^f \phi]^f$, i.e. when it is proper to say that "ϕ is fictional".

Everett's interpretation of the existing literature on principles of generation consists in distinguishing between three types of principles of generation, according to their formal complexity. His labels are "Type I", "Type II" and "Type III". Principles of generation are *inference schemas*; they can be thought of as defining a consequence relation between (fictional) propositions. Here are the general schemas before I go into the detail:

(T1) $[^f \phi]^f$

(T2) $\psi \rightarrow [^f \phi]^f$

(T3) If $\phi_1, ..., \phi_n$ are such that $[^f \phi_1, ..., \phi_n]^f$ *and* χ follows from $\phi_1, ..., \phi_n$ (in some specified sense), then $[^f \chi]^f$

We say that ϕ is fictional iff $[^f \phi]^f$ is derivable from an arbitrary long combination of the above principles of generation.

I will call "simple principles" the principles having the form of (T1) or (T2). I will call "elaborate principles" the principles having the form of (T3). In this chapter,

[1] This is Everett's notation. The superscript f stands for "fiction".

[2] I will not distinguish between the two expressions here. The second expression will be useful when the pretence operator operates on a subformula only, for it is easier to read, say, $\forall x A(x) \rightarrow [^f B(x)]^f$ using the second idiom. Nothing hinges on this for now.

I will begin by presenting and discussing the simple principles. Then I will present and discuss the elaborate principle.

The elaborate principles have received a lot of attention in the literature, contrary to the simple ones. I think it is useful to concentrate on the simple principles first, for it gives a feeling of the machinery and I think they raise interesting, undiscussed technical problems.

6.2 Simple principles of generation

6.2.1 Formalizing a simple pretence

Let us formalise Walton's pretence whose participants are Gregory and Eric and which is introduced by Eric stating the following rule: "Let's say that stumps are bears." There are, *prima facie*, two ways of formalizing this principle of generation (Sx stands for "x is a stump"; Bx stands for "x is a bear"):

(11) $\forall x(Sx \rightarrow [^f Bx]^f)$

(12) $[^f \forall x(Sx \rightarrow Bx)]^f$

(11) says that for every object, if it really is a stump, then it fictionally is a bear. (12) says that it is fictional that all stumps are bears. These two pretences, although intuitively similar, have very different consequences.

According to (12), it is fictionally the case that the set of stumps is a subset of the set of bears. Hence, if it is fictional that all bears are such and so, then it is fictional that all stumps are such and so. It so happens that all bears are mammals in the real world. And there is no reason to believe that this is not the case in the pretence.[3] So it is reasonable to accept that it is fictionally the case that all bears are mammals. Given (12), then, it follows that it is fictionally the case that all stumps are mammals. So the extension of "mammality" is strangely altered in this pretence, for it contains, among other animals, dead trees.

There is worse. Suppose there is a stump in front of Gregory. Let us call is s. Gregory would certainly like to infer that it is fictionally the case that there is a bear in front of him. But from $S(s)$ and (12) alone, Gregory cannot infer anything. Indeed, he needs an extra premise which is $[^f S(s)]^f$. One easy way of getting this premise would be to add a principle of generation to the effect that: $\forall x(Sx \rightarrow [^f Sx]^f)$. But then, it would be fictionally the case that s is a stump and that s is a bear. Although

[3] By virtue of a certain (T3) left implicit here. The reasoning is just like the one yielding "princes are human beings" in *Hamlet*.

such weird pretences are arguably possible[4], this is clearly not what Gregory and Eric have in mind when they play.

(11), by contrast, yields the expected fictional consequences. Stumps are stumps and bears are bears; and it is fictionally the case that stumps are stumps and bears are bears. Maybe, if the quantifier is understood without any contextual restriction, it so happens that, within the scope of the pretence, there are simply no stump anymore, i.e. it is fictionally the case that the extension of the predicate "stump" is null. More probably, Gregory and Eric meant to restrict the pretence to the forest they are in, so that there could be stumps elsewhere that are fictionally still stumps. But in general, since there is no reason to believe that the pretence messes up basic biological facts, it is very reasonable to think that: $[^f\forall x(Bx \rightarrow \neg Sx)]^f$. Hence, Gregory has it that $[^f Bs]^f$ (by (11)) and he also naturally infers that $[^f\neg Ss]^f$.

6.2.2 On distinguishing between (T1) and (T2)

Walton's "direct principles"

I understand that what I call simple principles correspond to what Walton calls "direct principles of generation" as opposed to (T3) which are "indirect principles of generation".

By Walton's standard, *directly generated* fictional propositions are those which do not depend on any other fictional propositions. *Indirectly generated* fictional propositions are those which depend on other fictional propositions. To get back to Gregory and Eric: "This stump is a bear" is directly generated, since it is derivable from real world information and a (T2) principle like (11). By contrast, "There is a bear in front of Gregory" is indirectly generated, since it is derivable from the two fictional propositions that "There is a bear there" and "Gregory is there".

This being said, one must refrain from identifying the directly generated fictional propositions with the explicit fictional propositions and the indirectly generated propositions with the implicit fictional propositions. As Walton readily observes:

> Any inclination to suppose that the propositions whose fictionality a literary work generates directly are simply the ones its words express, given the language in which it is written, dissipates quickly. Many of these propositions are not fictional at all, most obviously in the case of works with "unreliable" narrators (such as Ford's *Good Soldier*), and when they are fictional their fictionality is often implied rather than primary. It is

[4]Imagine a mad surgeon whose purpose in life is to make chimeras. He would surely try to make stump-bears.

> fictional that someone (the narrator) utters the words of the text, in many cases, and, if the narrator is "reliable," this implies the fictionality of what the words express. Is it always fictional at least that someone utters the words in question: Does *this* constitute the core of primary fictional truths? No. Sometimes it is fictional merely that someone thinks those words without uttering them, or that they express his fantasies or dreams or desires. Sometimes, perhaps, there is no narrator and it is simply what is expressed by the text (taken perhaps in one or another nonliteral manner) whose fictionality is generated directly. It is not uncommon for readers to be very uncertain which of these alternatives obtains.[5]

So there are some reasons to group (T1) and (T2) into a kind. Although Walton does not distinguish between the two, I think Everett has indeed good reasons to distinguish them.

(T2) is prop-relative

The above example illustrates the difference between principles of type (T1) and (T2). Roughly, one sees that (T2) principles are coordinating some real information channels with a fictional interpretation of this information. This is the formal rendering of the intuitive "count as" or the "is" of identity holding between props and their fictional counterparts as in expressions like: dolls "count as" babies, or mud globs *are* cakes. More precisely, it creates a mapping between the individuals of ordinary discourse and those of the fictional discourse. Consequently, talking *about* these stumps corresponds to fictionally talking *about* bears, although there is no actual bear.

Principles of type (T2) should thus be understood as modeling the function props have in a pretence. Props' function is precisely to "count as" such and such within the scope of the pretence. One can see how props ground a kind of "objectivity" within the scope of a pretence by coordinating real propositions with fictional propositions in Everett's framework.

Consequently, (T2) makes it the case that, in principle, many things that are true of the real-world prop will be duplicated in the pretence with the relevant fixes. If the stump is huge, it will prompt the imagining that fictionally the corresponding bear is huge. But not everything that is true will be thus duplicated. If it is true that the stump is a dead organism, it will not prompt the imagining that fictionally the corresponding bear is a dead organism. The line between what is so imported and left over is taken care of by (T3) principles, which we will discuss later.

[5] [Walton 1990], p. 170.

In practice, principles of type (T2) are either conventional as in the case of dolls, or they are deployed only in the context of a specific pretence as in the case of Gregory and Eric playing with stumps. What is so remarkable is that we are very good at taking up on non-conventional (T2) on the spot and even those which contravene conventional (T2). For instance, it is very easy to establish a (T2) principle to the effect that dolls are fictionally evil creatures and not babies.

By contrast, (T1) principles are stipulated. They generate fictional propositions by fiat, without any corresponding object in the real world. Everett's example is:

$$[^f Pigs\ can\ fly]^f$$

As such (T1) are much less "natural" that (T2) principles: the above example illustrates this forcefully. The labels follow *logical complexity* and not conceptual priority. It may very well be that the acquisition of (T1) principles is more difficult and comes later in development than (T2) principles as the level of abstraction suggests. But the labels are justified when we look at the complexity of the formulae.

Existentially conservative and creative pretences

I think Everett's distinction between (T1) and (T2) is even more substantial. It is, I propose, the formal rendering of a distinction made by Evans between existentially *conservative* and existentially *creative* pretences.

Conservative pretences are pretences which quantify over the same entities than ordinary discourse, whereas creative pretences quantify over some other entities, not in the domain of ordinary discourse. Evans gives interesting examples of very basic creative pretences. One is the following. Take a practicing boxer who is, within the scope of his pretence, fighting an opponent. He would deal and dodge blows to and from the imaginary opponent. Let us call our practicing boxer Bob. It seems that the natural way to formalise such that (with F standing for "fighting against" and B for "being a boxer"):

$$[^f \exists x(Bx \wedge F(Bob, x))]^f$$

i.e. it is fictionally the case that Bob is fighting with a boxer. But nothing corresponds to this boxer in the real world: there is nobody opposite Bob in the real world.

The practicing boxer's pretence is, in that respect, very different from Gregory and Eric's one. (T1) principles are essential to the formalisation of existentially creative pretences whereas (T2) principles cannot formalise them. Hence the importance of the distinction between (T1) and (T2).

One way of putting it is that existentially creative pretences free themselves from real features of props. In the practicing boxer pretence, *there is no prop* which stands for the imagined boxer opposite Bob. These pretences are arguably "less rich" that ones using props, for many things are left underspecified if not stipulated. For instance, one might say that the height of Bob's opponent can be inferred indirectly by Bob's punches: if $[^f$ he punches his opponent in the face$]^f$ then one can infer the height of his opponent, using a principle of type (T3). Same for his exact position. But as for, say, the skin-colour of Bob's opponent, it is arguably impossible to infer it. The only way to settle the matter would be to devise a (T1) principle to that effect.

This underspecification is a matter of degree though, for many things are also underspecified when there is a full-fledged prop. This phenomenon of indeterminacy in existentially creative pretences is at the core of the second part of this dissertation.

Existentially creative pretences and language

As for linguistic fictions, there are two cases which force existentially creative pretences. The first is when the ϕ in a (T1) principle is an existential formula. The second is when the ϕ in a (T1) principle contains a singular term which does not purport to refer to any real world individual. Indeed, as Evans eloquently puts it:

> The kind of pretence which is involved in the conniving use of empty singular terms is *existentially creative*; the pretence is not that something which there is is other than it is, but that there is something which in fact there isn't.[6]

Conversely, an existentially conservative pretence is one in which the following schema is valid:[7]

$$[^f \exists x \phi x]^f \rightarrow \exists x [^f \phi x]^f$$

This tells us how we should interpret principles of (T1). They are a formal tool to track down one's ontological commitments in arbitrarily complex pretences.

Here is an illustration of such tracking. Suppose, for instance, that one reads Leibniz's *Monadology* without believing it at all but as a fiction. One will, roughly,

[6] [Evans 1982], p. 358.

[7] Note that this formally corresponds to the Barcan formula in modal logic: $\Diamond \exists x Ax \rightarrow \exists x \Diamond Ax$, which has generated a lot of metaphysical debates about so-called "actualism" and technical developments in variable domain modal logics. See [Williamson 2013] (§2) for a discussion of actualism and [Priest 2008a (§15) for a discussion of variable domain logics.] This might be another interesting bridged between the modal account and the functional account.

have to imagine that for every real human being, there is a soul which is intimately related to it, and each one reflects the whole of the universe, etc. The principle of generation formalizing the first sentence will thus be (with H "being human", S "being a soul", and R "being intimately related to"):

$$\forall x(Hx \rightarrow [^f\exists y(Sy \wedge R(x,y))]^f)$$

This principle of generation is complex. Its overall structure is of (T2). However, one can easily derive from it that:

$$[^f\exists y Sy]$$

which is a (T1) principle. Such derivations of (T1) from complex (T2) principles are a formal tracking of one's ontological commitments within complex pretences.

This last remark gives good reasons for holding that (T2) principles are conceptually primary and "more natural" than (T1) principles which are conceptually derived, as suggested above. However, it also shows that they are a necessary feature of the general framework for one must be able to distinguish and model both the existentially conservative and the existentially creative pretences in pretence semantics.

6.2.3 Technical points

As mentioned above, simple principles have not received much attention in the philosophical literature. I think they raise very interesting technical issues though. These technical issues, I will try to show, point to substantial philosophical problems. I do not claim that they raise daunting problems for Everett's formalism though.

Two technical points I want to discuss now stem from the fact that pretence semantics, as presented above, is a semantics for a ordinary language supplemented with a fictional operator. The well-formed formulae of the extended language are thus the ones inherited from the base language plus the well-formed formulae containing the fictional operator. My first worry focuses on the non-logical vocabulary and my second worry is about the logical vocabulary of this extended semantics.

Non-logical vocabulary

Since pretence semantics is an extension of natural language semantics, natural language predicates carry over to pretences. In other words, predicates keep their meaning within pretences. Of course, natural language predicates need not have the same

extension within a pretence. As Gregory and Eric shows, the extension of the natural language predicate "bear" is significantly altered within their stump-bear pretence.

The general idea that the meaning predicates should be be the same outside and inside pretences is a good thing, for changing the meaning of words is usually not expected when engaging in a pretence. As Everett remarks:

> Suppose I am pretending that the earth is flat so that [^{f}the earth is flat]^f. If we do not hold the meaning of the empty vocabulary fixed then, at least arguably, the following [pretence could be] correct:
>
> **(F)** [T]he sentence "the earth is flat" [is] true [and] the earth would still be round but the word "flat" would have a different meaning.[8]

This does not mean that there are no metalinguistic pretences which precisely consists in pretending that words change their meaning. Such pretences are indeed possible as Everett points out.[9] But it means that such metalinguistic pretences are not expected as a default pretence. In general, without mention to the contrary, words used within a pretence have the same meaning as when they are used outside of the pretence.

However, there are pretences in which new predicates are introduced. Let me give an example first. In *Lolita*, the fictional narrator Humbert explicitly introduces the term "nymphet" in the following passage:

> Now I wish to introduce the following idea. Between the age limits of nine and fourteen there occur maidens who, to certain bewitched travelers, twice or many times older than they, reveal their true nature which is not human, but nymphic (that is, demoniac); and these chosen creatures I propose to designate as "nymphets".[10]

He subsequently discuss the meaning of the term and its extension at length.

Now, one would certainly want to make it fictional in *Lolita* that "Humbert says that Lolita is a nymphet". But since "nymphet" is not a real predicate, one cannot express such a sentence in pretence semantics.

[8] [Everett 2013], p. 24.

[9] Note that the possibility of such metalinguistic pretences is not a problem for Everett's view. Indeed, since natural language can express metalinguistic propositions, then an extension of such a language can express these metalinguistic propositions and make them fictional.

However, this is not expressible in first order logic. Although my oversimplification of the base language can suggest that Everett runs into trouble, it merely says that first order logic is not very powerful, which is a matter of course.

[10] [Nabokov 1955], §I.4.

The natural response is that since natural language is such that one can explicitly introduce new predicates, then so is the extended language. This, I propose, should give reasons to distinguish between predicatively conservative and predicatively creative pretences and generalise Evans's distinction discussed above.

Furthermore, it so happens that some fictional predicates *become* real predicates, so to speak. Indeed, "nymphet" now commonly appears in English dictionaries. Here is the Merriam-Webster online entry (in which no mention of the fictional origin is given!):[11]

> Definition of nymphet: a sexually precocious girl barely in her teens.
> *also*: a sexually attractive young woman.[12]

So it has to be possible that fictional predicates scope out of the pretence they originate from. This suggests some interesting, and potentially problematic, dynamics that Everett's formalism lacks. Accommodating Everett's framework to account for this linguistic phenomenon is an interesting research program for the future.

Logical vocabulary

For the same reasons, the logical vocabulary should be held fixed. And this is a good thing, for if one encounters "and" in a pretence, one is clearly expected to understand "and" as outside the scope of the pretence. Of course, as for metalinguistic fictions, weird metalogical fictions can happen, but this is not the default case.[13]

But this may appear as a problem when it comes to the meaning of quantifiers. Indeed, according to classical quantification theory, quantifiers are existentially loaded, which means that quantifying over individuals entails that they exist. This feature should be encapsulated within the scope of a pretence, for, by definition, there are things one can quantify over in an existentially creative pretence which do not exist in the real world. For instance, we can say, drawing on the practicing boxer pretence, that $[^f$there is a boxer opposite Bob$]^f$. But it does not follow that there is an individual in the world which $[^f$is a boxer and opposite Bob$]^f$.

This neutralizing the existential import of the existential quantifier is, some would argue, a change of meaning..[14] I will address these issues more specifically in part 3. But, for the time being, I want to make salient a potential problem about the interactions between the existential quantifier and the fictional operator.

[11] Though they mention that the term appears in 1955.
[12] https://www.merriam-webster.com/dictionary/nymphet.
[13] These metalogical pretences, for the same reasons, do not threaten pretence semantics.
[14] See Quine's argument that such things would "change the subject", in [Quine 1986], p. 81.

An existential quantifier cannot, in general, scope out of a fictional operator. Indeed, in an existentially creative pretence, such scoping out is not allowed. As seen above, the tracking of ontological commitments is supposed to tell us when such scoping out is not allowed. Such keeping track is thus very important.

From a cognitive viewpoint, it is not clear how such tracking is done, although it appears that we do it very well. For instance, in a complex pretence within which some real things and some fictional things are related we usually have no problem distinguishing between the two. Any historical novel containing both historical figures and purely fictional characters shows that humans perform very well in this tracking task.

This tracking problem in the abstract consists in extracting the normal prenex form of any fictional sentence, following the *Monadology* example given above. Since not all quantifiers can scope out of a fictional operator, it follows that not all formulae of pretence semantics will have a prenex form with all the quantifiers on the front of the formula. Introducing a fictional operator thus invalidates the basic theorem of predicate logic according to which every well-formed formula has a prenex form.

Given our cognitive ability to do the tracking, it seems that the tracking of ontological commitment is not completely unsystematic, though. So this suggests that one can try and give a basic theorem for pretence semantics which corresponds to giving as much a prenex form as possible. To illustrate this point, consider a toy example of a historical novel in which there are two characters: h a historical figure and f a fictional character and it so happens that they meet. This would be rendered in pretence semantics as (with M a relation standing for "meet"):

$$[^f hMf]^f$$

From this one would infer:

$$[^f \exists x \exists y xMy]^f$$

Given what we know about h and f one would then infer:

$$\exists x[^f \exists y xMy]^f$$

This is as close to a prenex form as one can get. It entails one ontological commitment:

$$[^f \exists y xMy]^f$$

I contend that giving such an algorithm for extracting near-prenex from from which one can read off one's ontological commitment is going to be very difficult in general. This is also an interesting research project for the future.

Here is one way one *should not* do it though. One intuition would be to say that there are two domains of individuals for the quantifiers. Let us call them the fictional (F) domain and the real domain (R). In complex pretences, by definition, $F \cap R \neq \emptyset$. At this point, it really sounds like the quantifiers *change* meaning when moving in and out of the fictional operator. If one want to fix the meaning of quantifier so that it does not change inside and outside a fictional operator, there is an intuitive way to do so which is the following. Define as the domain of quantification $F \cup R$ and then restrict the quantification in and out of the fictional operator, introducing predicates standing for "being a historical figure" and "being a fictional character" (or any other equivalent one might prefer). This nicely looks like natural language quantification where the domain is determined contextually.

But this entails a Meinongianism, according to which one can truly say that "there are things which do not exist". Indeed, among the characters in the history novel, some exist (the "historical figures") and some others do not exist (the "fictional characters"). Consequently, it is true that: "there are things which do not exist". Everett's formalism, ironically enough, would be a reason to adopt Meinongianism. This cannot be the case, for Everett is wholeheartedly committed to anti-Meinongianism.

I think the best way out of this is conundrum to deny that the quantifiers are existentially loaded in the first place. This entails that the meaning of quantifiers is given by free logic. I think that free logic can be fruitfully combined with pretence semantics. I will show how in part 3 of this dissertation.

6.3 Elaborate principles of generation

Let us now turn to elaborate principles of generation. Simple principles modelled the way we get fictional propositions from real propositions. Contrary to simple principles, elaborate principles are the principles which model how we get from fictional propositions from other fictional propositions. As such elaborate principles are the formal rendering of the intuitive idea following which one automatically fills in the background of the fiction using information about the real world. They are the key to the access to implicit fictional propositions which are at the core of the problem of fictional "truth". This is the reason why they have received more attention in the philosophical literature.

Here is how Everett introduces them:

> In addition to Type-I and Type-II principles of generation, our episodes of imagination and pretence often seem to be governed by various more general principles which allow us to fill in the background of

> our imaginative scenario on the basis of what the real world is actually like or what it would be like were the content of our pretence to really obtain, and so on.[15]

There are two main strategies to "fill in" the fictional background. Either one imports into the fiction what the world really is like, or one imports into the fiction how the participants think the world is like. The first way yields versions of what is usually called the "reality principle"[16], the second version of the "mutual belief principle".[17]

It should be clear that some mixture of the two principles are needed to model the way pretenders and readers "fill in" the background of their pretences. The equilibrium between the two principles is a delicate matter, it will be discussed in part two where the fine tuning of the theory is discussed. For now, I will consider Everett's version of the principles. I will then discuss whether cases which suggest that other elaborate principles of generation should also be considered.

6.3.1 The Reality principles

Here is the core intuitive idea behind the reality principles, expressed by Walton:

> The basic strategy which the Reality Principle attempts to codify is that of *making fictional worlds as much like the real one as the core of primary fictional truths permits*. It is because people in the real world have blood in their veins, births, and backsides that fictional characters are presumed to possess these attributes.[18]

Everett interestingly distinguishes two (T3) principles to formalise this intuitive idea:

> *Incorporation*
> If B is really true then $[^f B]^f$, unless there is a set of sentences $A_1, ..., A_n$ such that (i) $[^f A_1, ..., A_n]^f$ and (ii) if $A_1, ..., A_n$ were to be true then B would not be true.

[15] [Everett 2013], p. 22.

[16] As for alternative terminology; [Ryan 1991] calls it a "principle of minimal departure". [Woods 2018] calls it the "world-inheritance thesis".

[17] The terms come from [Lewis 1978].

There is a debate about whether the (T3) principles are principles which warrant *inferences* or if they merely *import* the fictional propositions, so that the inferences are done within the scope of the fiction and not outside of it. [Friend 2017b] alone defends the second view against the main tradition, calling it the "reality assumption". I tend to favour Friend's although nothing hinges on this here.

[18] [Walton 1990], p. 145.

Reality

If (i) $A_1, ..., A_n$ are the sentences that count as true within the pretence in virtue of Type-I and Type-II principles of generation, and (ii) if $A_1, ..., A_n$ were to be true then B would also be true, then $[^f B]^f$.

Incorporation is the idea of the reality principle applied to *facts*. Intuitively, if something is the case, then is is fictionally the case, unless contradicted. This principle brings in *many* fictional propositions. In particular, many propositions which are completely unrelated to the pretence. For instance, to take Walton's famous example, since it is a fact that Aristotle was the master of Alexander the Great, and nothing in Gregory and Eric's pretence runs against this worldly fact, then "Aristotle was the master of Alexander the Great" expresses a fictional proposition, even though Gregory and Eric have never heard of this philosopher before. So this principle over-generates. On the other hand, it is useful and intuitive in some respects: if it suddenly starts to rain in the forest, then it is fictional that it rains within Eric and Gregory's pretence.

Reality, by contrast, typically extends the pretence by taking into account what the pretence is about. Intuitively, it says that one should draw all the consequences of the fictional propositions, following the laws of reality, unless indicated explicitly to do otherwise. This is why, in Eric and Gregory's pretence, the bear is not a stump. Indeed, in reality, if something is a bear, it is not a stump. Hence in Eric and Gregory's pretence, without mention to the contrary, if this is a bear, it is not a stump. The principle also generate many irrelevant fictional propositions. For instance, it follows from *Reality* that Hamlet is either of blood group O or not. One might argue that this fictional "truth" has nothing to do with the fictional events.

Overgeneration is not a big problem at this stage. Indeed, other less general principles of generation can come to the rescue and restrict the general ones. We shall see in part two that the literary genre of a linguistic fiction arguably imposes such restrictions. Undergeneration is the issue, though, for it signals that some *general* principle is missing.

The typical example which shows that these principles of generation undergenerate is an implicit "flat-earth scenario". Imagine that a fiction originates in a society where everyone believes that the earth is flat. Suppose that in such a fiction, there was a fictional event which would make sense only if it is fictional that the earth is flat, although it was nowhere explicitly mentioned that, fictionally, the earth is flat. For instance, in the fiction, a boat sails until it reaches the end of the world and it is said to fall into the darkness. In such a case, the above reality principles are helpless. What happens is very clear though: people import in the fiction their erroneous beliefs. This leads to the so-called mutual beliefs principles.

6.3.2 The Mutual Belief principles

Here is the core intuitive idea behind the mutual belief principle, expressed by Walton:

> Mutual Belief Principle directs us to extrapolate so as to *maximise similarities between fictional worlds and the real world not as it actually is but as it is or was mutually believed to be in the artist's society.*[19]

Everett distinguishes two versions of this core idea:

> *Mutual Belief I*
> If (1) $A_1, ..., A_n$ are the sentences that count as true within the pretence in virtue of Type-I and Type-II principles of generation, and (2) those who created the pretence believe that if $A_1, ..., A_n$ were to be true then B would also be true, then $[^f B]^f$
> *Mutual Belief II*
> If (1) $A_1, ..., A_n$ are the sentences that count as true within the pretence in virtue of Type-I and Type-II principles of generation, and (2) those who participate in the pretence believe that if $A_1, ..., A_n$ were to be true then B would also be true, then $[^f B]^f$

Note first that Everett's formalisation is a qualification of his own *Reality*, but not of *Incorporation*, relative to creators and participants. It is thus possible to define another pair of Mutual belief principles in the following way:

> *Mutual Belief i*
> If B is believed to be true by those who created the pretence then $[^f B]^f$, unless there is a set of sentences $A_1, ..., A_n$ such that (1) $[^f A_1, ..., A_n]^f$ and (2) if $A_1, ..., A_n$ were to be true then B would not be true.
> *Mutual Belief ii*
> If B is believed to be true by those who participate in the pretence then $[^f B]^f$, unless there is a set of sentences $A_1, ..., A_n$ such that (1) $[^f A_1, ..., A_n]^f$ and (2) if $A_1, ..., A_n$ were to be true then B would not be true.

I think it is possible to distinguish between MBI/II and MBi/ii. That is: just as *Incorporation* and *Reality* often coincide, so will the MBI/II and MBi/ii. But in some cases, they will differ.

[19] [Walton 1990], p. 152.

For instance, suppose that Conan Doyle, while writing *the Adventures of Sherlock Holmes*, firmly believed that "the Loch Ness monster exists". But he never mentions the fact in his fiction, nor does he write anything that contradicts this belief. Then, it follows from MBi that $[^f$the Loch Ness monster exists$]^f$, but not from MBI.

The distinction between the I/II and i/ii versions of the principles will be salient when the creator's ideology greatly differs from the participants. This is, of course, typically the case when we appreciate fictions which originate in remote cultures (in time or space). Some arbitration between the two versions should be made explicit. Again, this is a task for a more fine-grained view on the problem, which is discussed in part 2.

6.3.3 Other elaborate principles?

As such, the general principles are far too general. But the systematic feature of the principles should be appreciated as a framework in which things can be restricted by less general principles. The question is rather whether this framework still under-generates. As Everett puts it:

> There are no doubt other general principles of this form which might be deployed. Note that in general a pretence might not employ any of these principles in an unrestricted manner; perhaps *Reality* should be applied to certain sorts of "basic" truths within a given pretence to generate further truths, but not to others. Indeed it might be that for some pretences *Reality* should be applied to certain sorts of "basic" truths while a version of the Mutual Belief Principle should be applied to others. Moreover the output of these principles might be restricted in certain sorts of ways, for example perhaps in some circumstances we employ *Incorporation* but restrict its output to exclude truths about remote areas of space-time, or other truths that are of no relevance to our pretence. So the same pretence might involve the application of a number of distinct Type-III principles of generation, and those principles might not apply unrestrictedly.[20]

There are indeed many interesting complex cases whose study will give us a sense of how the framework could be handled. Walton discusses some under the label "other implications" in his [Walton 1990] §4.3. The examples suggest some elaborate principles which are different from the reality and the mutual belief ones, although I doubt that they are as general, in the sense of being present in *every* pretence.

[20] [Everett 2013], pp. 23-4.

Iconic principles of generation

By iconic principles of generation, I mean some features which suggest "iconically" that such and such proposition is fictional without being either "realistic" or pointing to any background belief. Comic books provide many beautiful examples, some of which I want to discuss.

Whirls and lines around on object can be used to suggest that some object is moving in such and such direction or how fast it moves or the noise it makes within the pretence. For instance, a bell ringing is usually represented as a bell with lines around it. Another example, to represent that a character jumps, usually one draws some lines below it. But these lines are not realistic. Indeed, it does not correspond any visual representation. Neither is it believed to accompany the movement or the sound it suggests. There are "iconic" in an intuitive sense.[21]

Arguably, such iconic principles of generations can be found in linguistic fictions as well. For example, one can play with the fonts and written symbols in a very powerful way to generate fictional propositions. It is very usual to associate the sise of the font with the volume of what is said: the bigger the louder; and often, some unpronounceable symbols suggest that a character is fulminating against somebody or something.[22] The generation of these fictional propositions has nothing to do either with reality or with a mutual belief.

These two examples of "tricks" to generate pretend propositions suggest many others and the reader will easily find their preferred examples. I think these point to the practice of caricatures in general whose main aim is to suggests things while flying in the face of any sort of realism qualified or not. As such, I doubt that these are *general* principles of generation. They seem to be tied up with a genre, akin to

[21] Another beautiful example of such line are the cars in *Tintin* by Hergé: here are some examples: http://dardel.info/tintin/tout.html. Hergé had an even more impressive trick to suggest the fact that, within the pretence, the car is accelerating a lot: he would tilt the car forward, as if the car was trying to go faster than its wheels. This is by no means realistic, nor is it believed that cars are deformed by acceleration to such a visible degree. But the fictional inferences are very easy to get, and no one will end up imagining that the car is physically distorted in the fiction though this is what is represented. I think one of the reasons why the inference is easy to get is that it recalls the feeling of inertia when the car is suddenly braking, for instance. Note however that, if Hergé was following physics, the chassis of the car should be tilted *backward* to suggest acceleration.

[22] A very striking example of this can be found in *the Adventures of Asterix* written by René Goscinny and illustrated by Albert Uderzo. The systematic use of different typeface fonts is used to prompt the imagining of different phonological accents: German-speaking characters would "speak" in a Gothic typeface font, Greek-speaking character use an adapted typeface from the Greek alphabet, and so on. This trickery is "iconic" is an intuitive sense as well, although it is not a matter of figurative resemblance.

caricature, which broadens the generative features of the general framework.

These iconic principles also suggest that there is something about the interaction between the medium and the content. Numerous fictional propositions are obtained by some thinking about the medium itself, its expressive power and its limits, following tentative inferences. For instance, if you are at the beginning of a book and the main character is in danger, you might correctly infer that the character will not die *now*, since there is so much left to be read. There are straightforward *prima facie* formalisations of such inferences in the pretence semantics framework. I doubt that all such inferences can be grouped into a plausible general principle of generation, though.

Conventional principles of generation

Conventional principles of generation rely on more or less explicit conventions. For example, it is a convention in many pretences that witches wear pointy hats; it is a convention for many Western movies that goodies are dressed in white.[23] So that if you see a woman wearing a pointy hat, or a cowboy dressed in white, you can safely infer that, within the relevant pretences, the woman is a witch and the cowboy is good. Of course, this has nothing to do either with real facts or with mutual beliefs.

One might think that these are typically (T2) principles of generation since the convention systematically links a real feature of the prop (the pointy hat drawn or represented in the picture or the movie) and a fictional proposition (the woman wearing it is a witch).

But the fact that these conventions are tacit and occur in many fictions invite us to consider that they are more general. Indeed, (T2) principles as defined above are tied up to a pretence. So one should have a way to explain how such and such (T2) principle can become "famous", so to speak, and be in force in many different fictions. This points towards the notion of a genre, a less general set of principles of generation which does not apply everywhere, but sufficiently often that it gains recognition in the form of a (T3) principle.

One way one might want to formalise such genre principles in Everett's formalism is to introduce the idea of a "super-pretence" which is a set of pretences sharing some (T2) principle. A super-pretence is a pretence which is part of many individual pretence. For instance, there would be a super-pretence within which witches wear pointy hats. There are indeed many individual pretences according to which there is a (T2) principle to the effect that witches wear pointy hats. For example as my little cousin's drawing a witch is one; and *Harry Potter* is another. The (T2) principle

[23]The examples are taken from Walton.

would look like this:

$$\forall x(Pointy\ hat(x) \rightarrow [^{f}Witch(x)]^{f})$$

This way of formalizing genres is not implausible, for the genre conventions are not completely general either: it is easy to imagine a story in which there are some witches but they never wear pointy hats, or maybe some do and others don't depending on their personal taste. The study of the relation between pretences and super-pretences, as well as the dynamics of super-pretences corresponds to the conceptual studies about genres and their sociology. It is a difficult and interesting matter which, I suggest, can very well be combined with Everett's formalism.

With this in mind, we can explore more complicated cases of conventional principles of generations worth discussing. First, there are sets of conventions governing the identification of saints. In order to recognise a Christian saint when you look at pictures, looking for their attributes is generally helpful. For instance, Saint Sebastian is usually very easy to identify since he is the one who has one or several arrows piercing his body; Saint Francis of Assisi is often represented with animals since it is said that he could talk to them. The set of conventional attributes associated with a character is a subject of study in and of itself. Many historians and art critics study it and the field is famously full of controversies and debates of identification.[24]

The interesting part is that the vast symbolism of Christianity has been slowly and intentionally constructed by the Roman Church, among other authorities, for many centuries. Historians show that this symbolism, that is the set of conventional emblems, has systematic connections with other symbolisms coming from other cultures and religious beliefs. So the conventions have a complex history which may influence the generated, fictional propositions. In such a representation, because the character is represented with vine, one might argue that it is Dionysius or Christ, depending on the "conventional background". These cases are extremely tricky.

In order to see these links between sets of conventions, one has to step back from the idea that everything is rooted in beliefs or facts. This is why these kinds of conventions seem to fall out of the scope of the reality and mutual-belief principles. For instance, Christian painters in previous centuries used Greek and Roman symbolism to generate fictional propositions about their fictional characters or their political figures without mutually believing in Greek and Roman mythology. Such prince is fictionally represented as Hercules and that fictionally tells us something about the prince himself: namely that he is glorious, or whatever. Conventional principles

[24]See for instance what Carlo Ginzburg describes about the methodology of the late 19th century art critic Morelli in his "Traces" in [Ginzburg 1989].

of generation, I contend, run much deeper than the atemporal principles we have considered so far and they are needed to infer some fictional propositions.

Another interesting example of such *ad hoc* conventions which is not of type (T3) is that of the *leitmotiv*. Leitmotivs are musical phrases associated with some person, place, event or idea. Hearing a leitmotiv, one can infer, for instance, that, within the pretence, some particular person is about. Or course, nobody believes that people, or events, or ideas in the real world are accompanied with a particular tune. So the above inference is not licensed by any (T3) principle. Rather, there is an implicit convention, following which if you hear this tune, then this character must be about. Leitmotivs have been theorised and widely used by Richard Wagner in his operas. An extreme example of leitmotivs can be heard in *Peter and the Wolf* by Sergei Prokofiev in which the fictional characters, so we are required to imagine, *are identical* with their associated tune.

Leitmotivs can be implicit to a high degree which might question how one should render them in pretence semantics. Indeed, leitmotivs are sometimes implicit to the point of being completely unconscious: the underlying mechanism being that of reinforcement learning. The audience automatically associates this with that, makes the correct inferences in their mind, without having the slightest idea why this is fictionally so. Some not very subtle cinema music works this in using this automatic mechanism in order to help the audience orient in the plot, or maybe to grossly manipulate the audience into imagining this and that.[25]

This automatic feature of the inference seems to clash with one basic feature of pretence semantics which is the possibility of *misimagining*. To use Walton's terminology presented in the previous chapter, this seems to conflate *de facto* and *de jure* imagining. I such, I suppose the pretences underlying these uses of leitmotivs would yield special cases of pretences in which fictional propositions are generated unconsciously.

6.4 Conclusions

This concludes the presentation of Everett's general framework in which all kinds of pretences can be modeled. This framework is called pretence semantics, because it aims at yielding fictionality-conditions just like natural-language semantics aims at delivering truth-conditions for sentences of the language.

The core of pretence semantics is the possibility of formalizing principles of generation to generate fictional propositions. There are two important distinctions to

[25]The point was made by Adorno in *In Search of Wagner*

be made between kinds of principles of generation. First, one should distinguish between *simple* and *elaborate* principles of generation. Simple principles of generation are principles of generation which tell you how to get fictional propositions from real propositions only. Elaborate principles of generation are principles of generation which tell you how to get fictional propositions from real and fictional propositions. Elaborate principles, because they model inferences from explicit fictional propositions to implicit ones, have received much more attention in the philosophical literature.

The second distinction is internal to the simple principles. It is the distinction between (T1) and (T2) principles. (T2) principles aim at formalizing the *function* of props. As such, they are used to generate fictional propositions from the recognition of real features of props. (T1) principles, by contrast, do not concern the props of a pretence. They are crucially related to what Evans called existentially creative pretences. I suggested that they should be used to formally track down our ontological commitments when we are engaged in complex pretences like that of reading a historical novel where historical figures meet fictional characters. This naturally connects to problems about reference within the scope of pretences which are going to be discussed at length in part 3. As such, discussing the technicalities of simple principles is philosophically substantial, even though it has not been done much in the philosophical literature. I pointed to open problems and showed how the framework can improve by producing new distinctions and calling in free logic.

As a general framework, pretence semantics should be able to model *all* kinds of pretences. I suggested that some pretences, at first sight, resist being expressed in the general framework for the kinds of inferences they licence seem to be of neither type. I also showed that they could probably be accommodated into the framework, given a precisification of the *prima facie* problematic inferences.

As can be seen from this exposition, pretence semantics is not a mature formalism. I think this should give reasons to contribute and improve it. Having a general framework for all pretences is useful to discuss cases and solve philosophical problems like that of fictional "truth". In the following chapter, I will thus focus on fictional "truth" by applying the general framework to linguistic fictions like *Hamlet*, as opposed to other pretences like the mud-pie game. Il will show that they exhibit a common two-fold logical structure which is very interesting to investigate.

Chapter 7

From pretence to fiction

The principles of generation discussed so far are general, which means that they should apply to all kinds of pretences alike. Fictions, and in particular linguistic fictions, are pretences. So the general principles of generation should apply to fictions.

In this chapter, I will try and find the (set of) principles of generation which are constitutive of fictions, i.e. the generation principles which apply to all fictions *qua* fictions, as opposed to other kinds of pretences. These principles of generations should set apart children's games of make-believe and fictions. I will focus on linguistic fictions, but the following arguments should apply *mutatis mutandis* to other media as well.

It can be argued that such an inquiry is bound to fail, because the word "fiction" in natural language is too versatile and does not denote a clearly defined set of objects or practices.[1] Though the boundary between fictional pretences and non-fictional pretences is clearly vague and breeds many borderline cases, I think it is fair to say that there are also clear-cut cases. Gregory and Eric are not engaging in a fiction, in the ordinary sense of the word. By contrast, reading *Hamlet* is clearly engaging oneself in a fiction. Maybe what I am after should better be termed *stories* or *narrative* fictions. As we will shortly see, the "fiction principles" aim at formalizing the distinction between the author(s) and the readers(s) of a fictional text. It seems that in the case of Eric and Gregory a clear cut distinction between author and readers is not relevant, for they are both creators and participants in their game. By contrast, in the case of (published) linguistic fictions, the division of labour between creating and enjoying a pretence seems to be brought to completion. Readers are, in general, not expected to participate in the sense of putting forward new principles of generation. They are, in general, expected to draw inferences from principles of

[1] See [Renauld 2014] for such an argument.

generation which are *already there*. But there are counterexamples, which will be studied in due time.

The methodology I adopt here is thus constructive and not descriptive. I aim at applying the general framework of pretence semantics to linguistic fictions, not at describing what are linguistic fictions from a theory-neutral perspective. In so doing, I hope to show that the general framework of pretence semantics is applicable and provides a way of clarifying intuitive ideas spread out in the philosophical literature. Eventually, I will show that one can find a common logical structure for fictions in general. This logical structure is quite complex and is reminiscent of narratological models available in the literature. This will provide a solution to the problem of fictional "truth" as well as a way of solving other linguistic and narratological problems which are well known in the literature. I will thus conclude that one should adopt this model for fiction, given the theoretical benefits.

First, I will present the two-fold structure which is, I shall argue, constitutive of fictions. The discussion will focus on the so-called "report principle". The term comes from Everett, although it originates in the work of Searle. I will argue that, in order to answer the problem of "reporting the unreported", one should construct this two-fold structure. Then I will explain how it can be put to work. Finally, I will show that the problem of "reporting the unreported" might be thought to re-appear, in a sophisticated form. But I will argue that one can block the problem, by appealing to the "appearance principle". This last move was suggested by Evans. It is a rather tentative move, as I will show.

7.1 In search for the fiction principle

7.1.1 Genesis of the report principle

[Searle 1975] is a seminal article on the logical status of fictional discourse. Searle made a point about fiction being *pretend assertion*. Given his own theory of *speech act* (assertion is a speech act) and *intentionality* (pretence is intentional), Searle can give a "logical status" to fictions in his broad theory of meaning. Introducing the speech act theory and his own theory of intentionality would drive us very far from the subject. But I think Searle's fundamental idea is simple and fruitful even taken outside its original theoretical framework.

The idea is that fictional discourse is parasitic on non-fictional discourse: it is conceptually second. What is conceptually primary is talking about the real world. That is, natural language *can* be used, and often *is*, to talk about the real world. Indeed, one can successfully tell you what I ate this morning using natural language.

Let this be called the "testimonial ability" of language.

Fiction, that is the use of natural language which is at the core of writing and reading stories, according to Searle, is the pretence of this ability. In fictional contexts, we are using the testimonial ability of language for other purposes than genuine testimony: we make *as if* we were relating facts about the world.[2] The author pretends to give a testimony and the audience pretends to hear a testimony. Fictional discourse is thus akin to lies: lies are just half way into fiction, for, in lies, the audience does think the liar produces a genuine testimony.[3]

As Borges once wittily put it in "Everything and Nothing", a short story which is a pretend testimony of Shakespeare's life, fiction is a "controlled hallucination". What Borges writes about the fate of the actor generalises to fictional discourse:

> [...] the actor, who on a stage plays at being another before a gathering of people who play at taking him for that other person.[4]

I think this idea is rather uncontroversial, although it leads to interesting problems when it comes to making it precise. Everett's formalism makes it the case that fictional discourse is parasitic on non-fictional discourse, for the language of pretence semantics is an extension of a base language containing no pretence operator. It remains to be seen how the testimonial ability of language should be rendered fruitfully.[5]

7.1.2 The report principle

This leads to the "report principle". Here is how Everett introduces the report principle:

> We will treat the fictional text or narrated story essentially as if it were purported factual report. This is in certain ways close to the "report

[2]Note that this idea grounds Lewis's expression "told as known fact" discussed above. Lewis does not cite Searle, though.

[3]I am ignoring the difficult case of "bald-faced lies" here. See [Sorensen 2007] for a seminal discussion of bald-face lies. See [Maier 2018] for a distinction between lies and fictional discourse which takes into account complications relative to bald-face lies.

[4][Borges 1964], p. 248.

[5]Note that Searle's idea can be applied to other media provided there is a "testimonial ability" in this non-linguistic medium. One can easily see how drawing comes with such a comparable ability, hence the fiction / non-fiction distinction applies there as it applies to language. Conversely, if a medium has no such ability, the distinction between fiction and non-fiction vanishes. For instance, music, under a naïve understanding of what music is, would not ground such a distinction.

> model" of our engagement with literary fiction, on which we pretend or imagine that the fictional text we are reading is a factual report.[6]

The report principle is thus a principle of generation which says, informally, that readers of fiction make as if the fictional text was a true report of fact. The report principle is a close cousin of Lewis's "told as known fact" expression, although it is not used to construct a counterfactual statement.

Let me quickly dismiss a false objection here. Unreliable narration does not contradict the report principle, quite the contrary. An unreliable narrator can be conceived of, in this context, as analogous to a lying witness: they tell a distorted version of the (fictional) events. But for the lie to work, there must be (i) something like the testimonial ability of language, (ii) a norm akin to Grice's maxim of quality, according to which one ought to tell the truth in one's testimony. In the tribunal, for instance, this norm is usually made explicit. Consequently lying in the tribunal is possible (and severely punished). Conversely, if lying is possible in everyday use of language, this means that language has a testimonial ability *and* that something like Grice's maxim is in force.

In fiction, the testimonial ability of language is pretended and the norm is precisely the report principle. So unreliable narration is possible (though no punishment will happen). Therefore, far from contradicting the report principle, unreliable narrators presuppose it.

The interesting, tricky point about unreliable narration as opposed to false testimony is to explain *how* readers discover that the narration is unreliable. In reality, one can, in principle, catch the lying witness red-handed by confronting their testimony with the facts (or other witnesses' testimonies). In fiction, there are no such facts independent from the telling of the facts (and very often, there are not several testimonies). In practice, the display of unreliable narration is often very subtly encoded within the *style* of narration, hence the literary interest.

Formalizing the report principle

When one tries to formalise the report principle into pretence semantics, one is faced with two distinct interpretations of it. As Everett points out:

> In fact the report model can be understood in two different ways, depending on whether or not we understand the texts themselves to be present within the scope of the imaginings they prompt.[7]

[6] [Everett 2013], p. 32.
[7] [Everett 2013], p. 32.

So either *it is fictional that* the text is a report of fact, or it is not so. The first interpretation is called the "strong interpretation", and the second the "weak interpretation".

Many novels seem to follow a principle of generation like the strong report principle, according to which the fiction requires the reader to imagine that the text they are reading is, fictionally, a true report of facts. For instance, *Lolita* by Nabokov, is such a novel. The text one reads is presented as the memoires of the fictional narrator Humbert. He writes them in prison. And the reader is explicitly required to imagine that they read *these* fictional memoires. There are numerous other examples. But the strong interpretation of the report principle is a special case, for there are many stories for which it does not apply.

It so happens that *the Adventures of Sherlock Holmes* became the paradigmatic example of fiction discussed by philosophers. Conan Doyle's stories happen to abide by the strong report principle. Therefore, many philosophers were led to believe that the strong report principle was constitutive of all fictions. But it comes from the fact that Conan Doyle's story are *not* paradigmatic at all. As Proudfoot forcefully puts it:

> [...] philosophers who treat the Sherlock Holmes stories as a *standard* case of fiction misidentify the respect in which they are run-of-the-mill: they are so in their banality, but the fact that the stories can be told as known fact is a highly unusual feature.[8]

Indeed, there are fictional text which do not abide by the strong report principle. *Ad hoc* examples can be easily created when one adds at the end of a story "everybody died and no one was left to tell the tale".[9] But there are many other counterexamples to the strong report principle which are not *ad hoc*. These are the fictions in which it cannot be the case that the fictional text is not part of the fictional world. My favourite examples is Roy Lewis' book entitled *What we did to father*, which tells the story of a family of Australopithecus who have not invented writing yet. Yet, the "testimony" is in a written form. So if the strong report principle was in force, then the witness could not have been a member of the family. This clearly goes against what is fictionally "true" in the novel.

This is the so-called "reporting the unreported" problem. The term comes from [Walton 2013] (§3). In this article, Walton argues that the problem also happens in other media. For instance, a fictional drawing of a dinosaur does not license to infer that it is fictionally the case that there is a human being drawing the dinosaur.

[8] [Proudfoot 2006], p. 34.

[9] This *ad hoc* counterexample to the strong report principle comes from [Lewis 1978].

Another example is *2001: A Space Odyssey* in which it is fictional that the spaceship Hal is in a region of space where there is no one. Yet, the spaceship is seen from the outside. But it is not fictional that there are people watching the spaceship in this region of space. Because of the reporting-the-unreported problem, the strong report principle does not generalise to all fictions.

So are we left with the weak report principle as a candidate for generalisation. The weak interpretation of the report principle is designed to answer the reporting-the-unreported problem. Here is how Everett describes the weak report principle:

> Understood in the second, weaker way, the report model holds that when we consume fiction we treat the text as if it were a source of factual information, although the text itself is not a denizen of my pretence. [Footnote: The difference between the strong and weak version of the report model can be seen as a difference between whether our imaginative engagement with the relevant text is governed by a type-I or a type-II principle of generation. I note that there may also be mixed cases where, during the course of our reading a text, we will take the text to be part of the world we imagine at certain times but not at others (perhaps *Frankenstein* is an example of this).] Rather we are to let the text guide our imaginings as we would let a real text guide our beliefs.[10]

This passage is quite obscure. I will try to elucidate it. Then I will argue against Everett's view and propose another.

7.1.3 Strong and Weak

Weak troubles

Everett asserts, in his footnote, that the difference between the strong and the weak interpretation can be cashed out in his formalism as a difference between a (T1) and a (T2) principle.

Here is, I suppose, what Everett suggests:

> (Strong): $[^f f$ is a report of fact$]^f$
>
> (Weak): For all sentence A, if A would be true under the supposition that f were a true report of fact, then $[^f A]^f$.[11]

[10] [Everett 2013], p. 33.

[11] Note that Weak is not strictly speaking of type (T2), but I cannot see how one should translate "being a true report of fact" into first order logic supplemented with Everett's operator.

It seems that (Strong) is formalised as should be. By contrast, I think that formalizing (Weak) in this way simply begs the question.

Indeed, what does "under the supposition that f is a true report of fact" mean? Remember that f is the real text, for Weak is supposed to be a principle of type (T2). But f is not in fact a true report of fact, so the supposition yields a counterfactual statement. But then, we are left with the same problem that was used as a direct argument against Lewis's counterfactual treatment of "being told as known fact". There is no way of supposing that f is a true report of fact without dramatically changing the semantic content of f. Weak was supposed to explain how the reader is to suppose this. But it presupposes what it is designed to explain.

To make this vicious circle plain, here a more brutal rendering of Weak into a proper (T2) principle:

$$\text{(Weak')}\colon\ (\mathit{Report}(f) \wedge f \to p) \to [^f p]^f$$

The intended meaning of (Weak') is the following: if the proposition p is implied by the fiction f and f is a true report of fact, then p is fictional. But this is flatly circular: that p is implied by f is precisely what should be *generated* by Weak'.

Taking off the vicious circle

I argue that we can solve the puzzle by introducing a second pretence to the first's rescue. The idea has been proposed in several places, especially in [Walton 2013]. I think the general framework of pretence semantics can make this proposal precise.

Let us call *primary* pretence the pretence in which the fictional events occur. Let us call *secondary* pretence the pretence according to which the fictional text is a factual report of the events occurring in the primary pretence. This terminology comes from [Vuillaume 1990].[12]

The secondary pretence is a pretence in which the fictional text is a factual report of information. There are two main characters which I will call the *narrator* and the *narratee*. In general, the main fictional events of the secondary pretence is that of the narrator *telling* a story to the narratee, be it oral or written. The way the story is told, i.e. with or without interruption, in what order, reliably or not, are fictional events of the secondary pretence. The personality of the narrator, their sex, age, sense of humour and biases are also fictional features of this secondary pretence. The

[12]There are different terminologies in the literature. Predelli nicely talks about the "fiction proper" and the "fictional periphery". [Walton 1990] introduces the *work world* and the "game world". [Walton 2013] uses "primary story world" and "secondary story world".

study of the only secondary pretence is sometimes the object of literary criticism.[13]

Then, there are two cases which correspond to both interpretations of the report principle. In one case, the narrator and narratee are fictional characters of the primary pretence. In this case, the secondary pretence should be thought of as a part of the primary pretence. Indeed, one is to imagine that the narrator and narratee are characters among the other characters of the primary fiction. The narrator thus came to know about the events and later reports them to the narratee.

In the second case, the narrator and narratee pretend to be fictional characters part of the primary pretence. In this case, it is not the case that they are fictional characters of the primary pretence but it is fictional (according to the primary pretence) that they are. So from the real world point of view, it is primary-fictional that the narrator came to know about the secondary-fictional events and then reports them to the narratee. In this case, it is not possible to conflate the two pretences into one as in the previous case.

This double structure thus naturally unifies both interpretations of the report principle. Such a double structure has been investigated by narratologists for a long time, since it has been claimed that it explains many cases of "meta-fiction" on which I will focus in the next section. It elegantly solves the reporting the unreported problem by making explicit a hidden layer in the appreciation of fiction. We will see shortly that something like the reporting the unreported problem comes back in this framework, though.

I claim that fictional pretences as opposed to non-fictional pretences have this distinctive double structure. Consequently, fictions (in the ordinary sense of the word) should be seen as complex games of make-believe in which there is at least one embedded pretence.[14] Contrary to Gregory and Eric who straightforwardly pretend that this stump is a bear, the readers and spectators of *Hamlet* first pretend that they (that is the narratee whose role they play) receive information from a source (that we call a narrator) and then pretend that the source is a causal source of information of the primary-fictional events. In other words, what Borges says of the actor generalises to readers. To paraphrase him taking the reader's point of view: the reader plays at being a person (the narratee) listening to a person (the narrator) who plays at being a reliable, causal source of information who produced the written text one is reading.

[13]For example, feminist literary criticism originates in the study of the general setting of the secondary pretence in nineteenth century English novels. It was shown how such pretences where almost always featuring a male narrator and a male narratee, with the accompanying sexual biases. See [Gilbert and Gubar 2000].

[14]Stories within stories show that there can be more than one embedded pretence in a fiction.

Here is a schema of this double-structured pretence:

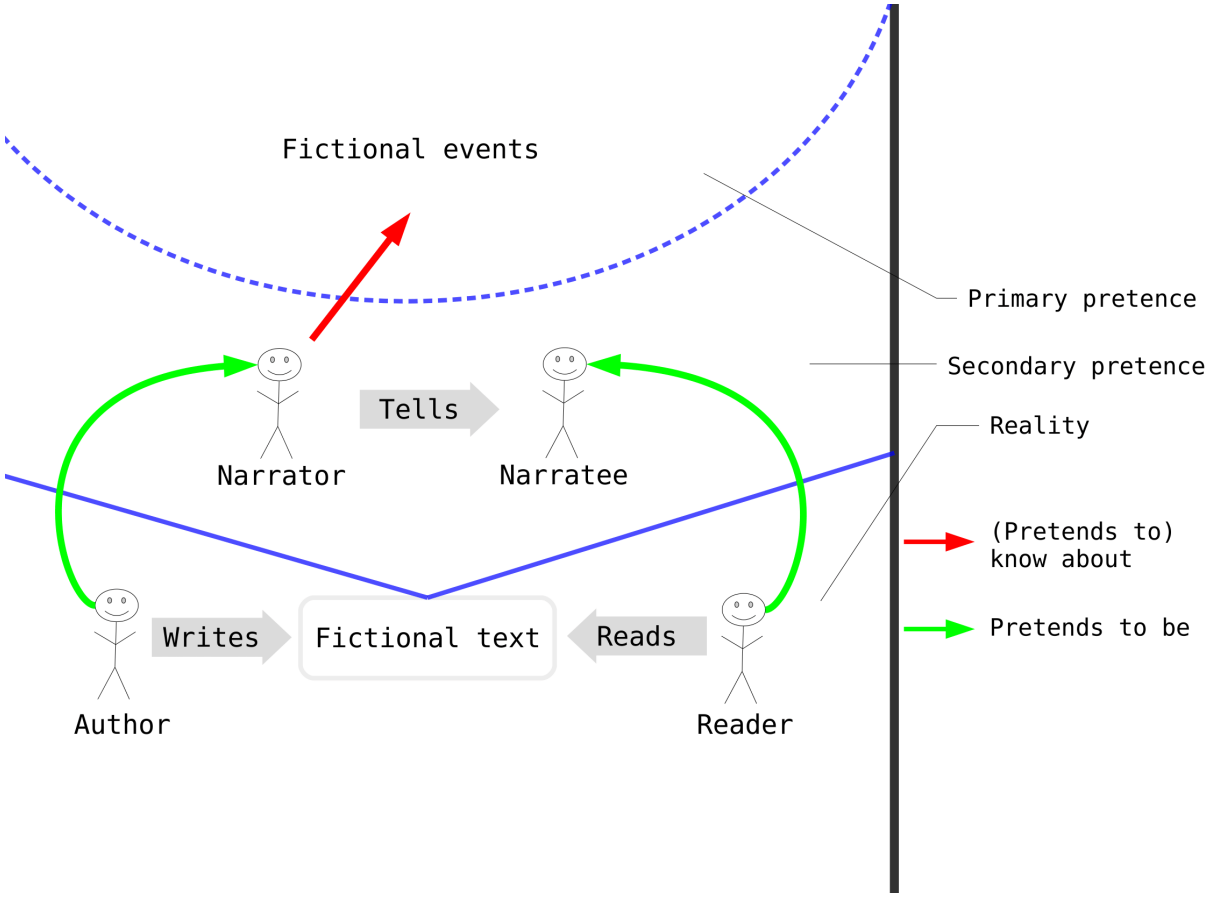


Figure 7.1: Double-structured pretence distinctive of fiction

The blue lines represent metaphysical borders. The metaphysical border between the primary and secondary pretence is a dashed line, because when the report principle is interpreted as STRONG requires, then the primary and secondary pretence collapse into one and the same pretence. In this case the narrator *knows* about the fictional events. Consequently, both the narrator and narratee are fictional characters. By contrast, when the report principle is interpreted as *Weak* requires, the primary and secondary pretences are distinct, and the narrator *pretends to* know about the fictional events.

This distinctive double structure has been investigated by theoreticians of litera-

ture. For instance in [Ryan 1991][15] and in [Eco 1994][16] one can find several schemas which reveal a similar two-fold structure. However, I can see two differences between my schema and theirs:

(i) it captures both STRONG and WEAK into one single schema, thus showing how the two interpretation of the report principle are structurally related;

(ii) the focus here is on pretence, hence on the fictionality-conditions of what the narrator tells to the narratee and not on what kind of "world" the fictional text "projects".[17]

Continuity and discontinuity between fictions and other games of make-believe

This structure, I contend, explains both the continuity and the fundamental difference between "ordinary" games of make-believe and fictions. Fictions are structurally more complex than other games of make-believe. They should be thought of as embedded games of make-believe. It is constitutive of linguistic fiction that they show a double-structured pretence. But it is also possible that they display even more sophisticated embeddings, with more levels of pretence. Games of make-believe, by contrast, are not necessarily double-structured. But it is possible to find such embeddings in non-fictional games of make-believe.

The continuity between the two kinds of activity is thus explained. Walton expresses some puzzlement concerning the (lost?) continuity between children's games of make-believe and fictions at the beginning of [Walton 1990]:

> Children devote enormous quantities of time and effort to make-believe activities. And this preoccupation seems to be nearly universal, not peculiar to any particular cultures or social groups. The urge to engage in make-believe and the needs such activities address would seem to be very fundamental ones. If they are, one would not expect children simply to outgrow them when they grow up; it would be surprising if make-believe disappeared without a trace at the onset of adulthood.
>
> It doesn't. It continues, I claim, in our interaction with representational works of art (which of course itself begins in childhood). The forms

[15] See her chapter 1 for discussion and p. 30 for her concluding schemas.
[16] See his 1st lecture entitled "Entering the woods" and especially his figures 1-5, pp. 19-23.
[17] The terminology is theirs.

> make-believe activities take do change significantly as we mature. They become more subtle, more sophisticated, less overt.[18]

I think we now have a clear idea of what "subtle, more sophisticated, less overt" means: it means double-structured in the above sense.

7.2 More on the secondary pretence

7.2.1 A toy example

Interestingly, children's literature often play with this double-structured pretence. Shel Silverstein is an expert in the subject of playing with the secondary pretence. Here are two dual examples which will serve as toy examples to understand what the secondary pretence is.

First, the poem "Slithery Dee". In this narrative poem, a character is trying to escape a sea monster called the "Slithery Dee". The first-person narrator is telling us how the monster chased "all the others". Toward the end of the poem, the narrator says that the Slithery Dee "ate all the others / But he didn't eat me". The punchline of the poem is the following:

> He ate all the others
> But he'll never eat — (abrupt stop and a strange noise)[19]

The effect works because it seems that the secondary fiction and the primary fiction are one and the same thing to a degree that if the narrator dies in the primary fiction, then the secondary fiction also ends. This is a highly unusual fictional setting, hence the comic effect. It thus wittily points out that a lone primary pretence (in the first person) cannot represent the death of its narrator.

Second, the story "A Giraffe and a Half". In this story, it seems that there is only a secondary pretence and no primary pretence. The story unfolds by stretching the antecedent of a conditional, starting with "If you had a giraffe ..." But no consequent is ever told. Indeed, the first half of the story unfolds many extraordinary things happening to the giraffe in this growing antecedent. But the second half of the story consists in undoing all these extraordinary things by subsequent events within the same antecedent. Consequently, one is left at the end of the story with nothing having happened in the story! There is just this hypothesised giraffe. The story thus

[18] [Walton 1990], p. 11-2.
[19] There are several versions of this poem which was later adapted into a song.

unfolds as a potential story only. One is stuck to the secondary pretence in which it is possible to tell primary-fictional things in ever so many ways.

These jocular fictions breed on the availability of both the primary and secondary pretences in fiction. As such, they are extreme cases which reveal the hidden complex structure of fictions.

7.2.2 A linguistic puzzle explained

Marcel Vuillaume has found and solved an interesting linguistic problem when it comes to fiction.[20] The problem concerns the tenses in fiction. It so happens that the past tenses co-occur with adverbs like "now" or "today" in the same sentence very frequently in fiction. In non-fictional contexts, by contrast, the presence of such an temporal adverb forces the present tense. In other words, having a VP in the past tense with an adverb like "now" or "today" in non-fictional contexts in either judged ungrammatical or borderline grammatical. In fiction, it happens all the time and we do not even notice it. Here is Vuillaume's favorite example, it is an excerpt from *the Red and the Black*, by Stendhal:

> Our hero was awkward enough to stand close to the little straw chair, which had once witnessed such brilliant victories. No one **said** a word to him, **today**; his presence was imperceptible – or worse.[21]

This puzzle is easily solved once we recognise that there are always *two* timelines available in the fiction and that these timelines are systematically correlated. There is the timeline of the primary and that of the secondary pretence.[22] In this example, "today" refers back to the timeline of the secondary pretence while the past tense in "said" refers to the primary pretence.

This idea also explains many other kinds of surprising collocations in temporal linguistics. For instance a sentence like the following is straightforwardly understandable:

(13) Anna Karanina met Vronsky at the train station where she will commit suicide towards the last chapters of the novel.

[20] In [Vuillaume 1990]. See also [Iatridou 2000] for a similar line of research investigating the relationship between tense, mood and counterfactual statements.

[21] *The Red and the Black*, chapter 20. Translation by Burton Raffel, ed. the Modern Library New York.

[22] For the record, Vuillaume tries and formalise this double structure using the so-called "mental space account" as developed in [Fauconnier 1994]. He also attempts to generalise this structure to non-fictional reports of fact.

Vuillaume's inquiry is based on many examples from nineteenth century French novels that led him to investigate the secondary pretences of these books. He shows how many novelists play with their secondary pretences, inviting the reader to ironically imagine what the secondary pretence is like by having their narrator explicitly addressing the narratee. The narratee is thus invited to follow the narrator, to closely look at the fictional character, to leave the room, and so on. And sometimes the narrator wittily remarks that the narratee is an invisible seer of the fictional events. This contributes to a shared pretence (among several novels) that the narrator and narratee are like ghosts in the primary pretence. They can see and comment on (especially the narrator, of course) the primary-fictional events without being seen or heard by the primary-fictional characters. They can also travel in the primary-fictional space and time without difficulty. Interestingly, the effect of such a "betrayal" of the double structure of the fiction does not destroy anything. Quite the contrary, it has been claimed to enhance the overall "realism" of the fictions in which they occur.[23]

7.2.3 Two narratological puzzles explained

Some works of fiction play with the double structure of fictional pretences. They are sometimes called "metafictions". They usually challenge narratological explanation. Although there are multifarious examples, I will explain two. This will show how the fictional principle can be applied.

In *In on a winter's night a traveler*, Italo Calvino plays with the double structure of fictional pretences by shifting the attention to the secondary pretence in a very original way. As seen above, it is very common for the secondary pretence to be assimilated as a part of the primary pretence. This yields the strong interpretation of the report principles. Calvino ingeniously tries to assimilate the secondary pretence with reality.

Indeed, the secondary pretence, as seen above, is intermediate between the reality in which the actual reader and author dwell and the primary-fiction where the primary-fictional characters dwell. The narrator and narratee, the secondary-fictional characters can easily become primary-fictional characters. After all, they are fictional characters, it suffices that both fictions be merged in the imagination of the reader.

Calvino proposes that we merge reality and the secondary fiction. He does so by using the second-person pronoun to address the reader. He describes the narratee

[23] I say "realism" with quotation marks because it is precisely these *unreal* devices that ground the "realistic" effect. The realistic effect was studied along these line, for instance, in [Barthes 1968].

thus addressed as the actual reader of the book. He speculates in different places about the detail of the actual reader: is it a "he" or a "she"? In the meantime, the primary pretence is comically malfunctioning: the reader (that is *you*) is systematically coming across unfinished stories due to some mischievous misprints. As such the secondary fiction becomes a search for the primary fiction, the actual reader is looking for the book entitled *If in a winter's night a traveller*, that is the one he is actually reading. One now gets the meaning of the title which stops half way through, so to speak.

Of course, this merging of the secondary fiction and the real situation is impossible. By definition, the fiction is not reality. And Calvino is just cleverly using the communicational situation to generate some key fictional propositions which are also true propositions: "You are reading Italo Calvino's new book *If on a winter's night a traveler*". Of course, nobody is tricked into believing that the story is about oneself and it is part of the fun. The cornerstone of Calvino's idea is that the pronoun "you", in the context of the secondary fiction, refers to the narratee *and*, given that the actual reader plays the role of the narratee, also homonymously to the actual reader.[24] This is because the reader "is" the narratee, i.e. whoever is the reader "count as", or fictionally is the narratee.

This interpretation of Calvino's novel, naturally, is open for narratological disagreement depending on the detail of the text. But I think this is basically what happens. One can appreciate how the pretence semantic framework is applicable and useful here.

Another nice application of the framework concerns a well-known narratological problem in *Madame Bovary* by Flaubert. The novel famously opens thus:

> We were at prep when the Headmaster came in, followed by a "new boy" not wearing school uniform, and by a school servant carrying a large desk.[25]

The pronoun "we" used here forces the strong interpretation of the report principle: the narrator is a character in the story, and its report is thus part of the primary fiction. Given the context, we naturally infer that the narrator is one of Charles Bovary's classmates back then. He thus knows about Charles, who is going to be one of the main characters in the story.

But this first-person narrator gradually fades out so that after a few pages the narrator of the story is clearly a third-person omniscient narrator. As such, it suggests a weak interpretation of the report principle, with a distinct, discreet secondary fiction.

[24] Had we two pronouns "you" and "zou" to define both, the whole joke falls down.
[25] Translation Margaret Mauldon, ed. OUP 2004.

The secondary fiction is distinct, for the narrator is omniscient and has information about primary fictional events which had no witness whatsoever, like the numerous moments of Emma's solitary thinking or her moments of adultery love making which, given the fictional detail, could not have been known by a hypothetical classmate of Charles.

I suggest that Flaubert is here "zooming out" of the primary pretence, so to speak. The subliminal change, created by Flaubert's mastery, has a number of literary effects, the first of which is the complete oblivion of the secondary pretence and the entire focus of attention on the primary pretence. In other words, the reader will not even raise the "how do you know?"-question which would inevitably distract him from the primary-fictional events. Flaubert thus has the narratological cake and eats it: he gets the narratological freedom a secondary pretence permits without raising the reader's suspicion or disinterest.

As such, Flaubert's trick consists in rendering the secondary pretence seamless, invisible, i.e. phenomenologically intractable. It is thus the dual of Calvino's narratological trick which consisted in rendering the secondary pretence full fledged. Notice that this interpretation squares with Flaubert's stated aesthetic standards when he famously writes that he wrote that the author in its work was to be like God in His creation: "everywhere present but nowhere to be found."

Again, this narratological interpretation of *Madame Bovary* is open for rational disagreement. But I take it that it explains what could really be mysterious if one did not allow for the double structure presented above.

7.3 The revenge of the reporting-the-unreported problem

7.3.1 Virginia Woolf and the report principle

In *Jacob's Room*, Virginia Woolf explicitly constructs a narrator who is reporting the events in the primary-fiction. But the narrator is not to be trusted as a primary-fictional character, for the "reporter" knows much too much about the fictional characters and their inner thoughts. Here is a passage which illustrate this going too far. The context is the following: Betty Flanders is Jacob's mother. Jacob left his hometown, Scarborough, for a tour of Europe in the early 1910's. Jacob and Betty exchange letters regularly, although Jacob responds less and less to his mother's letters. In commenting on one of Betty's letters left unanswered, the narrator says:

> Meanwhile, poor Betty Flanders's letter, having caught the second

> post, lay on the table – poor Betty Flanders writing her son's name, Jacob Alan Flanders, Esq., as mothers do, and the ink pale, profuse, suggesting how mothers down at Scarborough scribble over the fire with their feet on the fender, when tea's cleared away, and can never, never say, whatever it may be – probably this – Don't go with bad women, do be a good boy; wear your thick shirt; and come back, come back, come back to me.

The narrator explicitly says that Betty "cannot say" what is in her mind. Though the narrator crosses the line after "probably this" by saying what is on Betty's mind unbeknownst to her.

Consequently, the reader accommodates to place the narrator in a secondary pretence, where they can enjoy the prerogative of omniscience, where it is possible to search the heart and examine the mind without going into narratological trouble. What is interesting about this narrator is that they *present themselves* as having no such prerogative. It is as if they had relapses of omniscience again and again. This makes the secondary pretence very interesting from a psychological point of view and creates situations like the one above where the irony is everywhere.[26]

So far, so good. But a more perverse situation emerges when we understand that the narrator is trying to make a point (or is it Woolf's?). When they try and justify their relapses, they come up with a thesis which threatens the very possibility of a true report or a faithful testimony. Here is, to illustrate this, a sentence which appears several times and becomes like a motto:

> It's no use trying to sum people up. One must follow hints, not entirely what is said, nor yet entirely what is done.

This blurs the important conceptual distinction between the telling of facts and the facts themselves. Any workable notion of report presupposes that one can distinguish between the two. Here, when it comes to *people*, the narrator holds that such a distinction does not hold. Truly reporting on people's action is hopeless, so to speak; the best we can do is "to follow hints", that is, in concrete terms, a mixture of facts and telling of the facts.

The fact that no true report can be done in the real world is a controversial thesis, to say the least.[27] But the narrator rather makes it secondary-fictional that there

[26]Note that Virginia Woolf was translating and giving the first English edition of Freud around the same time she was writing *Jacob's Room*.

[27]Although some philosophers have defended it. The so called "school of suspicion" to borrow Ricœur's expression designating the philosophical works of Marx, Freud and Nietzsche.

can be no true report of facts, which should not be controversial. One can see how this breaks the whole machinery down...

7.3.2 Woolf-like counterexamples

Woolf's *Jacob's Room* is not a clear counterexample to the report principle as interpreted involving a double-structured pretence, but it is now easy to construct pure cases.

Imagine a story in which the secondary pretence is such that the language used has no testimonial ability. In other words, it would be secondary-fictional that there can be no report using language. Consequently, there can be no pretence of true report. Using Lewis's terminology to make things even clearer: in the secondary pretence there is no knowledge, so there can be no telling of things as if there was known fact. Such radical fictions would arguably be utterly difficult to imagine in practice, but they are conceptually possible. I call them "Woolf-like counterexamples" to the double-structured analysis of fictions.

I think this revenge of the reporting-the-unreported at the level of the secondary pretence has never been discussed in the philosophical literature. Although I think it was somewhat anticipated by Byrne 1993 when he talks about the "supernatural epistemological power" of one of Iris Murdoch's narrator in *The Book and the Brotherhood*. But he makes it a problem at the level of the primary pretence. And "this objection misses the mark" as [Hanley 2004] remarks, because of the double structure underlying storytelling that I made explicit in Everett's framework. Same for Everett who gets something like the revenge problem when he talks about what we think about the *authors* and not the fictional events in the following:

> If we really did read Joyce or Woolf as purported factual report we would reject them as frauds on the grounds that no one can have the sort of access to others' streams of consciousness that they seem to have. If we read Calvino or Burroughs or Kafka as purported factual reports we should take the authors of those reports to be simply insane.[28]

I think he points out to what I was talking about about *Jacob's Room*. But the rest of the paragraph makes it quite clear that he refers to the reporting the unreported problem for the primary fiction only. Although such talk about the actual authors by name make a weird signal, especially when he designates the *narrators* of these stories as "authors of those reports" in such a context. As for [Walton 2013] who

[28] [Everett 2013] p. 33-4.

sees and responds to the reporting the unreported problem in a very similar fashion as is done here, he simply fails to remark the problem's revenge.

I think this revenge problem is real and challenging for the view I put forward. It is of little comfort that I did found real instances of such radical counterexamples in the literature, for if it is not already in the loose, it will be. As far as I can see, one could try several strategies to circumvent the Woolf-like counterexamples.

One can try and argue that, in fact, this class of counterexamples actually *presuppose* the double-structured report principle. One would have to exhibit the typical narratological structure of such stories, in order to back up this claim. This strategy is analogous to the treatment of unreliable narration which is a pseudo-counterexample to the report principle in its simple version, although it should be done at the level of the secondary pretence.

One can try and argue that these counterexamples are really *impossible* stories, or contradictory fictions which are to be analysed along side with the other kinds of impossible stories. They would typically require another level of pretence in which a meta-narrator and meta-narratee tell each other how the narrator and the narratee have trouble telling the primary fiction. Of course, one would need to have an argument concluding that no revenge problem can appear at this level.

One can bite the bullet and conclude that the double-structured report principle is not the general principle of generation of fictions. Some fictions would fall short of this principle, though they are few.

I will *very tentatively* take the second route in response to Woolf-like counterexamples to the double-structured report principle. The needed argument that no revenge of the reporting the unreported problem can happen will rest on Evans's "appearance principle".

7.3.3 The appearance principle

The strategy is to build in a principle of generation into the picture so that Woolf-like counterexamples could not arise. The best they can achieve is to yield contradictory pretences, which can be analysed away. This general principle can be called the *appearance principle* and be put along with or in place of the reality and the mutual belief principles.

Here is a situation Evans considers:

> Two men, let us suppose, both seem to see a little green man on the wall, and are persuaded, with reason, that there is no such thing; they are, they think, victims of a trick, perhaps involving the use of holograms. There will be two different ways of telling the story; one in

> which they are right in thinking this, and one in which they are wrong. For the time being, let us suppose they are right. This shared perceptual information provides an ideal backdrop for an existentially creative game of make-believe. The basic stipulation – that things are as they seem – immediately generates a whole mass of make-believe truths. (e.g. that [^{f}the little man on the wall has a beard]^f, etc.).[29]

The appearance principle Evans's proposes is thus:

> [^{f}Things are as they seem]^f

He illustrates how this principle works with perception, which is pretty straightforward, but how is this principle to be applied in the case of linguistic fictions?

What we need to make sense of the appearance principle in the case of fiction is a distinction between illusory and non-illusory linguistic content. This, Evans claims, is given by the causal theory of semantic information he advocates in his book. Let us focus on referring expressions once again, since they define a fundamental subset of the semantic information on which things are easier to define. Definite descriptions can give rise to semantic illusions, as opposed to proper nouns. Definite descriptions, indeed, descriptively refer. That is, the referent, if there is one, is picked up by description, i.e. it is whoever uniquely *satisfy* the description. By contrast, proper nouns are directly referential. That is, the referent, if there is one, is picked up directly, i.e. there must be a relevant causal chain which links up the utterance of the name and the individual named, irrespective of what the speaker might think true or false of the referent.

There is a clear analogy with perception. One can see directly some individual or see them through a representation of them, like a drawing or a hologram. In the first case, the perceptual information is gathered causally, whereas in the second case the link to the individual is causal up to the representation and intentional afterward.

For instance, suppose our two protagonists now suffer a linguistic illusion. They would both interpret the sentence: "The average English man drinks tea in the morning". They would, by definition, understand they are suffering a semantic illusion and none of them would believe that there is a man, who is both English and average, and that *this man* drinks tea every morning. They need not know exactly what the trick is, just as in the case of the hologram. But the shared semantic illusion, if accepted, also "provides an ideal backdrop for an existentially creative" pretence. Given what

[29] [Evans 1982], p. 360. For the record, Evans symbolises the fictional operator using * * instead of brackets. It corresponds precisely to Everett's operator, so no harm is done to the original text here.

the words mean, especially the word "average", one can generate a "whole mass" of fictional propositions when applying the appearance principle. Of course, all this would be up for disagreement, but the person is clearly a man, probably white, something like 40 (I guess), voting for Brexit (but that already is open for disagreement I suppose), etc. The point is that what the word "average" provides an "informational background" in the required sense.

This principle of appearances blocks Woolf-like counterexamples. Indeed, the testimonial ability of language is encoded in the meaning of certain words. So if $[^f$things are as they seem$]^f$, then a secondary fiction in which the language used makes it impossible that reports of fact can be produced is simply impossible.

Is there a revenge reporting the unreported problem? I do not think so. Try and imagine a world in which there was no distinction whatsoever between appearance and being. It is either a world in which there were no appearances and only pure beings. Or a world in which there were no beings and only appearances. The first case sounds like Plato's realm of ideal Forms. But this world could not be *told*. That is the whole point of Plato's doctrine. The second case sounds like Berkeley's metaphysical picture in which *esse est percipi*, without the second part *aut perpicere*. Again, I think such a world could not be *told* or represented in any way. This is precisely why Berkeley needs *both* the things perceived *and* the things that perceive as primary entities in his ontology.

Therefore, I think it is possible to block Woolf-like counterexamples and thus meet the revenge of the reporting-the unreported problem. Instead of a report principle, one should adopt the more abstract appearance principle. It is not that the fictional text is a report of fact in the weak sense. It is that, when engaging in a fiction, it is fictional that things are as they seem. Since, without indications to the contrary, the testimonial ability of language seems to operate when assertions are uttered, it follows that the report principle is a special case of the appearance principle. This idea suggests a now familiar picture of language in keeping with a causal theory of semantic information. Evans, once again, was an interesting lead to follow on this.

7.4 Conclusions

The aim of this chapter was to find (the set of) fiction principles, i.e. the principles of generation which are constitutive of fictions as opposed to other kinds of pretences.

The starting point of this inquiry was Searle's seminal idea that fictional discourse is pretend assertion. It follows from this idea that what I called the "testimonial ability" of language is duplicated or simulated within the scope of fiction. This led

to the formulation of a report principle which says, informally, that the fictional text should be thought of as a pretend report of facts.

In trying to formalise this intuition, one is forced to distinguish between two interpretations of the report principle depending on the fictional status of the fictional text. If it is fictional that the text is a report of fact then one interprets the report principle in the strong sense, if not in a weaker sense. In light of the so-called reporting-the-unreported problem, the only candidate for generalisation was found to be the weak report principle.

I then discussed Everett's proposal to distinguish between the two interpretation of the report principle using principles of type (T1) and (T2). I showed that this was at best misleading. (T2) principles are principles which govern the function of props and fictional text are indeed props. But I showed that in order to avoid the reporting-the-unreported principle, one has to distinguish between two levels of pretence. I called these two intertwined pretences the "primary" and "secondary" pretences, following Vuillaume. I proposed that this two-fold structure should be constitutive of all fictions. And I illustrated the interest of this logical structure by discussing a linguistic problem relative to the use of linguistic tense in fictional discourse and well known narratological problems. I suggested that this double-structure was already familiar to narratologists. The main improvement on these already existing works is that pretence semantics avoids the problematic notion of "world" altogether.

Finally, I suggested a revenge problem of reporting-the-unreported which appears at the level of the secondary pretence. This problem, to my knowledge, has received little to no attention in the philosophical literature. I suggested that a tentative way of meeting the problem would be to shift from the weak report principle to a more general appearance principle which originates in the work of Evans. I tried to give a precise formulation of it and showed that this principle adequately prevents the revenge of the reporting-the-unreported problem.

This concludes the solution to the problem of fictional "truth". For linguistic fictions, because they exhibit this double-structure, are formalised into pretence semantics in such a way that inferences like the one concluding to the fictionality of (1) and not of (2) are completely available. In part 2, I will show how this framework can be fine-tuned so as to get more substantial fictional propositions out of fictional texts. I will show this by focusing on the practice of literary criticism and its underlying rationality.

Chapter 8

Readers as fictional truth-makers?

To end this chapter on fictional "truth", I would like to briefly consider the relationship between fictional propositions and readers. I suggested at the beginning of the previous chapter that, in fiction, the division of labour between creating and enjoying a fiction was complete. This suggests a view following which the author is active and the reader is passive. I want to substantially qualify this claim in this chapter by insisting on the activity of the reader when they read. This is a side issue which does not, strictly speaking, concern the problem of fictional "truth". However, I will show that it nicely connects with it in the pretence semantics framework.

Reader-response theory, originating in the work of Louise Rosenblatt[1], has many times been claimed to undermine an "objective" theory of fictional "truth" like that of the modal or the functional theorist. The motto is: readers are the real fictional truth-makers. I am not sure that such debates are well posed, though, as should be clear from the above discussion.

However, one interesting intuition coming from reader-response theory is that readers *do* cooperate in order to construct the fictional propositions. I think pretence semantics is precisely the framework in which such intuition is rendered precise and effective. All pretences are *games* which are empty shells if not played. This by no means implies that one can play the game without any formal constraint and that any proposition can be made fictional. But it is true that some practices of readers consists precisely in "playing" with what counts as a fictional "truth". I think these practices raise numerous interesting questions which could in principle be formalised and rigorously discussed in the pretence semantics formalism. I will thus mention some as doorway for future research.

[1]See for instance [Rosenblatt 1938]

8.1 Fictions without authors

Principles of generation come with an air of authoritarianism. Someone has got to be in charge! This was traditionally the prerogative of the author. But the functionalist framework does not, in principle, require that there is an actual author, setting up the rules and listing the principles of generation. Indeed, most of the principles of generation that we discussed are "built in" the framework and thus can be in force by the mere fact that one engages into a pretence. One can thus argue for the possibility of there being fictions without any actual author.

There is a debate over the possibility of conceiving of a fiction which has no author.[2] A fiction without author can be either a "natural fiction" or a "automatic fiction". Natural fictions are natural events or objects on which humans "read" stories. The paradigm case for Walton is to see pictorial representations in passing clouds: one can discuss and coordinate imaginings while watching the passing clouds; it is a very common game of make-belief. These are short-lived personal fictions, but some similar fictions have been steadily institutionalised. It is very clear that constellations of stars in a clear night should evoke a whole range of pictorial representations of mythical creatures. Here is another interesting example: in Savoy, France, one mountain represents very famously a pregnant woman in profile; she naturally has the traditional headdress of Savoy women.[3]

These are pictorial fictions, what about linguistic fictions? Here, examples are a little far-fetched but they exist. Suppose ants are walking on a beach. The traces they left on the sand are complicated. Suppose, that these traces end up forming the following sentence: "In a village, the name of which I have no desire to call to mind, there lived not long since one of those gentlemen that keep a lance in the lance-rack, an old buckler, a lean hack, and a greyhound for coursing." This would surely count as an author-less fiction prompted by a "natural" prop.

On the other hand, automatic fictions are fictions produced by a machine. It is easy to conceive of software which would randomly create sentences. After all, word predictors on smart phones produce texts which are comparable to *cadavres exquis*. These can surely enjoy the status of fiction. Another real life example is Queneau's *Cent mille milliards de poèmes* which is a set of ten sonnets each of whose verses can be combined with each of the others. Queneau thus produced 10^{14} well-formed sonnets. But is he the author of each and everyone of them? Answering "yes" would be to commit oneself to the idea that one can be the author of a text one has never

[2] [Currie 1990], [Walton 1990].

[3] See for instance: https://www.montagne-cool.com/donnees/photos/MOR_26786_1375978147.jpg_mini.jpg

entirely read, which sounds very odd indeed.

Some theorists are not happy with the idea of author-less fictions, but there is no particular conceptual problems with them. So I see no reason for turning them down.

The double-structured report principle works just as fine with author-less fictions. Indeed, it suffices to get rid of the author in Figure 7.1. I think this shows an interesting feature of the above analysis of fictional discourse. It is usually thought that the narrator is a projection of the author, a fictional *Doppelganger* of the real author. But actually, what these author-less fictions show is that the narrator is a fictional entity which is presupposed by the report principle without any connection to a real author. In other words, the narrator originates in the text, not in the author.

8.2 When readers are authors

There are numerous examples of well-known fictions which are authored by several people, from the *Iliad* to most contemporary TV series. There are also many examples of participative fictions, from choose-your-own-adventure books to improvised storytelling on the audience proposals. I think one can even go a step further in considering fiction-writing practices in which the readers are explicitly expected to be creating truths in fiction by writing them out.

One such project comes from the group of Italian authors Wu Ming who published a novel entitled *Manituana* first published in Italian in 2007. The book was accompanied with a website where contributions to writing missing chapters, prologues or making some pictorial or musical contributions were possible.[4] What is true in this fiction ultimately depends on both the authors and several readers' contributions. This was planned ahead, as is shown by this excerpt of an interview of Wu Ming:

> When we wrote *Manituana*, we built a web site in which anyone could contribute to the narration, but not only in written form: drawings, music, videos, any contribution that would multiply and develop the fictional world we opened with the novel. We believe that the novels we write do not define the boundaries of a fictional world, but they open the door on this universe. We explore a part of it, but readers keep exploring some other parts together. It is a never-ending process.[5]

[4] www.manituana.com

[5] Collectif Mauvaise Troupe 2014 *Constellations. Trajectoires révolutionnaires du jeune 21e siècle*, Constellation "La folle du logis. Imaginaires, récits, fictions" (https://mauvaisetroupe.org/

I think this practice is interesting in that it intentionally blurs the distinction between authors and readers, prompting readers to become authors.

8.3 A legal prequel for future fan-fictions

Another practice I want to discuss comes from fan-fictions. Fan fictions are fictions which take up on a popular work of fiction, and they are written by fans rather than their original authors. There is already a large literature on the phenomenon, although nothing really systematic yet.[6]

For instance, fans of *Star wars* have produced numerous stories taking place aside the Skywalker saga in the same fictional universe. They can be stories which do not interfere in anyway with the canonical stories. Or they can be stories in which the fictional characters of the fan-fictions cross path with the characters of the canonical fictions.

So far, nothing really interesting happens conceptually. Fan-fictions are derivative or dependent on their original fictions. This can be modeled very easily by picking up on an existent pretence and modifying the prop by adding some new sentences for instance. Authors writing series of adventures have been doing that for a long time.

Within fan-fiction practices, one literary device is interesting in the debate about fictional "truth", though. *Retroactive continuity* (or Retcon) is a literary device in which established facts in a fictional work are adjusted, ignored, or contradicted by a subsequently work which breaks the continuity with the former. One usually writes a retcon to connect several fan-fictions together for instance or fan-fictions with the canonical fictions. Retcons are also widely used to correct some errors or identified problems with a fiction after its publication, because, for instance, scientific discoveries are challenging already published anticipation works of fiction. Sometimes, they are also written so that the received interpretation of a fiction is changed, for

`spip.php?article96`). My translation. Here is the French version of the interview:

> Quand nous avons écrit *Manituana*, nous avons créé un site internet où n'importe qui pouvait apporter sa contribution narrative, mais pas seulement écrite: bandes dessinées, musique, vidéo, n'importe quel type de contribution qui multiplie, amplifie l'univers narratif que nous avions ouvert avec le roman. Nous croyons que nos romans ne définissent pas les frontières d'un univers narratif, mais ouvrent une porte sur cet univers. Nous en explorons un bout, mais les lecteurs continuent d'explorer d'autres bouts ensemble. C'est un processus qui peut continuer.

[6] See [Uckelman 2019] for a review of the literature.

instance be revealing the true (that is fictional) identity of a famous character. In those cases, retcons do contribute to establish the fictional "truths" as it were. Of course, as most fan-fictions, retcons are not produced by the authors.

Jean Rhys wrote such a novel in 1966 entitled *Wide Sargasso Sea*, which is a retcon in the form of a prequel of Charlotte Brontë's *Jane Eyre*. Jean Rhys makes it fictional in *Jane Eyre*, so to speak, that Bertha Antoinetta Mason, the madwoman in the attic, Rochester's first wife, became insane because of some voodoo spell which went astray. All this happened in Jamaica in the first years of their marriage, and Jean Rhys's novel well explains why it did not affect Rochester. Of course, this is crucially important to *Jane Eyre* because it undermines Rochester's telling of the story and explains his latent brutality.

Retcons naturally prompt many debates in the literary and fan communities alike. Not all such modifications are accepted. Dissident versions of what is the fictional "truth" appear and fight over the internet. That is fair enough. But the puzzling fact is that some fictional "truths" are thus added to the canon by the majority voice. One could marvel at this new democracy of readers against the old tyranny of the author.

The puzzle is elsewhere. It comes from US law. Indeed, when it comes to fan-fictions of mass productions, the copyright law aims at being strictly applied, since a lot of money is at stake when it comes to commercial derivative products. In these cases, companies hold the rights to sell and distribute not only the fictions but also all the toys and games associated with it. It is thus crucial to know *what* they (and only they) have the right to sell. This depends on what is "true in the fiction". So the law, in some cases, gets to decide what is fictional. Interestingly, section 107 of the Copyright Act defines a so called "fair use" of some copyrighted material, including some fictions. In the US law self-laudatory voice:

> Fair use is a legal doctrine that promotes freedom of expression by permitting the unlicensed use of copyright-protected works in certain circumstances.[7]

The fair use doctrine is the section that, for instance, permits that scholars use some copyrighted material for research or educational purposes.[8] When it comes to fan-fictions, and retcons, things tend to get tricky. But in principle, some fan-fictions can be made *legally* part of the canon, so to speak.

Of course, the trouble with this is that the court is only getting to arbitrate when there is a conflict. That is why one of the owners of a given copyrighted fiction

[7] https://www.copyright.gov/fair-use/more-info.html.
[8] Should we thank the law?

gets to sue a fan for producing such and such fan-fiction. This is very unlikely to happen in normal circumstances, because the amount of fan-fictions is usually seen as *advertisement* for the copyrighted canon. So nobody loses either fame or money in this kind of activity.

Here is a very relevant case from 1983, though. The key facts are the following:

> Plaintiffs, Warner Bros., Inc. and D.C. Comics, Inc., owned copyrights for the character "Superman" and works embodying Superman, including comic books, animated and live-action television series, and a motion picture. Defendant, ABC, Inc., created a television series called *The Greatest American Hero* which starred the fictional superhero "Ralph Hinkley." Plaintiffs alleged that defendant's Ralph Hinkley character and related promotional campaign materials infringed its Superman copyrights. Plaintiffs appealed the district court's ruling that public confusion as to the origin of defendant's superhero was unlikely as a matter of law.

The holding is the following:

> The court found that the characters Superman and Ralph Hinkley were not so substantially similar that consideration by a jury was required. Although the fair use defense was not raised or at issue, the court briefly addressed "fragmented literal similarity," which was the duplication of exact, or nearly exact, wording of a fragment of the protected work. The court noted that with respect to such claims, the fair use defense can be raised. The court pointed out that it can be expected that phrases and other fragments of expression contained in successful works will become part of the national language, but that does not mean that such expression loses all protection. On the other hand, although the author of a well-known copyrighted phrase "is entitled to guard against its appropriation to promote the sale of commercial products," original works of authorship with elements of parody have a stronger fair use defense against unauthorised use. The court emphasised that "[i]t is decidedly in the interests of creativity, not piracy, to permit authors to take well-known phrases and fragments from copyrighted works and add their own contributions or commentary or humor. After all, any work of sufficient notoriety to be the object of parody has already secured for its proprietor considerable financial benefit." This discussion did not have any bearing on the outcome of the case, but provides insight on the role of parody in a fair use analysis.[9]

[9] https://www.copyright.gov/fair-use/summaries/warnerbros-abc-2dcir1983.pdf.

Let us be warned.

Of course, one can question the rationality of such a "fair use" analysis. On the other hand, if this is a thing and if it is true that people disagree about fiction so as to take lawyers to prove the other wrong, then one is desperately in need of a rational framework to conduct one's arguments. This, I think, I provided in general here. In the next chapter, I will discuss less dramatic disagreements although equally conceptually challenging. I will try to provide a fine-tuning of the functional analysis.

8.4 Open-ending

These considerations show that the idea that fictional propositions originate in the reader's activity is perfectly compatible with the pretence semantics framework and the analysis of fictional discourse I have been proposing in this framework. In a sense, it is tailor made for it.

Fan-fiction practices and especially retcons show an interesting interaction between what is fictionally "true" and how the community of readers can contribute to it. I do not mean to say that it threatens the framework, quite the opposite. The sociology and dynamics of what is recognised as fictional in a community is a very interesting phenomenon which has become a theoretical object of study in the last decade, the result of which are sometimes difficult to grasp. My suggestion is that one can ground this object of study in the underlying principles of generation in force within fictions. Negotiations about which principles are precisely in force in which fiction reflect fan creative discussions and contributions as well as potential forks of canonical stories.

As an application of this negotiation phenomenon, I suggested one can get a closer look at copyright issues and the so-called "fair use" of copyrighted material. Indeed, what is copyrighted is sometimes what is fictional. In potential conflicts of interest, it is thus important to have an objective framework in which one can discuss about what is fictional and what is not fictional in a given fiction. Lawyers and judge might use intuition in empirical situations. I suggest they should rather try to formalise their argument within pretence semantics.

The next part will focus on a similar case where two readers disagree about what is fictionally "true". This part should be seen as an case-study in which I will have the opportunity to fine-tune the model I proposed in this first part. However, the second part is also largely self contained so that I will avoid as much as possible making reference to the points discussed here. The connections should be clear, though.

Appendix A

"Truth in fiction" as an unsupervised learning task

I want to defend the idea that the problem of truth in fiction, as defined in chapter 1, is conceptually identical with the problem of unsupervised learning. I conclude from this analogy that both research communities would gain in getting interested in one another, despite the *de facto* compartmentalisation of as distinct disciplines as computer sciences and philosophy. The following remarks are inspired by communications from Yann LeCun's course on deep learning at collège de France in 2016[1] and Stephane Mallat's 2018 one "L'apprentissage profond face à la malédiction de la grande dimension".[2]

The problem of "truth in fiction" in the abstract, as observed in chapter 1, consists in sorting out the fictional propositions from the non-fictional ones given the set of explicit fictional sentences. Humans are very good at doing this, from a very young age. And it seems that one can provide general heuristics for this sorting problem. The natural question a computer scientist would raise is the following: can one train a machine (say a neural network) so as to recognise fictional propositions given a fictional text? Thinking about the problem of fictional "truth" from this viewpoint is, I suggest, very much the right thing to do. However, I think, the problem of fictional "truth" is a difficult problem, as least as difficult as unsupervised learning tasks which are still open problems in computer sciences today.

[1] See https://www.college-de-france.fr/site/yann-lecun/course-2015-2016.htm.

[2] See https://www.college-de-france.fr/site/stephane-mallat/course-2017-2018.htm.

Supervised learning

Machine learning is a term denoting the field of research aiming at studying and developing algorithms and software which automatically and dynamically organise data. These algorithms are of interest to computer scientists and applied mathematicians as well as engineers working in the multifarious domains of digital industry. The tasks and problems solved by learning algorithms can be grouped according to their complexity: problems are "harder" than others in the sense that more powerful algorithms are needed to solve them. In the same spirit, an algorithm is said to be "better" than another when the former solves a problem faster or using fewer resources than the latter. The point of comparison with humans (or living animals) is important to define "being good enough" for an algorithm: given a task in an empirical setting, an algorithm is said to be "good enough" when it solves the problem in a comparable amount of time (or faster) with a comparable score as a representative group of human beings.

For instance, suppose the task is to distinguish spoken language from mere noise, given an audio input; humans are very good at this task, meaning that they rarely err and they give their response very quickly; an algorithm is good enough if it does at least as well as a human being both in success and time response. Another very famous example is that of the game of Go in which some algorithms were good enough for a long time and one algorithm (implemented in AlphaGo, developed by DeepMind) became much better than the best humans quite recently.

There is an important distinction between tasks of *supervised* learning and *unsupervised* learning. Very recently, algorithms became very good at most supervised learning tasks, but the best algorithms are still very bad when it comes to unsupervised learning tasks.

Supervised learning tasks are situations where the machine is given some feedback. The algorithm can thus use the feedback from the environment to correct itself by estimating and reducing to a minimum its predicted error. There are two kinds of supervised learning which we can call *pure reinforcement* and *reinforcement learning*.

Pure reinforcement are cases when there is little feedback, so the machine needs to be trained a lot. The paradigm example is "playing algorithms", like AlphaGo. The only feedback is whether the game was won or lost. The machine then plays millions of games against itself to improve.

Reinforcement learning are cases when there is as much feedback as there is data. The paradigm example is recognition algorithms. Each image is annotated. During the training, the machine predicts the presence of objects in the photograph and receives feedback in each trial. To be properly trained, the machines need a lot of

annotated data.

In the recent years, artificial neural networks have become a very popular method of solving supervised learning problems. Reinforcement learning tasks like object recognition or speech recognition have been solved very efficiently by neural networks. Convolutional network, also known under the name "deep learning" because they contain multiple hidden layers, were the first to achieve results comparable to humans in recognition tasks in 2012. Yann LeCun is the first to propose an algorithm based on the method of back-propagation which can be implemented efficiently in convolutional networks.

Typical cases of object recognition tasks are the following: find all the letters in a given document. The convolutional network is trained on a huge set of labelled photographed (labelled by humans). It somehow gets some characteristic features of "being a letter in a document". Then, it is tested on a set of unlabelled documents. Convolutional networks have been shown to be better than humans in this task. This reliability is now used industrially: software taking any document as input and giving a full texted document as output are legion (they are called Optical Character Recognition software); machines which automatically "read" cheques are now widespread in banks.

Interestingly, there is a lot of mathematical research to understand how these networks actually work. The recent performances of convolutional networks are difficult to explain mathematically, although there are many heuristics which are robust enough to put them into practice in the industry.

Unsupervised learning

By contrast, *unsupervised* learning are situations in which no feedback is given to the algorithm. The algorithm's task is to find out the underlying structure of the input without the help of labelled feedback. One of the difficult problems facing computer scientists is to give a precise evaluation of the possible outputs. Since there is no *a priori* "correct answer"[3], there is no straightforward way of evaluating and comparing the underlying structure of the data the algorithm could propose.

An abstract way of seeing this problem consists in trying to find the probability density function of the space in which the pieces of data "live". For instance, pictures can be seen as enormous sets of pixels, thus defining a space that can be defined and studied mathematically. The algorithm "observes" many pictures, but the density function (the predicted distribution of each and every one of the pixels) is unobserved:

[3]This is what labelled data provides you with.

one would need to have a gigantic number of disconnected examples. The task is thus to infer, from the data, the function according to which the population of examples constituting the data is distributed. The difficulty lies in the fact that the distinctive features which explain the distribution of the examples is not known *a priori*.

For example, if the task is to distinguish the foreground from the background of a moving picture, the algorithm will have to find out what the distinctive features of "being background" (and conversely "being foreground") are. But these can be very diverse in nature, here are some clues: being often out of focus, moving against a stable foreground (or the opposite), containing smaller ordinary objects than the foreground, etc.

There is no canonical way of solving unsupervised learning tasks at the moment. The underlying mathematics falters over the scarcity of possible examples compared to the huge diversity of potential distinctive features.

The problem of truth in fiction as an unsupervised learning task

Here are two examples taken from LeCun's lecture on April 8th 2016[4] which illustrate the challenges of unsupervised learning.

One is the continuation of video shots. Take as input the very short video of someone letting go of a ball down to an inclined plane, and predict the next few seconds. Humans are very good at doing this: we have strong intuitions that the ball will fall and bounce in the direction roughly perpendicular to the inclined plane. But maybe the ball is very flat and it does not bounce, or it is full of helium and it would go up. So there is a bunch of possible continuations of the movie which are very different. A good algorithm will have to define what "possible" precisely means here. But you can see how this becomes extremely difficult when you have to predict all the pixels on the screen (which is what machines do) according to the "possibles". The intuitive physics we, humans, use does not predict such low level properties, but rather abstracts away from the negligible features of the video and emphasises the crucial elements of it.

The second example is the modelling of "common sense". Take as input a natural language sentence like: "Gérard took his hat from the table and left the room", and predict all the inferences one can draw from this sentence. There are many. Several inferences are plausible: Gérard extended one of his arm to take his hat; if he was sitting, then he would have stood up; he opened the door before leaving the room;

[4] https://www.college-de-france.fr/site/yann-lecun/course-2016-04-08-11h00.htm.

etc. Several inferences are implausible: Gérard used a fishing rod; he crawled to the door; he destroyed the door open; etc. Several inferences are very implausible: Gérard put his hat on using telekinesis, and dematerialise out of the room. Note that any combination of inferences can actually by met by an *ad-hoc* scenario: very implausible inferences will naturally yield explicitly fictional scenarii. Interestingly, these three sets of inferences cannot be distinguished by their cardinality: there are just as many plausible inferences as there are implausible ones. The underlying metric of the plausible needed here is very difficult to formalise so as to make an algorithm out of it.

Neural networks are nowhere near the abilities of humans in these tasks of unsupervised learning yet.

Here is, then, the problem of truth in fiction presented as a task of unsupervised learning. The input for the reader is the text, that is a set of sentences organised in the form of a book. From this data, the reader manages to extract all the fictional propositions, or rather a *substantial* part of them. The precise definition of "substantial" here is just as difficult as the definition of "possible" and "plausible" in the above examples.[5] Note that the *a priori* knowledge for this kind of fiction-reading is very large and diverse in nature: the reader has a linguistic competence, meaning they can read and understand sentences of a natural language, thus mastering in an integrated manner all the strata of linguistic meaning: phonology, syntax, semantics, pragmatics, rhetoric, etc. But they also need to know enough of intuitive physics, intuitive behavioural science, psychology, etc. to imagine the possibility and plausibility of the fictional events. And the reader should also know some things about the distinction between fiction and nonfiction, some things about the medium, and about the genre conventions necessary to make the good inferences.

Here is a caveat though: the problem of truth in fiction is more complex in practice in that it integrates a great many other unsupervised learning tasks. The solution to the problem of "common sense" is, in a clear sense, presupposed in the problem of truth in linguistic fiction, just as understanding natural language is presupposed by reading. But this does not imply that the problem of fiction is conceptually more complex. It may simply be the coordination of several problems of the same complexity. This, of course, betrays my optimism which the above analogy is trying to back up.

I think it is helpful to see fiction-reading as a problem of unsupervised learning for two reasons: First it helps clarify the kind of cognitive tasks it relates to, even though the complexity of fiction reading may be higher than "simple" unsupervised

[5] One can see how the great beetle debate is such a relevant case of study to define what "substantial" means here.

learning tasks. Second it suggests both that the philosopher of fiction should get interested in machine learning, and that machine learning should get interested in fiction reading practices as studied by philosophers and literary theorists.

Part II

The great beetle debate

Chapter 9

Debates over disagreements

9.1 A logical space for disagreements

9.1.1 Faultless disagreements

Here is how Kölbel defines the notion of *faultless disagreement*:

> A faultless disagreement is a situation where there is a thinker A, a thinker B, and a proposition (content of judgement) p, such that:
>
> (a) A believes (judges) that p and B believes (judges) that not-p.
>
> (b) Neither A nor B has made a mistake (is at fault).[1]

Kölbel's paradigm examples are aesthetic disagreements:

> For example, imagine that Olivia believes that Matisse is better than Picasso while Felicity believes that Picasso is better. Suppose that both have had sufficient opportunity to sample the works of both artists and have given the matter enough consideration. Then it may well be that Olivia and Felicity each have exactly the view they ought to have on the question of whether Matisse is better than Picasso, and that for both of them changing their belief would constitute a mistake.[2]

Faultless disagreements are problematic for an ordinary semantic framework where the truth of a proposition is interpreted as a set of possible worlds. Indeed, if a proposition is a set of possible worlds, then it is independent of the speaker (by a

[1] [Kölbel 2004], p. 53.
[2] [Kölbel 2004], p. 54.

standard definition of possible worlds); it thus becomes difficult to see how both speakers can faultlessly disagree. This is why Kölbel takes faultless disagreement to be an argument in favour of a *relativist semantic framework*, according to which the truth of a proposition depends on parameters of evaluation distinct from possible worlds. The truth of a proposition, according to relativists, should thus be relativised to times, standards of taste, moral standards, epistemic standards or informational states. Neither of these notions can easily be interpreted using the sole notion of possible worlds.

Some philosophers, however, resist this argument for *semantic relativism* and they typically underline the fact that the paradigm examples of faultless disagreements are not disagreements in any ordinary sense.[3] Olivia and Felicity surely have a difference in opinion, but is it really a disagreement? Or maybe they are talking past each other? There is a strong intuition that *disagreeing* means disagreeing *about the truth-value* of a proposition. In other words, some theoreticians want to keep the word "disagreement" to denote only (linguistic) conflicts between people which involve a notion of truth. Olivia and Felicity are clearly in conflict but not necessarily in disagreement.

The class of aesthetic "disagreements" is controversial, because the truth-value of propositions like "Picasso is better than Matisse" is difficult to formalise. It contains an evaluative word, viz. "better", whose semantics analysis is greatly debated.[4] Whether Kölbel's example is an acceptable example of a disagreement is up for another debate. As a working hypothesis, I will follow Kölbel here, and use a loose enough notion of disagreement to cover cases like Olivia and Felicity's.

9.1.2 Fictional disagreements

Analogies and disanalogies

Imagine you just watched the latest film adaptation of your favorite novel. It is not implausible you would discuss some of the choices the film director made regarding, say, some detail of a character's physical aspect. For instance, you might have been puzzled by the character's hair colour, thus realizing you had imagined him say blond, and he appeared ash brown. Let us suppose the hair colour was left unspecified in the original novel. In this case, it seems that the content of your imagining and the content of the movie are at odds; it is hard to say *who* is right, or even *that one* is right.

[3] See for example [Stojanovic 2017].

[4] See for instance Andrés Soria Ruiz's recent dissertation [Soria Ruiz 2019].

This case strikes as analogous to Kölbel's example of an aesthetic disagreement. The analogy is clear: it seems that the director and you are in conflict; and no one has made a mistake, since the hair colour was not specified. But there are two disanalogies which have to be emphasised.

First, there is no evaluative term involved. The reason why one can think of a faultless disagreement here seem to depend on: (i) the fact that the novel and the movie are fictions; (ii) the fact that the hair colour is not specified in the original novel. So the intuitive "faultlessness" of this disagreement is not grounded in the use of an evaluative term. Rather, it seems to depend on the fact that the subject matter is *fictional*. Such a disagreement should thus be called a "fictional disagreement", leaving open the question of it being faultless or not.

Second, the two appreciators of the fiction seem to entertain *imaginings* rather than *beliefs* (or judgments). While reading the original novel, the movie director and you form two incompatible imaginings about the hair colour of the character. Imaginings are cognitively grounded in the interpretation of a work of fiction rather than on a "traditional" epistemic relation (like perception, memory or testimony). This might cause some trouble when it comes to Kölbel's second part of the definition, because it is not obvious what should count as a "mistake" in the context of imaginings. Consequently, before claiming that such fictional disagreement is faultless, one will need to define what is a "mistake" (or a "fault") in the context of imagination for Kölbel's definition to apply. I will construct such a notion in due time. For this reason, this case-study will amount to fine-tune the functional theory given in part 1. The distinction between imagining and misimagining is one of the basic distinction as discussed in subsection 5.2.1: defining what count as a "mistaken reading" should thus inform the general framework.

The Great Beetle Debate (GBD)

Stacie Friend has recently put to the philosopher's attention a similar kind of fictional disagreement in [Friend 2011]. She calls it the "Great Beetle Debate". It is a disagreement between two literary critics who read *the Metamorphosis* by Franz Kafka. They are disagreeing over the ultimate nature of Gregor Samsa's insecthood. One critic, Nabokov, claims that Gregor is a beetle; while the other, Smith, says that he is a cockroach; and none of them seem to have made any mistake in their reading.

This chapter is dedicated to a detailed analysis the GDB. In particular, I will ask whether the GBD is a faultless disagreement or not. I will defend that it is an interesting faultless disagreement which reveals much of how the complex processes of fictional imagination at play when one reads fiction.

9.1.3 The logical space

These primary considerations provide us with two orthogonal distinctions defining a logical space in which I will be working.

Following Kölbel, one can distinguish *faultless* from *faulty* disagreements. A *faulty* disagreement is a situation in which two subjects hold incompatible views, and at least one of them is at fault. In other words: they cannot be *both* right. But, of course, they can both be wrong. A *faultless* disagreement is a situation in which none of them is at fault. An example of a faulty disagreement would be arguing over the number of trees there are in the Jardin des Plantes; an example of a faultless disagreement would be arguing over the fact that the trees in the Jardin des Plantes are beautiful.

Following Friend's lead, one can also distinguish *real* from *fictional* disagreements. Real disagreements are disagreements about the truth-value of a proposition whereas fictional disagreements are about the fictionality of a proposition. The *truth-value* of a proposition is grounded on worldly facts, or the knowledge we have thereof. The *fictionality* of a proposition, on the other hand, boils down to what is prescribed to be imagined by a work of fiction.[5] Both examples involving the trees of the Jardin des Plantes are real disagreements. By contrast, the GBD and the movie adaptation scenario are fictional disagreements.

It seems clear that fictional disagreements can either be faultless of faulty. Arguing over an unspecified detail intuitively yields a faultless disagreement, as seen above. By contrast, in arguing about something that is explicit in the fiction, it seems that one party must be making a mistake. For instance, suppose a reader reads *Anna Karanina* so poorly as to genuinely think that (fictionally) Anna Karanina is a man.[6] Now, two readers can disagree about Anna being a woman or a man: that is, by definition, a fictional disagreement. It also is faulty, since one party is clearly at fault. The disagreement should be settled once we sit down and read relevant passages of Tolstoi's novel.[7]

As a result, here is a table of the logical space of disagreements thus defined:

[5] See chapter 5 for a discussion of this definition.

[6] I originally thought I could prove that it was impossible to have a faulty, fictional disagreement. But Stacie Friend, whom I thank, produced this very intuitive example.

[7] A faulty disagreement, in other words, is one that can be settled by such a Leibnizian idealised situation in which the two parties sit down and say: *calculemus*!

	Faultless	Faulty
Real	Disagreeing about the beauty of the trees in the Jardin des Plantes.	Disagreeing about the number of trees in the Jardin des Plantes.
Fictional	Disagreeing about the unspecified hair colour of a fictional character.	Disagreeing about whether Anna Karenina is a woman or a man.

Table 9.1: A logical space for disagreements

Now the question is: where should we put the GBD? I emphasised the similarity between the GBD with the unspecified hair colour disagreement. But the GBD is actually a little more complex than it first seems. Before I can do a precise analysis of the GBD, I need to set up the debate in precise terms.

9.2 The Great Beetle Debate

9.2.1 Todorov's definition of the fantastic

> "J'en vins presque à croire": voilà la formule qui résume l'esprit du fantastique. La foi absolue comme l'incrédulité totale nous mènerait en dehors du fantastique; c'est l'hésitation qui lui donne vie.
>
> [Todorov 1970], p. 35.

In the introduction of a lecture on *the Metamorphosis*, Nabokov starts with several remarks on the methodology of literary analysis.[8] One of the first things to do, Nabokov emphasises, is to define the *genre* of the text, which is to guide the interpretation of the text. *The Metamorphosis* is a *fantastic* story, Nabokov argues.

The term "fantastic" has become a technical term of literary analysis denoting a specific genre, thanks to the work of Tzvetan Todorov. In 1970, Todorov published a famous essay entitled *Introduction à la littérature fantastique*, translated into English three years later under the title *The Fantastic: A Structural Approach to a Literary Genre* (translated by Richard Howard in Cleveland: Case Western Reserve University Press). The logic of literary genres is notoriously difficult to systematise. Todorov's proposal is to tackle the definition of the fantastic by contrasting it with other close literary genres like the marvelous or some science-fiction. He then tries to put forward positive features of the fantastic.

[8] Nabokov's lectures are transcribed and published in [Nabokov 1980].

In the following, I will define the "T-fantastic" (T for Todorov), which is my interpretation of Todorov's definition.[9] I am not sure how Todorov himself would welcome my definition of the T-fantastic: I will say a word about it later.

A story is *T-fantastic* when something "uncanny", apparently super-natural happens in a world in which the "laws of reality" as we ordinarily experience them remain intact otherwise.[10] So the T-fantastic is to be sharply distinguished with the marvelous, or the oneiric, in which the "laws of reality" are explicitly altered. For instance, talking animals, magical creatures or metamorphoses of any kind are very common fictional events in fairy tales, and this shows that the fictional events do not occur in a world nomologically similar to ours. Maybe some "laws of reality" remain unchanged in some marvelous stories, for example the law of gravitation. Indeed, objects and people regularly fall in fairy tales. But there is a very clear sense in which the worlds to be imagined in appreciating marvelous stories are nomologically different from the world of everyday experience: extraordinary, magical things also happen in fairy tales.

The same holds for some science-fiction stories in which the reader is explicitly required to imagine that some "laws of reality" are altered. For instance, in some science-fiction stories, one can be teleported away, or travel back in time, or plug one's consciousness into a computer and "live" forever. These are explicitly going against both ordinary experience and current scientific results.[11]

This is not what happens with the T-fantastic. The typical T-fantastic story presents a world which could very well be our world for all we know, were it not for this one phenomenon which is blatantly at odds with everyday experience. For instance, in *La Vénus d'Ille*, a short story by Prosper Mérimée, an ancient bronze statue appears to have closed its hand by itself: a character put a ring on its finger at some point in the story, and later on the statue's hand is closed on the ring.

[9] Todorov's definition, though central to the subject, is not the last word when it comes to defining the genre *fantastic* in the field of literary studies. I will not discuss all the available definitions and their subtle differences. For other interesting definitions, see for instance [Caillois 1965], [Brooke-Rose 1983], [Malrieu and Vercier 1992], [Mellier 2000].

[10] Caillois uses the term "inadmissible" to express the same idea in [Caillois 1965]:

> Tout le fantastique est rupture de l'ordre connu, irruption de l'inadmissible au seuil de l'inaltérable légalité quotidienne.

My translation:

> The fantastic is essentially a violation of the known order, an irruption of the inadmissible into the unalterable everyday legality.

[11] Some "optimistic" minds would rather say "ahead" instead of "against"...

Consequently, the characters in the story are deeply puzzled by this fact, and the narrator of the story doubts that this is what actually happened. He considers several rational explanations without coming to any convincing conclusion. Given the puzzlement of the fictional characters, a supernatural explanation as in a fairy tale is not straightforwardly available. Besides, such appeal to magic would simply ruin the whole point of the short story. Indeed, the point of the T-fantastic as a genre is that the reader is naturally trying to look for a rational explanation underlying the uncanny fictional event. By definition, the conventions of the genre *trigger* rational thinking on the part of the reader, by challenging it. This challenge to rational thinking implies a feeling of hesitation about what is true in the fiction, such hesitation usually develops into a feeling of unease, which is characteristic of the T-fantastic.

Todorov and the T-fantastic

I have to say that Todorov is less explicit than I in distinguishing the uncanny fictional event being part of the fictional world and what the reader actually imagines when they read. I put most of the weight of my definition in the reader's stance towards the text, trying to hold fixed the idea that the world of the fiction and the real world differ little in their nomological structure.

Todorov, on the other hand, takes a narratological point of view and he insists that the fictional world depends, in some relevant sense, on what the reader actually imagines. So that the distinction between what is a fictional event and what is in the reader's imagination flickers. In my definition the felt *hesitation* on the part of the reader is a *consequence* of the way the fictional world is presented to the reader as being. Todorov, on the other hand, suggests that the felt hesitation is *definitional* of the genre.

It is worth reading Todorov on this:

> Here is the core of the fantastic. In a world which is indeed ours, the one we know, without devils, sylphs, nor vampire, there happens a singular event which cannot be explained following the laws of this familiar world. Those who perceive the event must opt for either of these two possible solutions: either it is an illusion of the sense, a product of the imagination and the laws of reality remain what they are; or the event actually occurred, it is part and parcel of reality, but then this reality follows laws unknown to us. Either the devil is an illusion, an imaginary being; or it actually exists, as any other living being: with this little reserve that one rarely encounters it.

> The fantastic fills the time of this doubt; once either of the answer is selected, one leaves the fantastic and enters a neighbour genre, the strange or the marvelous. The fantastic is the felt hesitation of a being who knows no other laws than the natural laws, when faced with an apparently supernatural event.[12]

In order not to be committed to any definite interpretation of this complex and subtle definition, I introduced my definition of the *T-fantastic.* Though, I think I am quite faithful to the Todorov's spirit, if not to it's letter.

9.2.2 Application of Todorov's definition to Kafka

Nabokov starts his lecture on *the Metamorphosis* by pointing out that it is a *fantastic* story. I will claim that it is T-fantastic to be more precise.[13] The "apparently supernatural event" of *the Metamorphosis* is in the title and in the opening sentence:

> One morning, upon awakening from agitated dreams, Gregor Samsa found himself, in his bed, transformed into a monstrous vermin.[14]

[12]This is my translation. Here is the original paragraph in French, [Todorov 1970] p. 29:

> Ainsi se trouve-t-on amené au coeur du fantastique. Dans un monde qui est bien le nôtre, celui que nous connaissons, sans diables, sylphides, ni vampires, se produit un événement singulier qui ne peut s'expliquer par les lois de ce même monde familier. Celui qui perçoit l'événement doit opter pour l'une des deux solutions possibles: ou bien il s'agit d'une illusion des sens, d'un produit de l'imagination et les lois du monde restent alors ce qu'elles sont; ou bien l'événement a véritablement eu lieu, il est partie intégrante de la réalité, mais alors cette réalité est régie par des lois inconnues de nous. Ou bien le diable est une illusion, un être imaginaire; ou bien il existe réellement, tout comme les autres êtres vivants: avec cette réserve qu'on le rencontre rarement.
>
> Le fantastique occupe le temps de cette incertitude; dès que l'on choisit l'une ou l'autre réponse, on quitte le fantastique pour entrer dans un genre voisin, l'étrange ou le merveilleux. Le fantastique, c'est l'hésitation éprouvée par un être qui ne connaît que les lois naturelles, face à un événement en apparence surnaturel.

[13]History is harsh: Todorov's work was published after Nabokov's lecture, so it would be a mistake to claim that Nabokov had Todorov's definition in mind. I do not know precisely what Nabokov's definition of the fantastic would be, nor do I know of any scientific work on the subject. I am quite confident to follow Nabokov's spirit in these lines, though.

[14]This is Joachim Neugroschel's translation in [Kafka 1915], p. 90. Here is the Kafka's original wording in German:

> Als Gregor Samsa eines Morgens aus unruhigen Träumen erwachte, fand er sich in seinem Bett zu einem ungeheueren Ungeziefer verwandelt.

The rest of the short story prescribes no other supernatural event to be imagined. That is to say, apart from the first sentence, the world of Gregor Samsa appears to follow the same "laws of reality" as we experience them in our world. To be really precise, Kafka's story does not even require to imagine the metamorphosis of a man into a "monstrous vermin" *qua* super-natural *process*, but only the result of the metamorphosis. So the supernatural event is merely *presupposed*: the story begins after the metamorphosis is complete. I think this detail makes it an even more paradigmatic case of the T-fantastic.[15] There is a clear sense, however, *the Metamorphosis* by Kafka is of the T-fantastic genre, since the reader is required to imagine that a man named Gregor Samsa metamorphosed into a "monstrous vermin" during the preceding night, in a world roughly similar to ours in terms of natural laws.

Kafka makes it difficult to interpret the metamorphosis as a bad dream. Indeed, in the second paragraph, it is explicitly said that "it was no dream". Moreover, Kafka describes the room in some detail so that Gregor's phenomenology does not match that of a dream. Grgor recognises his "human" surroundings and concentrates on what he can see in a very ordinary way:

> "What's happened to me?" [Gregor] wondered. It was no dream. His room, a normal if somewhat tiny human room, lay quietly between the four familiar walls. Above the table, on which a line of fabric samples had been unpacked and spread out (Samsa was a travelling salesman), hung the picture that he had recently clipped from an illustrated magazine and

[15]Todorov raises two issues when it comes to classifying *the Metamorphosis* as fantastic. The first is that the supernatural event is not, strictly speaking, part of the story but it happens before. It is an issue for him because it runs against his further idea that the plot of a fantastic story finds its acme when the supernatural event occurs. The second problem with Kafka's story is that the characters in the story are not so puzzled, which is puzzling. Consequently, it seems that the readers are not hesitant as they should be if the story was fantastic in his sense.

On the other hand, Todorov remarks that the effect of Kafka's story on the reader is that of "adaptation" which he takes to be the dual of "hesitation" ([Todorov 1970], p. 179-180). He then follows Sartre in considering that Kafka's handling of the supernatural should be regarded as a completion of the fantastic. As Todorov puts it ([Todorov 1970], p. 182):

> With Kafka we are confronted with a generalised fantastic: the whole world of the book and the reader itself are part of it.

He finally gets away with it by distinguishing a "classic" fantastic from Kafka's fantastic. The former entails hesitation while the latter turns the fantastic's hesitation into an other form of unease, another kind of metamorphosis.

By contrast, these two issues do not arise if one uses the definition of the *T-fantastic*.

inserted in a pretty gilt frame.[16]

As for the possibility of a metaphorical interpretation, it is not easily accessible.[17] The word "vermin" has a metaphorical sense designating a despicable human being. Here is an example of such use taken from the Merriam-Webster dictionary of English online: "The *vermin* who looted abandoned houses after the hurricane." But, in the opening scene, considerations about the number of his legs, the sounds he produces and the disgust his family feels when seeing him for the first time strongly suggests that Gregor has turned into a non-human creature.

Maybe some readers will resist imagining such an impossible metamorphosis while maintaining that the "laws of reality" be roughly the same as in our world. Resistance is surely expected, and that, I think, is part of what Todorov means by the reader's "hesitation". Refusal to imagine, however, is not expected. I will have to say more about what the reader is expected to do when they interpret a fictional text, but let us contend ourselves for the moment with the idea that reading fiction comes with a charitable attitude on the part of the reader so that they accommodate what is written in the book, and do not drop the book at once when they can see a conflict between what is supposed to be the "laws of reality" imported into the fictional world and some specific fictional events.

9.2.3 Nabokov's comment on *the Metamorphosis*

Now, one plausible "law of reality" is *determinacy*, meaning that for any given set of incompatible properties, an object has one and only one of them. If we take incompatible properties to be contradictories, it means that for a given object and a given property, the object either has the property or the negation of the property.[18] One simple way of interpreting *determinacy* is to say that there is an answer to questions of the form: "Has O the property P?"

[16] [Kafka 1915], p. 90.

[17] Although is has been proposed in [Kazin 1941], p. 147:

> The genius of this story is that which transforms a figure of speech ("You treat me as if I were an animal!") into a parable of the inner distance between human beings.

[18] See for example a formulation of such a universal principle in Kant's *Critique of Pure Reason*, section "The transcendental ideal" A573/B601. Translation by Norman Kemp Smith, 1929:

> According to this principle [viz. *determinability*], of *every two* contradictorily opposed predicates only one can belong to a concept. This principle is based on the law of contradiction, and is therefore a purely logical principle. As such, it abstracts from the entire content of knowledge and is concerned solely with its logical form.

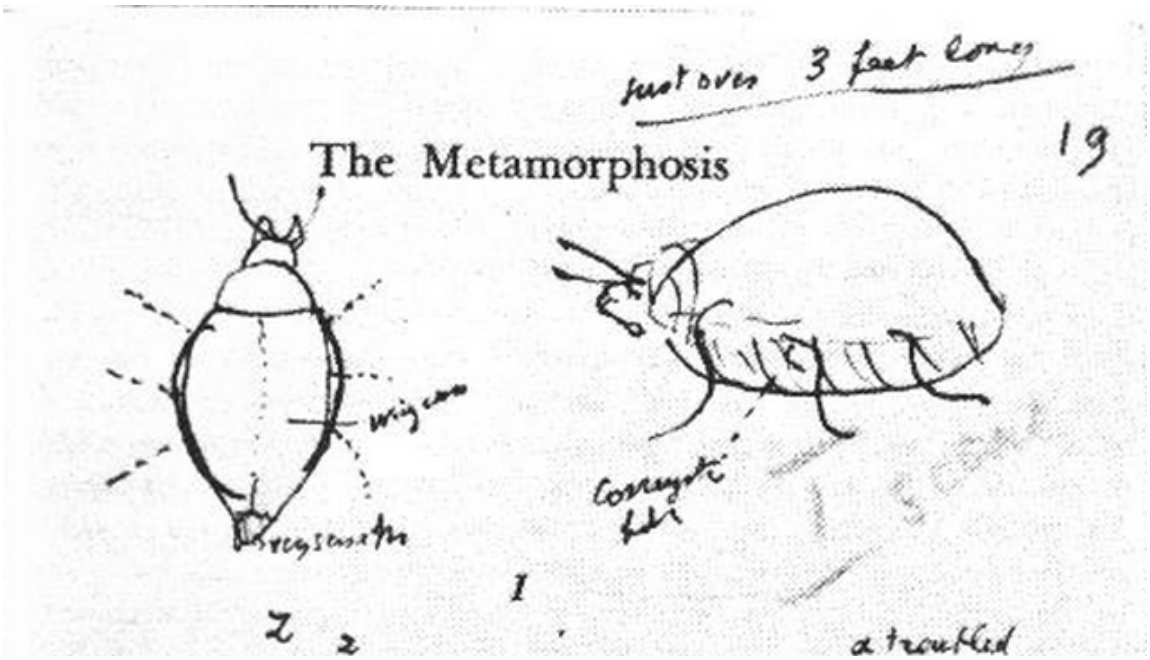


Figure 9.1: Facsimile of Nabokov's lecture notes on *the Metamorphosis*.

At this point, Nabokov makes an argument. Given that *the Metamorphosis* is a T-fantastic story, objects in the world of *the Metamorphosis* are as determinate as objects in the real world. So Gregor is determinate. If Gregor is determinate, then one can in principle determine what kind of "vermin" he has turned into. Here is Nabokov's argument in the form of a Q&A:[19]

Q1: What kind of *vermin*[20] has poor Gregor turned into?

A1: (Nabokov:) "An arthropod" (the class of animals containing spiders, insects, crustaceans and millipedes).

[Why? because it is said in the text that: "His numerous legs, pitifully thin in comparison to the rest of his circumference, flickered helplessly before his

[19]I am responsible for the analytic reconstruction of Nabokov's argument here. This means that I explicitly formulate the questions which are merely implicit in Nabokov's text, and I put between quotation marks what I take to be Nabokov's "answers".

[20]The German word is *Ungeziefer*. For simplicity, I will henceforth use the English word only and ignore the specific overtones of the German word, for I do not really know how much they differ from English (including the semantic vagueness of the word).

eyes.", and the rest of the text makes it quite clear that he must be an insect (taking into account what he eats, the fact that he can walk on walls, etc.).]

Q2: What kind of insect has Gregor turned into?

A2: (Nabokov:) "Some say a 'cockroach'. But a cockroach is flat on both sides and has long legs. On the contrary Gregor is manifestly round, back and belly, and he has short legs (this is crucial to the understanding of the first scene in which Gregor has enormous difficulties to get out of his bed). [...] He must be a beetle."[21]

"Smith" is the name Friend gives to the critic disagreeing with Nabokov at the level of A2. Here is, in a nutshell, the Great Beetle Debate:

Smith: Gregor Samsa has turned into a cockroach.

Nabokov: No, Gregor Samsa has turned into a beetle.

[21]Nabokov also considers the information the text gives about his colour, has a brief comment about the possible wings of Gregor, discusses his size, finally draws how he imagines Gregor to be in his lecture notes. The drawing is reproduced in Figure 9.1. He also considers the fact that, at the end of the story, the maid calls Gregor a *Mistkäfer* (dung beetle), but the term in German is used as a pet-name in context so it should not be considered a strong argument.

Chapter 10

Imagining past each other

I think it is fair to say that for most readers of *the Metamorphosis* (among which I count myself) this entomological debate over the ultimate nature of Gregor, when heard for the first time, is somehow abstruse. The "average reader", if I may say so, is arguably closer to Smith than Nabokov in this debate. I suppose they would readily answer Q1 in a way similar to A1 (though undoubtedly with less precise a word than "arthropod", chosen by Nabokov); I also suppose they would not raise Q2 spontaneously, but if presented with Q2, they would answer somehow differently than Nabokov saying something like: "I don't know, probably a cockroach", with a small degree of confidence, and little "objective" evidence to display. Why "cockroach"? Either because anti-Nabokov critics have spread the word and it has already pervaded the collective mind, or more probably because the cockroach is one of the prototypical images of a vermin in everyone's mind.

This phenomenological description of the debate suggests a descriptive analysis: the GBD is nothing more than two readers imagining their way. One is imagining a cockroach, the other a beetle. But really, they are imagining past each other. The GBD seems to be a disagreement, because Nabokov and Smith think they are talking about the same thing, namely Gregor. But Gregor exists only in their imagination. What is going on is that Nabokov and Smith talk about the way they imagine Gregor to be. The only thing one can do is thus to explain why they imagine the way they do, if there is a why.

According to this line of thinking, the GBD is not really a disagreement since the two parties do not, strictly speaking, share the same ground on which they could disagree. I do not think this descriptive analysis of the GBD is exhaustive, for it crucially puts aside a normative dimension at play in the GBD. In the next chapter, I will show that there are good reasons to think that Nabokov and Smith do not

imagine past each other. For the time being, though, I will explore this descriptive analysis of the GBD. First, I think it will help clarify some important conceptual distinctions. Second, I think there is a very interesting follow-up on this descriptive analysis involving the concept of *imaginative resistance*.

10.1 Prototypes and fictional imagination

Prototypes have raised interesting conceptual and empirical investigations. I think they are the underlying phenomenon of the phenomenological data presented above.

One can find a seminal work towards a prototype theory of concepts in [Rosch and Mervis 1975]. Originally, it was designed to test empirically an idea which traces back to the second Wittgenstein: the semantic contribution of common nouns in natural languages is to bring up to mind, or evoke, or prompt prototypical exemplars associated with the common noun. Accordingly, in processing the expression, say: "a bird was eating" without any particular context, it prompts something like a robin or a sparrow eating in one's mind. By contrast, it is very improbable that the indefinite expression "a bird" be empirically associated with a sulphur-crested cockatoo or an Atlantic puffin, although such realisations would be correct if we only consider the truth-value of the sentence "a bird was eating". This fact was taken by the second Wittgenstein to be an argument for different kinds of pragmatic approaches to the meaning of common nouns expressions which "go beyond" truth-conditions (against the first Wittgenstein).

This observation suggests that the semantic content associated with a common noun has a specific structure: among the different possible instances of the kind designated by the common noun expressions, one (or a small group) is prototypical. But this structure is notoriously difficult to describe and investigate precisely. Here are three general problems: First, the "distance" of a non-prototypical exemplar to the prototype is a very tricky notion to define. For example, what is the "more prototypical" bird between a hummingbird and an ostrich? Or between an eagle and a vulture?

Secondly, prototypes vary a lot if we consider different people and cultures. Indeed, my prototypical bird is probably different from yours and a European's prototypical bird is probably different from an Australasian's one. But common nouns expressions do not seem to vary as much. We both use the same word "bird" to talk about the idiosyncratic prototype we have in mind and, similarly, the word "bird" has an equivalent in all cultures, more or less.[1] So in trying to correlate common noun

[1]Modulo Quine's wicked translations. See the "gavagai" example in [Quine 1960] and [Quine

expressions with prototypes, one has to account for this underlying psychological variability, which is a very difficult problem.

Finally, prototypes appears to be very malleable and changing so that it is doubtful that a proper compositional semantics can be built on top of them. By which I mean that inference patterns using prototypes are very difficult to systematise.

But if prototype theory may not fulfill its promises as a semantics for natural languages, it is arguably more appropriate to use it for a theory of fictional imagination.

What I have just said may suggest that prototypes are a theory of mental imagery. This is not the case. Prototypes thus defined aim at explaining what are the building blocks of the propositions expressed by natural language expressions containing common nouns. The empirical theory of prototypes hinted at above aims at being a theory of propositions. Hence, appealing to prototypes for a theory of fictional imagination aims at theorizing *propositional* imagination and not mental imagery.

A theory of propositional imagination is an answer to the problem of the fictional *inferential mechanism*. In other words, it consists in answering the question: what are the inferences licensed by the reading of a text of fiction? Such a theory should explain how all the fictional propositions accessible from the reading of a text of fiction are constructed out or retrieved from the set of real propositions and the specific cognitive mechanisms underlying fiction reading. There are competing theories, but most agree on the basic principles underlying this fictional inferential mechanism. These are the reality principle and the mutual belief principle. In order to describe the psychological data, some put more weight on the first than on the second, others do the opposite.[2]

Interestingly, the empirical theory of prototypes shows how the two principles come together very neatly. Here is a plausible principle:

> (*Prototypical fictional reading*) Whenever you come across a familiar common noun expression while reading a text of fiction, try to depart the least possible in imagination from the prototype it denotes in a non-fictional context (given the other prescriptions to imagine).

In this principle, both the mutual-belief and reality principle are interwoven.[3] However, it requires a precise theory of prototypes. This much is enough to interpret the phenomenological data of the GBD.

1969].

[2]See section 6.3 for a discussion of these principles.

[3]Note that it resembles the appearance principle as defined in subsection 7.3.3 without presupposing a causal theory of semantic information.

10.1.1 The average reader's prototypical vermin

I said that the "average reader" was closer to Smith than Nabokov in the GBD. This is an empirical question. One should test the readers of *the Metamorphosis* and focus on their interpretation of the common noun expression "a vermin". The question boils down to: what is the "most prototypical" vermin between a cockroach and a beetle? My intuition is that people would significantly prefer cockroaches to beetles as a prototypical realisation of the expression "a vermin".

For all that it's worth, Gabriel Garcia Marquez has a very humorous flight of lyricism on cockroaches being *the* prototypical vermin of the creation:

> The old bookseller, knowing about Aureliano's love for books that had been read only by the Venerable Bede, urged him with a certain fatherly malice to get into the discussion, and without even taking a breath, he explained that the cockroach, the oldest winged insect on the face of the earth, had already been the victim of slippers in the Old Testament, but that since the species was definitely resistant to any and all methods of extermination, from tomato dices with borax to flour and sugar, and with its one thousand six hundred three varieties had resisted the most ancient, tenacious, and pitiless persecution that mankind had unleashed against any living thing since the beginnings, including man himself, to such an extent that just as an instinct for reproduction was attributed to humankind, so there must have been another one more definite and pressing, which was the instinct to kill cockroaches, and if the latter had succeeded in escaping human ferocity it was because they had taken refuge in the shadows, where they became invulnerable because of man's congenital fear of the dark, but on the other hand they became susceptible to the glow of noon, so that by the Middle Ages already, and in present times, and *per omnia secula seculorum*, the only effective method for killing cockroaches was the glare of the sun.[4]

This explains Smith's reading in the GBD: it so happens that the semantic contribution of the expression "a vermin" without further context is a prototypical cockroach. This is why Smith, as a prototypical average reader, imagines the way he imagines.

Note that I have simplified the problem here, by considering only two possible realisations of the "a vermin", for simplicity. But, of course, one can consider other options in principle. Let me have three little notes on this.

[4][Garcia Marquez 1967] (ch. 19) p. 305.

A note about translations

The Metamorphosis has been translated into English a few times. *Ungeziefer* is a problematic word. It has been translated by "vermin" (trans. Willa Muir 1952, trans. Stanley Corngold 1972, trans. Joachim Neugroschel 1993) but also by "insect" (trans. Edwin and Willa Muir 1933, trans. Susan Bernofsky 2014), by "cockroach" (trans. Michael Hofmann 2007) and even by "bedbug" (trans. Christopher Moncrieff 2014). There are many more. Same for French: "vermine" has been chosen by Alexandre Vialatte, while Claude David translated the word into "cancrelat" (which is a synonym for cockroach). So translators are not unanimous. A comparative study of all the translations could be a precious work to consult.

A note about adaptations

The Metamorphosis has been adapted for the cinema and television a dozen times. In such adaptation to a visual media, one is forced to decide what Gregor *looks like*. Interestingly, the available adaptations are very diverse and the possible solutions also depend on the visual media.

In a comic book, Gregor has been drawn as a cockroach. If we compare the many different front pages of editions of *the Metamorphosis*, one can find many illustrations of cockroaches and beetles, but some editors chose to represent nonexistent vermin-like creatures, or other kinds of insects. On the cover of a paperback edition of Claude David's translation, one can even see a grasshopper...[5] There also have been several filmed adaptations and adaptations for the stage which make choices of their own.

Here is a first interesting example. The Royal Opera House produced a dance-theater adaptation of *the Metamorphosis* by the Portuguese choreographer Arthur Pita, created for the British ballet dancer Edward Watson.[6] Obviously, Gregor is represented as a human being, given this is a ballet.

Another interesting example is that of the flamenco dancer Israel Galván. In an interview, he describes how he thought of Gregor as a mix of different insects in his dancing vocabulary. Both the cockroach and the beetle are considered along with the praying mantis and the scorpion. I translate from the French version of the interview:

> When I read Kafka, I imagined how the insect would dance. I considered some change in dancing techniques, so that I could reach a new

[5] Which is actually incompatible with Claude David's translation!

[6] Here is the Royal Opera House documentation on the show and a filmed representation available on youtube.

> way of moving one's body. It cam from the fact that I was considered a "strange animal" (*un "bicho raro"*) in my family. On the other hand, as if to play, I began to see other dancers as insects... I classified dancers. For instance Mario Maya was the praying mantis, Farruco the scorpion, Enrique el Cojo the beetle, El Guïto and Vicente Escudero other insects... For "La Metamorfosis", I based myself on Mario Maya, Enrique et Cojo, El Guïto et Farruco.[7]

Again, a comparative study of adaptations for visual media could be a precious work to be done.

A note on Kafka's intentions

Kafka's intentions have *purposely* remained obscure. Discussing the front page of his book, Kafka wrote the following to his publishing house: " The insect itself is not to be drawn. It is not even to be seen from a distance." His wish was fulfilled in the first edition of the book in 1915 (see the original cover in Figure 10.1), but his wish is now long forgotten as mentioned above. His intentions, because their are negative, are difficult to interpret, though. Probably, he did not want to influence the reader's imagination when it comes to Gregor's appearance. Maybe he was trying to make a point about prototypical imagination...

10.1.2 Nabokov's prototypical vermin

> Suppose that everyone had a box with something in it which we call a "beetle". No one can ever look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But what if these people's word "beetle" had a use nonetheless?
>
> [Wittgenstein 1953], §293.

Nabokov was a lepidopterist.[8] Lepidoptery is a branch of entomology. As such,

[7] Thanks to Anne-Sophie Riegler for bringing this work to my attention.

[8] For instance, Nabokov published a little article entitled "The Female of *Lycaeides Argyrognomon Sublivens*" published in *the Lepidopterists' News* in 1952, introducing a new butterfly. He first found it when he was location scouting for *Lolita*. He was also curator of lepidoptery at the Harvard Museum of Comparative Zoology at the end of the 40s.



Figure 10.1: *Die Verwandlung* (original cover)

Nabokov was an expert in insects. As an entomologist, Nabokov's prototypical insect-vermin was very plausibly different from Smith's, who is not an entomologist (*ex hypothesis*). This difference is easily explained by appealing to prototypes. Given the variability of prototypes among individuals, it is clear that being an expert on something must have an effect on the kind of prototype one entertains. Nabokov thus markedly departs from Smith and other average readers when it comes to prototypes of insects.

One can now see how Nabokov and Smith are *imagining past each other*. Nabokov, due to his training in entomology, has a very refined imagination when it comes to imagining insects. Given (Prototypical fictional reading), the reading of *the Metamorphosis* would prompt different prototypes in Nabokov's and Smith's mind. Very relevantly, Nabokov is aware that cockroaches do not get stuck on their backs, but that beetles do; whereas Smith is not. Hence, for Nabokov, the most prototypical "vermin" matching Gregor's description in Kafka's text is a beetle, and not a cockroach.

Now you know enough of entomology, you should be inclined to imagine like Nabokov and depart from Smith. This is a striking observation which is in tension with our current hypothesis following which Nabokov and Smith merely imagine past

each other and, consequently, do not really engage in any disagreement.

This change suggests that Nabokov's rejection of Smith's reading is an instance of *imaginative resistance*. This getting together the GBD and the phenomenon of imaginative resistance should account for the agonistic flavour of the GBD.

10.2 Imaginative resistance

Imaginative resistance is a well-known phenomenon to philosophers of fiction. The term originated in the seminal article [Walton 1994]. Walton claims that the phenomenon was already spotted in [Hume 1757].

Imaginative resistance is the term of art which designates the phenomenon of spontaneously refusing or resisting to imagine that such and such event is fictional. One expects imaginative resistance to be a systematic response in two kinds of situation. The first one is when the fictional event to be imagined clashes with the reader's moral value(s). The second one is when the fictional event to be imagined clashes with some basic knowledge the reader has about the world. The first case is a case of *evaluative* imaginative resistance; the second is a case of *descriptive* evaluative resistance.[9]

The literature on imaginative resistance has mainly focused on evaluative imaginative resistance but it is now accepted that both kinds are a kind of imaginative resistance. There are arguably other types of imaginative resistance. For instance, Hume's original formulation was an aesthetic claim: he observed that we tend to judge that the fictions which go against our basic moral value(s) are *worse* than those which follow them. In the literature, it is now common to distinguish between the different "puzzles of imaginative resistance".[10]

In the following, I will show how the GBD can be interpreted as a kind of descriptive imaginative resistance. Although, the GBD is more complicated. Indeed, it is a conflict between *two* readers while imaginative resistance, in normal cases, involve only one reader. As such, the GBD can be seen as a challenge to extend current theories of imaginative resistance.

[9]This terminology is adapted from [Kim, Kneer, and Stuart 2018] who talk of "counter-*evaluative* and counter-*descriptive* propositions".

[10]See [Gendler and Liao 2016] for an exhaustive and opinionated review of these puzzles. See also [Weatherson 2004] for similar distinctions and [Walton 2006]. See [Reboul 2011] for a different perspective on the underlying problem.

10.2.1 The GBD as a case of descriptive imaginative resistance

The distinction between *evaluative* and *descriptive* imaginative resistance is intuitively the following: the first kind boils down to a *refusal* to accept that such and such is fictionally the case, and the second to an *incapacity* to imagine that such and such is fictionally the case.

Evaluative case

Here is the famous Giselda story, taken from Walton's seminal article:

> Giselda gave birth to her fourth child. In killing her baby, Giselda did the right thing; after all, it was a girl.

The reader will spontaneously refuse to imagine that it is fictional that "Giselda did the right thing". The processing of the second sentence should be jarring: this has been informally called the "wow effect". This phenomenon is robust and is now empirically investigated.[11]

There are natural ways to get around this jarring effect which are part of the *explanandum*. One would typically imagine that it is not fictional that Giselda did the right thing, but it is fictional that *she thinks* she did the right thing. Or one could imagine that Giselda lives in a society foreign to ours in which it is mutually believed that killing baby-girls is considered a "right thing to do".

This case of imaginative resistance is *evaluative* in that it conflicts with some moral value(s) of the reader. Here is a general definition:

> The *puzzle of imaginative resistance*: the puzzle of explaining our comparative difficulty in imagining fictional worlds that we take to be morally deviant.[12]

Descriptive case

By contrast, *descriptive imaginative resistance* is a case in which no such conflict with moral values is at play. Rather, what is required to be imagined is strikingly at odds with what the reader *can* indeed imagine. This kind of imaginative resistance typically arises with response-dependent descriptive concepts, such as shape.

[11] See for instance [Kim, Kneer, and Stuart 2018].
[12] In [Moran 1994], cited and adapted in [Gendler 2000].

A famous example comes from [Yablo 2002]. Yablo wrote a story about people trying to find an "oval-shaped maple leaf". Everyone knows (at least in Canada) that a maple leaf is five-fingered shaped: three large "fingers" in the middle and two small ones surrounding the stem. So imagining an oval-shaped maple leaf is something like a geometrical impossibility: we cannot do it, hence we resist imagining an oval-shaped maple leaf.

Here is Yablo's story prompting imaginative resistance:

> They flopped down beneath the great maple. One more item to find, and yet the game seemed lost. Circles, squares, octagons – all the other shapes there were plenty of. Why not the one on their list? Hang on, Sally said. It's staring us in the face. What kind of tree are we lying under? A maple tree. She tore off a leaf. Here was the oval they needed! They ran off to claim their prize.[13]

Arguably, the reader should resist in imagining the fictional situation, on the ground that it conflicts with some basic knowledge about maple trees. As Yablo remarks: "the leaf's physical properties having been settled, whether the leaf is oval or not is not up to [the author]."

The Great Beetle Debate

Now that you know enough of entomology, you can see how the great beetle debate can be seen as a case of descriptive imaginative resistance. Cockroaches, when put on their back, have no difficulties turning around.[14] So imagining Gregor as a monstrous cockroach *and* having difficulties turning around when he wakes up in his bed, is somehow similar to imagining an oval-shaped maple leaf. This suggests that Nabokov is experiencing a form of descriptive imaginative resistance.

Of course, there is a big difference with Yablo's story: it is not written in the text that Gregor is a big cockroach *and* it is stuck on its back. Kafka's text requires only that the reader imagines Gregor to be "a vermin". So the kind of imaginative resistance Nabokov experiences is towards Smith's reading of *the Metamorphosis* and not towards Kafka's text. This is a tricky extension of the problem of imaginative resistance that we need consider.

[13] [Yablo 2002], §21.

[14] If you doubt this fact, let us just suppose that this is true.

10.2.2 Interpreting imaginative resistance

In the recent [Altshuler and Maier 2018], Altshuler and Maier try to give an general explanation of the phenomenon of imaginative resistance. Their claim is the following: imaginative resistance triggers a kind of "wow effect"; this "wow effect" is the result of an accommodation mechanism much like a resolving of a presupposition; but the reason why it is so jarring is because this accommodation pattern involves the making salient of a hidden narrator. Their thesis is that imaginative resistance is a kind of accommodation involving the making salient of a narrator.

Accommodation is a term of art introduced in [Lewis 1979b], but it was first discussed in [Stalnaker 1977] and [Karttunen 1974]. It denotes a very general phenomenon in the use of language which Karttunen introduces thus:

> Ordinary conversation does not always proceed in the ideal orderly fashion. [...] People do make leaps and shortcuts by using sentences whose presuppositions are not satisfied in the conversational context. This is the rule rather than the exception. [...] I think we can maintain that a sentence is always taken to be an increment to a context that satisfies its presuppositions. If the current conversational context does not suffice, the listener is entitled and expected to extend it as required.[15]

For instance, if someone tells me: "Emar stopped smoking". If I did not know Emar was a smoker, I am to add this information to the conversational context or common ground. Such "extending" the common ground while processing sentences suggests a dynamic view of the common ground. There are several ways of modeling this. Altshuler and Maier use Discourse Representation Theory (DRT), so I will refer to it in my subsequent comment.[16]

The kind of accommodation involved in the phenomenon of imaginative resistance is a *narrator* accommodation. The point can be made in showing how it works on the Giselda story. While processing the first sentence of the Giselda story, the reader is expected to imagine that Giselda gave birth to her fourth child. In the processing of the second sentence, there is the jarring effect due to the fact that the reader resists imagining that it is fictional that Giselda did the right thing. In order to resolve the tension, Altshuler and Maier propose that the reader ascribes the jarring content to the narrator of the story, so as to get the distanced statement: "According to the (crazy) narrator of the story, Giselda did the right thing". In order to do this, they must make the narrator available so that they can ascribe it the jarring content.

[15] [Karttunen 1974], p. 191.
[16] See its original formulation in [Kamp and Reyle 1993].

This making salient of the narrator is an accommodation: the narrator was there all along in the background, only it is now explicitly put in the foreground. Just like "Emar used to smoke" was true all along in the background and becomes explicitly in the foreground when we process the sentence: "Emar stopped smoking".

In processing the first sentence, Altshuler and Maier hypothesised, as a default case, that the narrator be "neutral" or rather sharing our moral value(s) since no information about it was given. In the processing of the second sentence, the narrator proved to be very far from our moral value(s), hence the jarring effect. Technically, in DRT, accommodation consists in introducing a new discourse referent to denote the narrator (just like one would introduce a new discourse referent when coming across an indefinite description).

Following this picture, imaginative resistance consists in ascribing to a narrator the problematic evaluative judgement (for instance: "killing baby-girls because they are female is the right thing to do") or the problematic descriptive judgement (for instance: "maple leaf are oval-shaped"). This proposal thus treats in the same manner evaluative and descriptive imaginative resistance.

Narrator Accommodation and the GBD

I think Altshuler and Maier's idea is basically right. As hinted above, however, in the case of the GBD, the question is not to accommodate a *narrator*. The narrator merely says that Gregor has turned into "a vermin". The GBD is a debate between two readers. Extending Altshuler and Maier's idea to the GBD is thus not straightforward.

If one applies Altshuler and Maier's theory strictly to Nabokov's reading, one should make available the reader Smith as a discourse referent. Smith would thus be ascribed the problematic descriptive judgement "Gregor is a cockroach" that Nabokov resists to. But there is no reason to think that Smith is available as a discourse referent in Nabokov's reading of Kafka's text. Why would he?

The availability of a discourse referent for the narrator of a story is based on the fact that if a story is told, there must be a teller, even if merely in the background.[17] But Smith is just a reader. Consequently, if Smith is available as a discourse referent, then anyone could be and the number of discourse referents for a story will grow out of control. So it is not clear how Smith can enter the DRT analysis of imaginative resistance.

[17] Altshuler and Maier's account indeed implies that all stories have a narrator, since all stories, in principle, can yield imaginative resistance. This is not a problem given the theory of fictional discourse as a double-structured pretence defended in section 7.2.

I think the accommodation idea is relevant to the GBD. Indeed, we access some inferences which are, intuitively, of the same nature as those present in the imaginative resistance case. From the point of view of Nabokov, one can easily infer that Smith does not know much about cockroaches: indeed, he would not resist imagining a cockroach stuck on their back. This case is the same in the oval-shaped maple leaf example: if someone did not experience imaginative resistance in Yablo's scenario, one would immediately infer that they are not Canadian. In this sense, Smith betrays his lack of entomological knowledge, when he makes his reading overt.

Conversely, Nabokov betrays his own entomological knowledge as *basic* knowledge about the world. Hence, when Nabokov distances himself from Smith, he also betrays his desire to ridicule the inexperienced reader (implying something like "Smith does not know the basics"). This explains Nabokov's snobbish tone in this debate.

Interestingly, these kinds of inference generalise. A reader betrays a lot of themselves when they make their interpretation overt This comes from the fact that there are a lot of "gaps to fill" in imagination when one reads a fictional text. Crucially, they betray their background knowledge (or the lack thereof) about the world and their moral value(s) (or the lack thereof) as shown by this third-person perspective on imaginative resistance cases. Theorizing about imaginative resistance should thus naturally extend to these cases.

An outline of a general theory of "betrayal" can be found in [Walton 1990] (§10).[18] I suppose one way to go would be to say that accommodation is just one kind of betrayal. But betrayal is a broader notion. It extends to non-semantic, non-pragmatic content as in the case of the GBD where background information about the participants in the conversation become available. In the case of a fictional story, Altshuler and Maier convincingly argue that the narrator is always available as a discourse referent. I suggest that, given the availability of the narrator, one should also make available possible narratee. The "average reader" Smith is one salient possible narratee in this debate. This addition would explain the kinds of inferences one accesses when interpreting the GBD. But this implies a very liberal view of discourse referents which is largely unmotivated and hence problematic for DRT.

10.3 Conclusions

In this section, I have been exploring the idea that Nabokov and Smith are merely imagining past each other. This idea was prompted by the phenomenological data

[18]The context is very different though, for he tries to account for metafictional statements. For a discussion of these, see part 3 of this dissertation.

following which it is reasonable to think that the relevant difference between Nabokov and Smith in this debate is that Nabokov is an entomologist while Smith is not. This distinction was reflected at the level of semantic content, using the notion of a prototype. Indeed, if there are such things as prototypes, it must be that experts and neophytes do not share the same prototypes. This analysis of the GBD implies that Nabokov and Smith do not really disagree in any substantial manner: they merely imagine their own way.

This idea goes against the intuition that Nabokov and Smith are conflicting in some sense. In order to retrieve the conflict, I suggested we appeal to the notion of imaginative resistance. What would explain the conflict is that Nabokov resists imagining the way Smith does. But I showed that this line of thinking, though fruitful for the debate about imaginative resistance, goes against the promising analysis of imaginative resistance as narrator's accommodation. I suggested an extension of Altshuler and Maier's DRT analysis so that possible narratees can become available discourse referent. This extension is possible but implies a very liberal view of what counts as a discourse referent in DRT, which might be seen as a problem if it is not argued for independently.

In this chapter, I have taken a descriptive perspective on the GBD which is blind to the important distinction between what readers do (and can) in fact imagine and what is fictionally the case.[19] In other words, I focused on the empirical question: what *do* readers imagine when they read such a fictional text? But it is altogether another question to ask: what *should* readers imagine when they read such a text? I think the GBD essentially points towards the normative question, since Nabokov produces an argument in favour of his reading (and against Smith's). In other words, the GBD is not really a debate about the detail of Nabokov and Smith's psychology. This empirical investigation is merely a byproduct of the debate.

In the next chapter, I will focus on what is fictionally the case and put aside the empirical question. From a normative viewpoint, the GBD, far from being a case of imagining past each other, is a genuine disagreement about what is fictional in *the Metamorphosis*. Given our logical space, it is only natural to ask: is it a faulty disagreement or a faultless one? I will try to show that the GBD is a faultless disagreement.

[19]This is one of the basic distinctions between *de facto* and *de jure* imagination as discussed in subsection 5.2.1.

Chapter 11

The Great Beetle Debate as a faultless disagreement

As stressed above, Kölbel's distinction between faulty and faultless disagreements rests on the notion of *mistake*, or fault, or error. But this notion is not straightforward when we consider fiction reading. In the context of beliefs, a mistake has to to with *falsity*. But there is no *true* nor *false* imaginings; since imagination, by definition, does not necessarily aim at representing the world as it really is.

On the other hand, in the context of fiction, one has the intuition that some interpretations are *correct* and some *incorrect*: in other words, one can always *misimagine* what is fictional. In this chapter, I will show how one can construct the distinction and then apply it to Kölbel's definition of faultless disagreement. I think a good starting point is to consider how wrong, or incorrect, a reading can be.

11.1 Mistaken readers

11.1.1 Grading erroneous readers

In order to make a point, I will multiply readers.

$\mathbf{Smith_0}$

Suppose there is a reader called Smith_0. In reading the first sentence of *the Metamorphosis*, Smith_0 imagines that Gregor has turned into a monstrous *prime number*.

There is a clear sense in which Smith_0's reading is completely wrong. This is because a "prime number" does not fall into the familiar extension of "vermin", even

accompanied with the epithet "monstrous".

Although it is possible to imagine a far-fetched fictional scenario in which a prime number would be a vermin, it would be very difficult to argue that this far-fetched scenario is identical with *the Metamorphosis*.

Smith_1

Suppose there is a reader called Smith_1. In reading the first sentence of *the Metamorphosis*, Smith_1 imagines that Gregor has turned into a monstrous *rodent*.

In ordinary English, a rodent falls into the familiar extension of "vermin".[1] But there is a clear sense in which it is a mistake to imagine that Gregor has turned into a rodent. Indeed, the rest of the story mandates imagining some things which are clearly at odds with this interpretation: it is quite obvious that Gregor must be an insect, if we take into account the number of legs he has, the fact that he can walk on walls, and the kind of food he eats.

Although it is possible to imagine a far-fetched fictional scenario in which Gregor is a six-legged rodent able to walk on walls and eating what he eats, it would be very difficult to argue that this far-fetched scenario is identical with *the Metamorphosis*.

Smith_2

Suppose there is a reader called Smith_2. In reading the first sentence of *the Metamorphosis*, Smith_2 imagines that Gregor has turned into a monstrous *cyborg-pest*.

A cyborg is a creature which has both organic and mechanic body parts, they are nowhere to be found in the real world, but they are a classic kind of quasi-living-beings in many worlds of science-fiction since the 60s. If they were to be cyborgs, surely some of them would qualify as "vermin", in the sense that they would be as gross, as foul as our traditional organic vermin are. Smith_2's imagining would thus be Gregor turning into a monstrous one of them.

Of course, Smith_2 is more careful than the first two Smiths and he would imagine the right number of legs, as well as the walking on walls and the food eaten compatible with the cyborg-vermin features.

There is a sense in which Smith_2's reading is wrong, but I take it that this intuition is less robust than for the first two Smiths.

Since "cyborgs" were invented and popularised long after Kafka's death, it is very implausible that Kafka could have thought that the word "vermin" would designate

[1] Maybe the German *Ungeziefer* does not licence this interpretation; if not, please find an analogous example to make the point I am trying to make.

cyborg-pests. In this sense, Smith_2 is not *really* cooperating with Kafka's intentions: Smith_2's reading is essentially anachronistic.

I am not sure I can hold the view that *any* essentially anachronistic reading of a fictional text is wrong. Indeed, there is a well entrenched tradition of anachronistic interpretations of theater plays in Western countries, and I do not think this tradition is conceptually wrong or mistaken. But I am quite confident that one wants to keep a sharp distinction between anachronistic readings and non-anachronistic ones.

So maybe Smith_2 is not *wrong* in imagining Gregor turning into a cyborg-pest, but I take it that his being right is quite controversial.

Smith_3

Smith_3 is the Smith we already know, who imagines Gregor to have turned into a cockroach. As I suggested earlier, most readers would associate with Smith_3, this is why I called him a representative of the "average reader".

Aside: For those who have the intuition that this multiplication of indexed readers is somehow artificial, I have two things to say. First, it is merely a conceptual device to distinguish the different layers of meaning which are intertwined in the complex process of reading fiction. Second, think of some people who have a weak mastery of the language when they read *the Metamorphosis* in German for instance, and thus make clumsy mistakes: this should help to make these readers less artificial.

11.1.2 Why are the Smiths so wrong?

In order to define a notion of "mistaken reading", I need to apply the reality principles. A rough definition will suffice here. The reality principle says that when one reads a text of fiction, one should not in their interpretation depart from reality unless explicitly stated or needed. Note that the definition of the T-fantastic goes towards an even stricter formulation of a reality principle.

Smith_0

One can give a *semantic* story to explain Smith_0's mistake. As is commonly agreed, one tacit general rule for fictional imagination minimally requires that one imagines *what is said* in the text. There are two ways of making sense of Smith_0's reading as a correct one. But one goes against the tacit general rule, and the other goes against the reality principle.

One scenario would be that the expression "prime number" changes its ordinary meaning in the fiction to denote, say, a cockroach. This is surely possible to imagine such a fiction. And then Gregor would be a vermin, because he is a "prime number" in the fiction. This scenario goes against the general rule that words keep their ordinary meaning in fictional texts, unless there is an explicit mention to the contrary.

The other scenario would need some construction of a world in which prime numbers, while denoting the same set of abstract things as in our world, are vermin. For example, other numbers would be living peacefully in a mathematical realm, before the prime numbers invade this realm bringing lots of diseases and causing much damage to the live stocks of the numerous population. This is surely possible to imagine such a fiction. But it contradicts the "reality assumption" in many ways, and no explicit mention to go against the reality assumption is made in Kafka's text.

The same line of reasoning can be used to explain away a vast number of mistaken readings: Gregor turning into abstract objects, artefacts, non-living things, plants and non-vermin-like animals are immediately excluded in the same manner.

$\textbf{Smith}_1$

One can give a *pragmatic* story to explain Smith_1's mistake. Indeed, Smith_1 does not take into account all the information available in Kafka's text. Maybe some want to call it a *semantic* story, in that it focuses on the *inferences* drawn by the reader. I call it pragmatic, because it is not a mistake about *what is literally said*, but about what is *meant*. There are two ways of making sense of Smith_1's reading. One contradicts the reality assumption, and the other does not take into account all the information available in Kafka's text.

One scenario would be the far-fetched scenario in which a six-legged rodent can walk on walls and this conflict with the reality assumption following which rodents do not have six legs. Naturally, nothing in Kafka's text mandates the imagining of such a chimera. The other scenario consists in imagining an ordinary rodent, but it clashes with the information about the number of legs, the walking on walls, and the food eaten.

$\textbf{Smith}_2$

The story about Smith_2 is somewhat complex, as hinted above.

In order to "prove Smith_2 wrong" I think Nabokov would insist on the *genre* of the text. Imagining Gregor to have turned into a cyborg-pest tends to make *the Metamorphosis* become a science-fiction short story. And, presumably, all the atmosphere of the T-fantastic disappears. Of course, there is still a mystery: how

can a man turn into a cyborg? But once we have accepted cyborgs into the picture, then the explanation of this fictional event need not follow "the laws of reality" as we experience them. And then, why not explain this metamorphosis by a surgical intervention that happened the night before?[2]

One can see how unorthodox this interpretation would look like. This unorthodoxy coming from the violation of the *genre* of the story is not inaccurate as the interpretations violating the general rules of fiction. If the first two Smiths are *generally* incorrect because the violate some underlying general rules of fiction reading, then Smith_2 is *specifically* incorrect in that he merely violates a specific rule in force while reading *the Metamorphosis* as a T-fantastic story.

Smith_3

Nabokov argues that Smith_3 is wrong.

He has an argument: a cockroach is "flat on both sides and has long legs", and this is clearly incompatible with the first scene when Gregor has enormous difficulties getting out of his bed. Therefore, Gregor cannot be a cockroach. The closest insect one can think of is a beetle. Therefore Gregor is a beetle.

According to Nabokov, the reason why Smith_3 is wrong is very similar to the reason why Smith_1 is wrong. Based on the reality assumption and some entomological knowledge, constructing a fictional scenario in which a cockroach would have difficulties turning around is as far-fetched as imagining a six-legged rodent.

I think it is fair to say that there is something fishy about Nabokov's argument.

11.2 Saving Private Smith_3

> The Creator, if He exists, has a special preference for beetles.
>
> J.B.S. Haldane (This is referring to the fact that 25% of all know animals species are types of beetle.)

I do not think $\text{Smith}_{0,1}$ can be rescued in any way. Rescuing Smith_2 is a project for the literary theorist, since it has to do with orthodox and non-orthodox readings. This greatly surpasses my ability in literary theory. Rescuing Smith_3 is, I should claim, a philosophical enterprise: it will serve to establish the GBD as a faultless disagreement. I can see four ways of doing so.

[2] Actually, that would explain the "agitated dreams" which are probably due to the anaesthetic...

11.2.1 The way of the psychologist

Nabokov says Gregor cannot be a cockroach. But maybe Gregor is a very "stupid cockroach", or at least a strikingly clumsy one: he does not even know he can turn around using his long legs.[3]

Gregor clearly seems to have a human psychology in Kafka's story. Indeed, his dehumanisation is very progressive in the story. In the first scene, he *has* human urges: he wants to talk, he wants to open the door and explain why he is late for work, he can feel the shame and many other human feelings. So it seems possible he has not yet taken up all the reflexes cockroaches have, just as he has not given up all of his human urges.

Now, Nabokov has to tell us more about the psychology of cockroaches and beetles. Maybe he would have had something to say about it, had he been here to help me write this.

11.2.2 The way of the sceptic

Maybe Gregor is a yet unknown insect closer to a cockroach than a beetle, unable to turn around so easily.[4]

Entomological discoveries sometimes happen. Here is one relevant such discovery: The dictyoptera were a family of insects containing the blattodea and the mantodae. The various species of cockroaches are part of the blattodea, and the English common noun "(cock)roaches" is usually used to denote the entire family of the blattodea. Mantodae is the closest family of insects to the blattodea, and the most famous representative of this family are preying mantids. For a long time, this classification of the dictyoptera was thought to be exhaustive.

But, DNA analysis performed in the first half of the 20th century, and then additional phylogenetic evidence in the 60s put the termites into a subfamily of the blattodea, against the former hypothesis that they form a family of their own. It is now acknowledged that termites form a subspecies of social, wood-eating cockroaches; their common ancestor with the cockroaches is estimated to have lived during the Jurassic or Triassic, that is roughly from 150 to 250 million years ago.[5]

[3]Thanks to Markus Gabriel for this suggestion.

[4]Thanks to Paul Égré for this suggestion.

[5]See [Capinera 2008] entry "Cockroaches (Blattodea)". Here is one of the introductory paragraph:

> The cockroaches (also called roaches) are members of an ancient order of insect. The order name is derived from the Latin word *blatta*, or cockroach. They are closely related to the preying mantids (Mantodea), and often grouped with them (as suborder)



(a) Diploptera punctata (the Pacific beetle cockroach) – mimic of brown beetles



(b) Prosoplecta (cockroaches) – mimic of ladybird beetles

Figure 11.1: Two spieces of cockroaches which mimic beetles

I don't know how much Nabokov knew about all this, but it shows that scientific discoveries can change the picture substantively. And the scientific discovery which led to classify termites as cockroaches was not helped by a scrutiny of natural language.

Now, suppose entomologists found a new kind of cockroach in the future, with smaller legs and rounder body. Then Nabokov's argument would be made obsolete, and Smith_3 would, eventually, be correct. This puzzling relativity about what is fictional in *the Metamorphosis* seems to follow from Nabokov's rigorous stance.

One might thus adopt a sceptical position. We probably cannot know precisely the ultimate nature of Gregor, so that it is more rational to accept *both* Smith_3 and Nabokov's interpretation in the light of available "evidence".

Here is another interesting piece of entomological knowledge to strengthen this sceptic line of response: Some species of cockroaches evolved to mimic some species of beetle. The *Pacific Beetle Cockroach* mimics the common brown beetle. The *Prosoplecta* is a cockroach which mimics the ladybird. See Figure 11.1 for photographs of these kind of cockroaches. The world of insects would make us all sceptic, if only we knew more about it...

to form the order Dictyoptera. Though the mantids evolved from the cockroaches, they are a specialised group of predatory insects that warrant individual recognition. Termites (Isoptera) can also be placed in the order Dictyoptera, and are considered by some to be social cockroaches. The order name for cockroaches is given as Blattaria.

11.2.3 The way of the ignorant

Maybe it is fictional that Gregor is a beetle, maybe it is fictional that he is a cockroach, but it does not really matter for the reader to know.

Nabokov's implicit premise is to say that entomology can apply as it stands in the real world to the world of *the Metamorphosis*. So he suggests that an entomologist reader is more competent than a non-entomologist reader, when it comes to determining what is fictional about Gregor. This is quite objectionable on two grounds.

First, it is not even clear that Kafka himself had this sort of knowledge, so that Nabokov would have to defend the thesis that the reader should be more knowledgable than the author, making the author himself a poor reader of his own work. Maybe this line of defense is possible, but it is surely counter-intuitive.

Second, the amount of knowledge necessary to read all the fictions we read will inflate without control. It follows that most readers are incompetent readers; and probably, for reasons of time and interest, most competent readers of fiction do not read fiction. There are some good reasons to think that Nabokov agrees with this. I am not claiming that it cannot be defended on rational grounds, but surely it is a very counter-intuitive view which one might want to avoid.

Ignorance is bliss, when it comes to fiction reading. The way of the ignorant stresses this point in arguing that the practice of fiction reading is not the same as the practice of entomology or any other scientific investigations. Indeed, there is much more leeway in imagining a world of fiction than there is in the scientific description of nature; and this is what the disagreement between Smith_3 and Nabokov shows.

11.2.4 The way of the erudite person

Erudite people know of [Haldane 1926]. Haldane was a biologist. In 1926, he published an article entitled "On Being the Right Size" in which he argues that there is a "right size" for all kinds of living creatures, so to speak. He shows that the size of an animal is a zoological feature which obeys the same laws of evolution than the other biological features; this fact, Haldane remarks, we tend to overlook, making as if the size of an animal did not matter when it comes to zoology. Interestingly, Haldane illustrates his point by appealing to creatures of fiction. Here is the opening of his article:

> The most obvious differences between different animals are differences of size, but for some reason the zoologists have paid singularly little

> attention to them. In a large textbook of zoology before me I find no indication that the eagle is larger than the sparrow, or the hippopotamus bigger than the hare, though some grudging admissions are made in the case of the mouse and the whale. But yet it is easy to show that a hare could not be as large as a hippopotamus or a whale as small as a herring. For every type of animal there is a most convenient size, and a large change in size inevitably carries with it a change of form.
>
> Let us take the most obvious of possible cases, and consider a giant man sixty feet high - about the height of Giant Pope and Giant Pagan in the illustrated Pilgrim's Progress of my childhood. These monsters were not only ten times as high as Christian, but ten times as wide and ten times as thick, so that their total weight was a thousand times his, or about eighty to ninety tons. Unfortunately the cross sections of their bones were only a hundred times those of Christian, so that every square inch of giant bone had to support ten times the weight borne by a square inch of human bone. As the human thigh-bone breaks under about ten times the human weight, Pope and Pagan would have broken their thighs every time they took a step. This was doubtless why they were sitting down in the picture I remember. But it lessens ones respect for Christian and Jack the Giant Killer.[6]

What about insects? Interestingly, the size of insects is not limited by gravity. But the main constraint on the size of insects comes from the way they get the oxygen from the atmosphere into their bodies. As Haldane remarks:

> [...] while vertebrates carry the oxygen from the gills or lungs all over the body in the blood, insects take air directly to every part of their body by tiny blind tubes called tracheae which open to the surface at many different points. Now, although their breathing movements can renew the air in the outer part of the tracheal system, the oxygen has to penetrate the finer branches by means of diffusion. Gases can diffuse easily through very small distances, not many times larger than the average length traveled by a gas molecule between collisions with other molecules. But when such vast journeys – from the point of view of a molecule – as a quarter of an inch have to be made, the process becomes slow. So the portions of an insect's body more than a quarter of an inch from the air would always be short of oxygen. In consequence hardly any insects are much

[6] [Haldane 1926], p. 1.

more than half an inch thick.[7]

One can now see how Gregor would have trouble breathing to the point of asphyxia in a world where zoology was really the same...

Nevertheless, there are well-known cases of insect gigantism. In the Carboniferous (around 300 million years ago), insects were up to 10 times larger than the ones we can observe today. Some fossils have shown for instance that the Meganeura (a dragonfly-like insect) had a wingspan of about 70 centimeters; other fossils show that the Arthropeura (a millipede-like arthropod) was up to 2.6-meter-long. The main hypothesis for such a gigantism, in keeping with [Haldane 1926], is that the atmosphere in the Carboniferous contained about twice as much oxygen as today's one. Of course, there are debates about other possible causes of this gigantism which are nicely reviewed in [Harrison, Kaiser, and VandenBrooks 2010].

Perhaps we should imagine that Gregor's world has an atmosphere containing around 40% of oxygen so that we can justify his size. But if that were so, then Gregor's parents would be in bad shape: oxygen in such proportion is very toxic for humans; they would suffer severe dazzling and spasmodic vomiting.

At this point, I am not sure it is Nabokov's interest to claim that entomology should apply *as it really is* in order to justify his reading against Smith_3's...

11.3 Faultless disagreements and literary disagreements

11.3.1 The GBD is faultless

I gave a relative definition of what a mistake or a fault is in the context of fiction reading. According to this criterion, Nabokov and Smith_3 are not at fault, contrary to $\text{Smith}_{0,1}$. We saw that the case for Smith_2 is trickier which suggests that the distinction between correct and incorrect reading should be seen as a continuum between two extremes rather than a binary distinction.

Everything is in place for Kölbel's definition to apply, then. There is a reader Nabokov, a reader Smith_3 and a proposition expressed by the sentence: "Gregor Samsa has turned into a cockroach". Smith_3 imagines this proposition while Nabokov imagines the negation of this proposition, both are reading the same text, namely *the Metamorphosis*. Neither Smith_3 nor Nabokov is at fault, i.e. both their reading are *correct*.

As a result, the GBD is a faultless disagreement.

[7] [Haldane 1926], p. 3.

11.3.2 Interpretation and elucidation

I just defended that the GBD is a faultless disagreement. I think this result entails a thesis I will call *pluralism* about interpretation. *Pluralism* is the view that there are several, incompatible interpretations of a text of fiction. This view is controversial in literary criticism. First and foremost; I must clarify what I mean by "interpretation" here, since it may be subject to dangerous equivocations in the context of literary criticism.

Elucidation

In [Friend 2017a] one can find important distinctions. There are at least three kinds of interrelated interpretative activities in which literary critics are well-versed. One is *elucidation*: it consists in making explicit what is merely implicit in the fiction. The second is *explication*: it consists in "ascertaining the meanings and connotations of words, or passages" of a fiction. The third is *thematic interpretation*: it consists in "identifying the themes and theses in the work as a whole".[8]

Here is Friend's definition of elucidation, in more precise terms:

> To elucidate a work is to determine what is going on in the storyworld, what is "true in the story" – or as I prefer, *storified* – where this is not specified by the explicit text and may even contradict it (as with unreliable narrators).[9]

The GBD is thus a paradigmatic instance of a debate about *elucidation*.

Trivial and substantive elucidations

It is interesting to distinguish between two kinds of elucidations: the "trivial" kind and the "substantive" kind, so to speak.

Trivial elucidation is a case where what is elucidated has no bearing upon other kinds of interpretative activities. Literary critics are usually not interested in boring elucidation, although philosophers love it. To use a famous example from [Lewis 1978], it must be true in *the Adventures of Sherlock Holmes* that Sherlock Holmes lacks a third nostril and has never visited Saturn's moons. Since this is not explicitly stated in Conan Doyle's text, we get these fictional propositions by elucidation. But such implicit fictional proposition about Sherlock Holmes has no particular bearing on explication of thematic interpretation.

[8] [Friend 2017a], p. 388. Friend's terminology is adapted from [Beardsley 1958].
[9] [Friend 2017a], II, p. 388-9.

Substantive elucidation, by contrast, is a case where what is elucidated has crucial consequences which feed into the other interpretative activities. Here are some examples taken from Beardsley:

> Is Hamlet mad? What is Raskolnikov's real motive in killing the old woman? Where is the speaker and what is his situation in "Gerontion" or "Sailing to Byzantium"? What traits are basic to the character of Antigone?[10]

Most philosophers of fiction avoid these substantive cases of elucidation, because the answers to such questions are usually controversy-ridden and going into such literary controversies would drive them away from making a point about fictional "truth" in general.

Friend goes a step further, for she argues that some cases of elucidation are not so much essential for explication and thematic interpretation, but also for the mere *understanding* of the fiction:

> Many puzzling works demand efforts at elucidation. Anyone who fails to wonder who Godot is and why Vladimir and Estragon awaits him, or who is unperturbed by Bartleby's intransigence has simply not engaged with the relevant works.
>
> [Footnote: As these cases indicate, elucidation cannot always sharply be distinguished from other dimensions of interpretation or criticism.][11]

Controversial and non-controversial elucidations

This points to another distinction that Friend considers. It consists in wondering whether elucidation is controversial or not.

> Elucidation is more interesting when it seeks the unobvious story-truths, the ones that generate disagreement.
>
> [Footnote: The term *interpretation* is sometimes reserved for this kind of elucidation, contrasted with uncontroversial *description*. See, for example [Lamarque 2008], p. 152.][12]

This distinction is in principle orthogonal to the first one. Indeed, one can have disagreements which have no consequence whatsoever on the other interpretative

[10] [Beardsley 1958], p. 242.
[11] [Friend 2017a], p. 389.
[12] [Friend 2017a], p. 389.

activities. For example, the unspecified hair colour scenario I presented in the beginning of this chapter may have no consequence whatsoever. It is easy to imagine a story in which being dark-haired of auburn makes no difference whatsoever when it comes to explication or thematic interpretation. Still, one can disagree about it. So this is a trivial, controversial case of elucidation.

Of course, many substantive elucidations are also controversial. One can, for instance, disagree over the case of Hamlet's being mad.

Some substantive elucidations, on the other hand, are non-controversial. One can wonder about who Godot is without having any competing elucidation in mind.

What about the GBD?

11.3.3 The GBD as an interesting case of elucidation

The GBD is, by definition, a controversial elucidation, since the fictional detail subject to elucidation generates a disagreement. Is it a trivial kind of elucidation?

At first sight, one might want to say "yes".[13] But, one might argue, following Nabokov's example, that it is substantial. Indeed, Nabokov develops the elucidation about Gregor's insecthood into both an original explication and thematic interpretation. His explication is a positive proposal, while his thematic interpretation divides into a polemical and a positive proposal.

Bearing of the GBD on the explication of *the Metamorphosis*: Here is Nabokov's explication:

> [Gregor] has a tremendous convex belly divided into segments and a hard rounded back suggestive of wing cases. In beetles these cases conceal flimsy little wings that can be expanded and then may carry the beetle for miles and miles in a blundering flight. Curiously enough, Gregor the beetle never found out that had wings under the hard covering of his back. (This is a very nice observation on my part to be treasured all your lives. Some Gregors, some Joes and Janes, do not know that they have wings.)[14]

So, according to Nabokov, the fact that Gregor is a beetle allows one to conclude that he has wings.[15] Consequently, Gregor's failing to discover that he has wings

[13] In conferences, when discussing the GBD, I was several times said: "Is this *really* interesting?" To which I usually answered: "Thanks for suggesting everything I said is just uninteresting!"

[14] [Nabokov 1980], p. 259.

[15] Note that most cockroaches have wings, too. I suppose Nabokov has in mind german cockroaches (*Blattella germanica*) which, although they have wings, do not fly.

makes him a pitiful, almost tragic character. This is an explication of Kafka's text for it "ascertains the meaning and connotation" of the fictional metamorphosis.

Bearing of the GBD on the thematic interpretation of *the Metamorphosis*: Nabokov first argues against existing psychoanalytical interpretations of the work on the ground that these critics never ask themselves what kind of insect Gregor is. According to Nabokov, such critics merely ramble over the unconscious symbolism of the "vermin" Gregor became instead of taking a naturalistic stance on the fictional event:

> Freudian [critics], like Neider in *The Frozen Sea* (1948), contend, for example, that "The Metamorphosis" has a basis in Kafka's complex relationship with his father and his lifelong sense of guilt; they contend further that in mythical symbolism children are represented by vermin – which I doubt – and then go on to say that Kafka uses the symbol of the bug to represent the son according to these Freudian postulates. The bug, they say, aptly characterises his sense of worthlessness before his father. I am interested here in bugs, not in humbugs, and I reject this nonsense.[16]

Later on, Nabokov offers a positive thematic interpretation of Kafka's short story, it is delightfully naturalistic:

> Let us look closer at the transformation. The change, though shocking and striking, is not quite so odd as might be assumed at first glance. A commonsensical commentator (Paul L. Landsberg in *The Kafka Problem* [1946], ed. Angel Flores) notes that "When we go to bed in unfamiliar surroundings, we are apt to have a moment of bewilderment upon awakening, a sudden sense of unreality, and this experience must occur over and over again in the life of a commercial traveller, a manner of living that renders impossible any sense of continuity." The sense of reality depends upon continuity, upon duration. After all, awakening as an insect is not much different from awakening, as Napoleon or George Washington. (I knew a man who awoke as the Emperor of Brazil.) On the other hand, the isolation, and the strangeness, of so-called reality – this is, after all, something which constantly characterises the artist, the genius, the discoverer. The Samsa family around the fantastic insect is nothing else than mediocrity surrounding genius.[17]

[16][Nabokov 1980], pp. 255-6.
[17][Nabokov 1980], p. 260.

So the theme of Kafka's story would be this feeling of alienation which is common both to the travelling salesman who wakes up in unfamiliar surroundings (Gregor is a travelling salesman) *and* to the brilliant artist who comes from a philistine family. I let this insightful, ironical piece of criticism to the literary critic's judgement!

Conclusion: *Prima facie*, Gregor's insecthood could pass for a detail grounding a trivial kind of elucidation. But Nabokov connects it with explication and thematic interpretation, so that the GBD should rather be seen as a debate over an substantive case of elucidation. Consequently, one should not be too prompt to judge what counts as trivial or not. I suppose this shows that, in literary criticism, the devil is in the fictional detail.

11.3.4 Pluralism about elucidation

Here is, finally, an argument in favour of pluralism for elucidation which relies on the premise that the GBD is faultless.

There are, in principle, three possible theses about elucidation. There is *absolutism* according to which for all cases of elucidation, there is only one correct answer. There is *relativism* according to which for all cases of elucidation, there are several correct answers. There is *pluralism* according to which for some cases of elucidation, there are several correct answers (and for others, only one).

Given that the GBD is a faultless disagreement, absolutism is false. Relativism entails that there are, strictly speaking, no case of incorrect reading. Given that $\text{Smith}_{0,1}$ are incorrect, relativism is false. Indeed, whether Gregor is an insect or not is a case of elucidation. But this case of elucidation admits only one answer: Gregor is an insect. Therefore, pluralism about elucidation is the only live option.

Note that pluralism about elucidation does not entail pluralism about the other kinds of interpretative activities. However, it may be seen as a good reason to adopt pluralism all the way up.

This pluralism is argued for theoretically. In practice, one should want to discuss the width of the spectrum for correct elucidations. In the case of *the Metamorphosis*, we have considered two correct elucidations (Nabokov and Smith_3) and two incorrect elucidations ($\text{Smith}_{0,1}$); Smith $_2$ is a case for discussion. So I suggested one should consider a continuum of possible elucidations.

The anachronistic reading is merely a toy example. For more relevant debates about the acceptable width of elucidations, one should consider post-colonial studies or feminist studies of literature. Another interesting place to look at are the definition of a "canon" in fan-fictions, as discussed in chapter 8.

Pluralism also raises several theoretical problems. There is a problem about the relation between free will and imagination: how much is it possible to imagine what one wants? There is a problem about what constitutes the "non-negotiable" core of fictional propositions: this leads back to the problem of "truth in fiction". There is a problem about the shareability and commensurability of incompatible elucidations: do Nabokov and Smith_3 imagine *the same* Gregor?

I think the common underlying phenomenon of these philosophical question is the notion of *indeterminacy*. As Friend puts it: "That said, elucidation leaves a great deal indeterminate." In the next chapter, I will try to determine the origin of this indeterminacy. I will argue that such indeterminacy is the hallmark of fiction as opposed to non-fiction.

Chapter 12

The fundamental indeterminacy of fiction

A tempting interpretation of what has been said so far would be the following: Nabokov and Smith appear to disagree on a fictional event (viz. the kind of insect Gregor *fictionally is*); this is due to the fact that they interpret the text their own way, thanks to pluralism about elucidation; and the reason why they can do that is because *the Metamorphosis* is fiction.

There is certainly a clear intuition that this pluralism is a feature of fiction as opposed to non-fiction: in a newspaper article, or a medical report, even if the eye colour of some protagonist is not mentioned, it has to be *either* blue *or* brown (or some other possible colour), and this fact cannot depend on the reader. In fiction, pluralism entails that such fictional event *can* depend on the reader. In other words, Nabokov's and Smith_3's faultless disagreement would be but a byproduct of the fact that *the Metamorphosis* is a fiction.

This amounts to saying that what is fictionally represented is *in principle* indeterminate, uncertain, vague in a way real objects are not. Let us call this the "fictional uncertainty principle". If one wants to extend the metaphor, then one would say that fiction is like quantum mechanics where uncertainty is the rule. In this chapter, I will thus look for the fundamental inequalities which frame this notion into a meaningful picture.

I will first show that the notion of indeterminacy under discussion here is not linguistic in nature, nor is it ontological. Then, I will try to show that it comes from the pragmatic nature of the reality principle. The fundamental indeterminacy of conventions, interestingly, was a long-standing idea in Quine's philosophy. I will thus connect my analysis of the GBD with some of Quine's analysis of modality.

12.1 Friend on indeterminacy

Interestingly, contrary to what I have been arguing for, Friend presupposes that Smith and Nabokov have a faulty disagreement. She assumes that Nabokov is right and Smith is wrong, and that Nabokov has given convincing arguments in favor of his position.[1] As a matter of fact, she is not really concerned about the specifics of this debate, for she is investigating the phenomenon of co-reference with empty singular terms. The debate is only a case in point where Nabokov and Smith are both imagining the *same* thing (proof: they disagree). In my terms, the problem Friend wants to solve does not rely on the answer one wants to give to Q2, but on the fact that *there is* an answer to it. Here is Friend's paragraph on which I ground this interpretation of her intentions:

> Let's assume that Nabokov is right about his interpretation of the story, and thus about what the story prescribes imagining. Then Nabokov has said something true, while Smith has said something false. More importantly, Nabokov and Smith have disagreed: they are making incompatible claims *about the same thing*, even though there is no thing they are talking about. Nabokov claims that Kafka's story prescribes imagining that Gregor has been changed into one kind of insect, Smith that Gregor – the very same character – has been changed into a different kind of insect. We want a theory to explain the possibility of *intersubjective identification*: the phenomenon of thinking and talking about the same thing, even when there is no such thing.[2]

At this point, Friend considers and quickly dismisses a possible objection: it concerns the "indeterminacy" of Gregor; and she dismisses the point by appealing to the distinction between *indeterminacy* and *underdeterminacy*. Let me quote her again. I will then try to make explicit her argument, and explain why I think the argument does not work.

In her footnote 7, Friend writes:

> While Nabokov makes a good argument, there is no fact of the matter whether we are supposed to imagine that Gregor becomes a beetle or any other kind of vermin. This indeterminacy does not indicate anything mysterious about the nature of fictional characters, but is rather a result

[1] It does not matter for her argument to know *who* is right, it matters that they are not *both* right.

[2] [Friend 2011], pp. 187-188.

> of the fact that there is a limit to the amount of information a work of fiction gives us in prescribing imaginings.[3]

The indeterminacy Friend has in mind is most probably epistemic, in that she sees the cause of it in the essentially finite amount of information a text can give. I think that the phenomenon she points out is the semantic *vagueness* of the word "vermin": the idea is that the vagueness of the description *in the text* does not ground in any sense the ontological vagueness of Gregor *in the fiction*. The underlying argument is that the vagueness of a description *never* entitles someone to infer the vagueness of the thing described.[4] This sounds true: denying this would be to commit oneself to a generalised ontological vagueness (as is commonly agreed that most natural language descriptions are vague); for instance to the ontological vagueness of the wife of Socrates (since the only information we have about her are textual descriptions); but this is absurd since the wife of Socrates was no more vague than you and I.[5] Therefore, the vagueness of Gregor's description does not entitle anyone to infer the vagueness of Gregor. Let us call this argument the *argument from epistemology*.

I think this argument does not go through as it is. The reason why it does not work is that Gregor is a fictional character, and, arguably, it is possible for a fictional character to be vague, or indeterminate *in its fictional world*. In fact, the reason why Gregor is not vague, or indeterminate, is because there are no such things as "indeterminate objects" *in the world of* the Metamorphosis.

There is nothing wrong with the argument from epistemology in real contexts, but the argument has to be blocked if we consider fictional contexts. If in some fictional worlds some of the described characters are indeterminate, then an extra-premise should be added to the argument from epistemology for it to be valid in general. Namely something like: "and the relevant fictional contexts under discussion do not entitle anyone to imagine indeterminate characters in their own fictional world".

In the case of Gregor and *the Metamorphosis*, it is plausible that this extra-premise is true. But it has to be argued for. I will come to this argument later. First I need to argue for the claim which puts a severe logical constraint on the generality of the argument from epistemology. That is, I have to say more about these fictional worlds in which indeterminate characters dwell.

[3] [Friend 2011], p. 187.

[4] This argument is well argued for in [Russell 1923].

[5] Some philosophers have argued for a generalised ontological vagueness, but the argument for such a position does not appeal to the vagueness of natural language descriptions. See for example [Unger 1980].

12.2 Ontologically bizarre characters

12.2.1 Definition

In [Everett 2005], there is a paragraph on "ontic indeterminacy". He writes a story with two characters, Frick and Frack, according to which it is impossible to determine whether Frick and Frack are different or one and the same character. Frick and Frack are ontically (that is at the level of identity conditions) indeterminate. Here is Everett's paragraph:

> Now it is widely recognised that if there are fictional objects then it will at least sometimes be a vague or indeterminate matter as to whether two fictional objects are identical.
> [...]
> A story might describe two characters in such a way that it is left indeterminate whether or not they are identical:
>
> **Frackworld:** No one was absolutely sure whether Frick and Frack were really the same person or not. Some said that they were definitely two different people. True, they looked very much alike, but they had been seen in different places at the same time. Others claimed that such cases were merely an elaborate hoax and that Frick had been seen changing his clothes and wig to, as it were, become Frack. All that I can say for certain is that there were some very odd similarities between Frick and Frack but also some striking differences.
>
> I think it is pretty clear that in this story it is left indeterminate as to whether Frick is Frack.[6]

In Everett's story, there is a prescription to imagine that either Frick and Frack are two different persons or that they are only one. The identity condition of, say, Frack are indeterminate. I think the natural reading of this story is that *either* "Frick is Frack", *or* "Frick and Frack are two different people" *in Frackworld*. But we can force the unnatural reading following which *both* "Frick is Frack" *and* "Frick and Frack are two different people" are fictional in Frackworld. This reading makes Frackworld a world very different from ours: it is a world where it is indeterminate whether some people are one or two. In our world, thanks to numerical identity, one person can only be one.

[6] [Everett 2005], pp. 628-9.

This example aims at showing that any of the "laws of reality" can be altered in a fictional world. Even some laws about the metaphysical structure of the world. In such fictional worlds dwell "ontologically bizarre characters". *Ontologically bizarre characters* are fictional characters which contravene the ontological principles in force in the real world. There are, by definition, *impossible* objects.

12.2.2 More examples of ontologically bizarre characters

The example taken from Everett above might strike you, reader, as artificial. Frick and Frack is (are) not (a) fictional character(s) in the ordinary sense of the expression, they are rather characters made up by a philosopher of fiction who needed a clear-cut counter-example to oppose other philosophers of fiction. Actually, this prejudice in favor of the literary is unfair. For fictional authors are arguably as twisted as philosophers, when it comes to inventing ontologically bizarre characters. Here is a list of some of them taken from different kinds of fiction.

Nabokov's (T)wynn

Here is a page from Nabokov's novel *Pnin* published in 1957; it strangely resembles Everett's story to the point where maybe one ought to consider them *one and the same* story...

> Pnin and I had long since accepted the disturbing but seldom discussed fact that on any given college staff one could find not only a person who was uncommonly like one's dentist or the local postmaster, but also a person who had a twin within the same professional group. I know, indeed, of a case of triplets at a comparatively small college where, according to its sharp-eyed president, Frank Reade, the radix of the troika was, absurdly enough, myself; and I recall the late Olga Krotki once telling me that among the fifty or so faculty members of a wartime Intensive Language School, at which the poor, one-lunged lady had to teach Lethean and Fenugreek, there was as many as six Pnins, beside the genuine and, to me, unique article. It should not be deemed surprising, therefore, that even Pnin, not a very observant man in everyday life, could not help becoming aware (sometime during his ninth year at Waindell) that a lanky, bespectacled old fellow with scholarly strands of steel-gray hair falling over the right side of his small but corrugated brow, and with a deep furrow descending from each side of his sharp nose to each corner of his long upper-lip – a person to whom Pnin knew as

> Professor Thomas Wynn, Head of the Ornithology Department, having once talked to him at some party about gay golden orioles, melancholy cuckoos, and other Russian countryside birds – was not always Professor Wynn. At times, he graded, as it were, into somebody else, whom Pnin did not know by name but whom he classified, with a bright foreigner's fondness for puns as "Twynn" (or, in Pninian, "Tvin"). My friend and compatriot soon realised that he could never be sure whether the owlish, rapidly stalking gentleman, whose path he would cross every other day at different point of progress, between office and classroom, between classroom and stairs, between drinking fountain and lavatory, was really his chance acquaintance, the ornithologist, whom he felt bound to greet in passing, or the Wynn-like stranger, who acknowledged that somber salute with exactely the same degree of automatic politeness as any chance acquaintance would. The moment of meeting would be very brief, since both Pnin and Wynn (or Twynn) walked fast; and sometimes Pnin, in order to avoid the exchange of urbane barks, would feign reading a letter on the run, or would manage to dodge his rapidly advancing colleague and tormentor by swerving into a stairway and continuing along a lower-floor corridor; but no sooner had he begun to rejoice in the smartness of the device than upon using it one day he almost collided with Tvin (or Vin) pounding along the subjacent passage. When the new Fall Term (Pnin's tenth) began, the nuisance was aggravated by the fact that Pnin's class hours had been changed, thus abolishing certain trends on which he had been learning to rely in his effort to elude Wynn and Wynn's simulator. It seemed he would have to endure it always. For recalling certain duplications in the past – disconcerting likenesses he alone had seen – bothered Pnin told himself it would be useless to ask anybody's assistance in unraveling the T. Wynns.[7]

Calvino's Knight

Italo Calvino wrote a short novel called *the Nonexistent Knight* in which the protagonist is a nonexistent Knight, working in the army of Charlemagne, king of the Francs. He his but an empty armour. He is characterised among the other soldiers as being an over-zealous spoilsports. A woman soldier of the same army is in love with him, and wishes he was existent, for he would have a body which she would long for.

[7] Nabokov 1957 *Pnin* (§5.5), p. 148.

You can see how this character has a strange ontological status within Calvino's novel, and this is surely part of the fun.

Woody Allen's Harry

In *Deconstructing Harry* by Woody Allen, the character played by Woody Allen suffers from a strange illness: he suddenly becomes out of focus. It is clear that he is out of focus not only for the spectator, but also in the movie, since the other characters see him out of focus. As a consequence of this, he looses his job as a film actor for obvious reasons, and have much trouble with his wife and kids who get terrible headaches interacting with him. After he understands the psycho-somatic causes of this strange illness, everything comes back to normal and, towards the end of the movie, he is not out of focus anymore.

Borges' Aleph

In a short story entitled *Aleph*, Borges writes about a strange object called an "aleph". One of the main characters has an aleph in the basement, and invites another character to go have a look at it. The aleph is a small sphere lodged under the staircase. It contains the whole universe, and every single detail of the universe is there in front of the viewer. This aleph is surely a marvelous fictional object since it is strictly part of its fictional world and the world it is in is strictly contained in it.

Sylvan's box

This example was discussed in subsection 3.3.4. Graham Priest writes about an incredible box which is both full and empty at the same time. Graham imagines what would be his reaction to such an object, being one of the leading dialethists, he should be happy to discover an object which proves his metaphysical position right.[8] But the object is strangely disturbing to him: what would be the political consequence of his disclosing the news? The short story ends with ten morals to be drawn for the philosophy of fiction.

The alchemist's lab

In *One hundred years of solitude*, Gabriel Garcia Marquez writes about the twists and turns of the passing of time. Towards the end of the novel, the last of the Aurelianos

[8]Dialethism is the claim that some contradictions are true *in the actual world*, i.e. that the real world is an impossible world.

discovers a strange truth about the alchemist's lab where José Arcadio Buendia has withdrawn for a very, very long time, now being considered a crazy person. In this room, time stagnates. Here is the explanation of this strange phenomenon:

> In the small isolated room where the arid air never penetrated, nor the dust, nor the heat, both had the atavistic vision of an old man, his back to the window, wearing a hat with a brim like the wings of a crow who spoke about the world many years before they had been born. Both described at the same time how it was always March there and always Monday, and then they understood that José Arcadio Buendía was not as crazy as the family said, but that he was the only one who had enough lucidity to sense the truth of the fact that time also stumbled and had accidents and could therefore splinter and leave an eternalised fragment in a room.[9]

Conclusions

Ontologically bizarre characters are legion. Note that I have been using the term "character" to metonymically denote fictional *objects*. There are, to my knowledge, no exhaustive typology of ontologically bizarre characters, although many philosophers of fiction have their favourite. The closest to the attempt is [Borges 1967]. Although most of the imaginary beings listed in this book, strictly speaking, *ontologically* impossible.

12.3 On the indeterminacy of Gregor

The diversity of these examples show that *ontologically bizarre characters* are a broad category. I will now argue that Gregor Samsa is not one of them.

Gregor Samsa is not ontologically bizarre the way Calvino's nonexistent knight is, he is not ontologically vague, nor is he contradictory, or indeterminate in an ontic sense. Gregor is arguably *biologically* bizarre, in the sense that he is an insect the size of a middle-sized dog or so; and maybe insects, because of the way they are organised, cannot be that big.[10] This puts Gregor closer to a kind of chimera, or an imagined animal. But this is not *ontologically* bizarre.

Moreover, if *the Metamorphosis* is puzzling for the reader, this is precisely because Gregor is not ontologically bizarre. Indeed, suppose Gregor was like Calvino's knight,

[9] [Garcia Marquez 1967], p. 276.
[10] See [Haldane 1926].

a blatantly ontologically bizarre character. Then the reader would know from the start that the fictional world Gregor lives in is very different from our own. The "laws of this world" are nothing like the "laws of reality". Arguably, this would be much less puzzling and would trigger the kind of reaction we have when imagining something marvelous or magical; be it fear or laughter, but not unease. The puzzle comes from the conflict between: (i) the anomaly which is Gregor's metamorphosis; (ii) the fact that everything else seems familiar. If Gregor's world was utterly different, then maybe his metamorphosis would not be that puzzling anomaly. That Gregor is not ontologically bizarre is a condition of possibility of this feeling of unease the reader has in reading *the Metamorphosis*.[11]

Yet, there is a sense in which Gregor is indeterminate. Indeed, he is entomologically indeterminate. Is he a cockroach? Is he a beetle? Well both, or neither: it is indeterminate. Because both elucidations are correct.

But this does not entail that Gregor is ontologically bizarre. Indeed, suppose Gregor was an ontologically bizarre character. Then, one can hardly see why the entomological category Nabokov uses would make sense in the world of *the Metamorphosis*. In other words, any natural science is somewhat out of place in a world where "the laws of reality" are completely different. Suppose for instance I ask the blood type of the non-existent knight. Well, given that the nonexistent knight has no body, it seems straightforward that he has no blood, hence no blood type. Hence, elementary physiology is irrelevant in Calvino's novel: the question makes no sense. By contrast, one can see how Nabokov's question *makes sense*. Of course, Nabokov's argument is controversial; but the mere fact that it makes sense suggests Gregor is not ontologically bizarre.

This line of thinking points towards an interesting asymmetry. It seems that entomologically bizarre characters are not necessarily ontologically bizarre. But an ontologically bizarre character is *de facto* an entomologically bizarre character. There hides Gregor's indeterminacy: it is merely entomologically indeterminate, thus not zoologically indeterminate, not physically indeterminate, not ontologically indeterminate, etc.

I think the asymmetry reflects deep-rooted intuitions about epistemology and metaphysics which I will try to make explicit. This connection between the principles governing fictional imagination and epistemology and metaphysics will be important later on.

[11] This is very much the gist of the point inspired by Nabokov I made earlier about *the Metamorphosis* being a T-fantastic story.

12.3.1 Naïve epistemology

There are different areas of knowledge associated with different disciplines we call the sciences. The easiest way to describe this is to differentiate the sciences by their degree of abstraction. Logic is more abstract than entomology. This means that the body of positive knowledge that logic aims at structuring applies to a larger set of objects than that of entomology. "Larger" here is qualitative, not quantitative. Logic applies to more *kinds* of objects than entomology does. Indeed, the laws of entomology apply to all kinds of insects, while the laws of logic also apply to all kinds of animals, to non-living creatures, abstract objects, etc.

In other words, logic is, by definition, more universal than entomology: the domain of entomology is the set of all the insects, while the domain of logic is everything there is; hence the domain of entomology is a subset of the domain of logic. All sciences, understood as production of structured bodies of knowledge, can ideally be compared by their degree of abstraction. This is what I call naïve epistemology. I think this is an intuitive view.

Now, ontology is manifestly related to logic. Let us say content ourselves with the idea that logic is to ontology what mathematics is to the natural sciences. Consequently, ontologically bizarre characters are fictional objects which contradict some logical knowledge. By contrast, entomologically bizarre characters are fictional objects which contradict some entomological knowledge.

Given naïve epistemology, entomologically bizarre characters are not necessarily ontologically bizarre; but ontologically bizarre characters are necessarily entomologically bizarre.[12] So we get the asymmetry we looked for.

A nice consequence of this naïve epistemology is that there are as many ontologically bizarre characters as they are logical laws. In so far as we agree on the fact that "all existing individuals are determinate" is a law of logic, then any indeterminate fictional object in its own world is an ontologically bizarre character. Same with objects infringing numerical identity, the principle of contradiction, etc.

Naïve epistemology is controversial. My purpose is not to defend it, but to show that such a naïve view can yield the kind of indeterminacy we are looking for.

12.3.2 Naïve metaphysics

The modal space is ordered in some sense. The easiest way to describe this is to consider the space of all possible worlds in which the actual world is but one

[12]Maybe the latter statement does not really mean anything, for it may very well be that entomology is irrelevant to ontologically bizarre characters.

point and define a similarity relation between pairs of possible worlds. A world exactly similar to our world in which I am a woman is more similar than a world in which some knights are nonexistent. Since possible worlds are usually understood as being logically possible worlds, we should call impossible worlds the worlds in which ontologically bizarre characters dwell.

Following this intuition one easily welcomes the idea that impossible worlds are less similar to our world than any possible world.[13] Thus, ontologically bizarre characters live in the fringes of the modal space, and this is what makes them remarkable.

Consequently, when one is required to imagine an ontologically bizarre character, one should travel the modal space very far away. In the remote world(s) one ends up, all hell breaks loose. If one tries to keep cool-headed: what should one think of the distance between the actual world and an impossible world in which, say, non-existent knights are dwelling *and* the laws of entomology are the same as in reality? What about an impossible world in which non-existent knights exist *and* the laws of entomology are not the same as in reality? My intuition is that they are both very, very far away... In practice, one could really be imagining anything, if there is not more specific instructions.

On the other hand, when one is required to imagine an entomologically bizarre character like Gregor, there is not that much leeway. Without further notice, imagining ontologically bizarre characters is out of the question.

This naïve metaphysics thus explains the wanted asymmetry and points towards the *pragmatic* nature of fictional indeterminacy.

12.4 Finding the indeterminacy on the pragmatic nature of fiction

Friend gave an implicit argument leading to the conclusion that Gregor is not ontologically indeterminate: this is the argument from epistemology. I argued that the conclusion is true, but the argument is not valid in fictional contexts: one needs an extra premise. This is due to the fact that in some fictions, there are indeterminate characters. Consequently, the extra premise should say something like: "*The Metamorphosis* is not such a fiction". This extra premise is plausible in its own ground, once we agree that the genre of *the Metamorphosis* is of the T-fantastic.

This much insures that Gregor's indeterminacy is not ontological, but it does not say anything about Gregor's entomological indeterminacy. The reason why Gre-

[13] In [Nolan 1997], this is called "SIC: Strangeness of Impossibility Condition". See also [Berto and Jago 2019] for a discussion of this idea.

gor has one kind of indeterminacy and not the other, I suggested, comes from an asymmetry between the "priority" of ontology over entomology. I gave two sketchy possible theorizing about what "priority" means here: either by appealing to the notion of abstractness in a general epistemological framework or by calling in the notion of distance between possible worlds in a general metaphysical framework. I think this gives enough thickness to this crucial idea of priority which I will now use.

In the following, I will show that Gregor's (entomological) indeterminacy depends on the fine-tuning of this idea of priority in practice. I will show that this fine-tuning is a result of a pragmatic adjustment. Gregor's indeterminacy generalises to all fictional characters, by virtue of this pragmatic feature of fiction reading.

12.4.1 Two readings of the reality principle

Let us go then, back to Nabokov and Smith_3 having this GBD. They both agree that *the Metamorphosis* is a T-fantastic story, in which the "laws of reality" are those we experience in our world. Thus, they both read the text in such a way that *the Metamorphosis* requires imagining a world in which the "laws of reality" are those we experience in our world. This insures that nobody ends up in a remote impossible world.

But there are two steps in the debate. These are shown by the two subsequent questions Nabokov raises about Gregor. The point is that both Nabokov and Smith_3 answer Q1 in the same manner, according to their commitment to the genre of the story: it is fictional that Gregor is an insect. However, they answer Q2 differently, and they can justify their answers: Gregor is a cockroach according to Smith_3's reading and he is a beetle according to Nabokov's.

At this point, [Friend 2017b]'s terminology will be useful to distinguish between these levels of what is to be imagined when one reads a fictional text:

> I will say that a work *mandates* imagining that P if failure to imagine that P would mean falling below a minimum threshold for comprehension. A work *prescribes* imagining that P if we should imagine that P to have a full appreciation of the story. Finally, a work *invites* imagining that P on the following condition: if the question arises and we must choose between imagining that P and imagining that not-P, we are required to imagine the former.[14]

Imagining Gregor as in insect is clearly mandated by Kafka's text. By contrast, giving an answer to Q2 seems to be an invitation to imagine on the basis of Kafka's

[14] [Friend 2017b], p. 2.

text.[15] So we are not at the same level of obligation, so to speak.

Invitations, thus, are where the indeterminacy come from. I think Friend's definition of invitation contains the core feature of the indetermination we are looking for: "if the question arises and we had to choose". The case where the question does not arise is uninteresting, because in this case one cannot define what where the indeterminacy lies. If one never asked whether Gregor is a beetle or a cockroach, maybe it is fictional that he is a beetle unbeknownst to all reader's of *the Metamorphosis*. In this case, he is not indeterminate in the sense that both Smith_3's and Nabokov's elucidations of the text are correct and available, albeit incompatible. So the interesting part is "if we had to choose". This is because one is put under pressure that they imagine the way they imagine.[16]

On the one hand, the reader is already committed to what they are mandated and prescribed to imagine. On the other, the reader has to choose whether P or not-P. There are two cases, either the proposition to be imagine is logically independent from the prior commitments, or it is not. In the first case, one can choose either proposition without too much damage. In the second case, one has to either choose the proposition in keeping with the prior commitments, or revise one's commitments altogether.

Here are some examples: If a reader is asked whether Sherlock Holmes has a mole on his back or not, given that imagining this proposition or the negation of it is logically compatible with the reality principle and other information one has about Sherlock Holmes, this is a case of free choice. I think the GBD is an example of the second case: as shown in the discussion about Smith_3's reading, if one were to loosely apply entomology to the world of *the Metamorphosis*, then Gregor should be a beetle (because he gets stuck on his back). But if we were to strictly apply entomology to the world of *the Metamorphosis*, as shown by Smith_3's reading, Gregor could not even be an insect of that size. So we have a clash between a prior commitment and a proposition to imagine. Smith_3 resolves the clash by relaxing the prior commitment so that entomology *does not* strictly apply to Gregor. Consequently, the fact that Gregor gets stuck on its back is not necessarily a fact of entomology and can be explained otherwise. This is indeterminate.

[15] Arguably, Nabokov would say that this is a prescription, given that he makes a lot of what seems to be a detail.

[16] At this point, one should appreciate Friend's choice of word when she writes "invitation": in fiction, if you are invited, you have to choose.

12.4.2 Quine's notion of indeterminacy

This constant negotiation between abiding to an explicit prescription and prior commitment echoes Quine's peculiar notion of indeterminacy which appears in different places of his work. One place is the so-called "indeterminacy of translation", where a field linguist is supposed to make choices given prior commitments about the grammatical structure of the foreign language they use.[17] Another place is Quine's criticism of crucial experiment in the natural sciences[18]. Quine considers the idea that "total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field."[19] In both cases, we have to keep the balance between pressure coming from the outside and prior commitments which are negotiable.[20]

Quine uses this special notion of indeterminacy to produce an argument against the possibility of a truth-functional analysis of counterfactual statements. Consider the following pair of statements:[21]

(14) If Caesar had been in charge [in Korea], he would have used the atom bomb.

(15) If Caesar had been in charge [in Korea], he would have used catapults.

Suppose the speaker here is following a reality principle, so that they depart minimally from the actual world in a world where the antecedent is true.

What the pair of statement shows, Quine argues, is that given two relevant parameters of evaluation, it is possible to comply to the reality principle in two radically different manner. Either the "priority" is given to Quine's actuality: thus we bring Caesar in the sixties in our imagination and wonder what Caesar might have done with the US military technology. Or the "priority" is given to Caesar's actuality: thus we bring the Korea war into Caesar's time in our imagination and wonder how Caesar would have conducted his own troops. Quine's point is that the required idea of "priority" is pragmatically indeterminate: it depends what one means, i.e. what kind of imaginative project one has in mind.

[17] These commitments are called "analytical hypothesis" in [Quine 1969]. For the first exposition of Quine's example, see [Quine 1960].

[18] The so-called "Duhem-Quine thesis" or what is also called "epistemic holism". This thesis comes from [Quine 1951].

[19] [Quine 1951], p. 39.

[20] These "quinean" situations correspond to Goodman's notion of "reflexive equilibrium" as put forward in [Goodman 1955]. Goodman's examples are more convoluted though. For the record, when I say "Quine's notion of indeterminacy" I should perhaps say "Goodman's" if he is the one who invented the notion in the first place.

[21] [Quine 1960] § 46.

12.4.3 Gregor is pragmatically indeterminate

I think the GBD is a sophisticated instance of the same phenomenon. Given that all fictions use some principle of reality and that there are always several ways of implementing this principle, the same kind of indeterminacy pervades the fictional imagination in general.

This indetermination is not semantic, nor is it ontological. I would characterise it as pragmatic. In [Lewis 1973], Lewis proposes to interpret this pragmatic indeterminacy as sensitivity to the context of utterance. I think this is compatible with what I have said in this part. The fact that we read a text in the context of a literary debate or for the passing of time prompts different problems of elucidation.

Quine's epistemological picture allows for a nice connection between the natural sciences and the interpretative practice of elucidating texts of fiction. The hinge for such a connection is his original notion of indeterminacy.

12.5 Conclusions

In this chapter, I argued that indeterminacy is the hallmark of fiction as opposed to non-fiction. As such, Gregor is fundamentally indeterminate with respect to insecthood in a way real insects are not. This is the reason why the GDB is faultless.

I showed that Gregor's fictional indeterminacy was not a result of the vagueness of natural language expressions. Then I showed that it is not an ontological indeterminacy either. These considerations are not trivial, for I showed that in many fictions dwell "ontologically bizarre characters", some of which are *literally* indeterminate.

Rather, I insisted that Gregor is *entomologically* indeterminate. The fictional indeterminacy thus originates in the specific background knowledge one needs to import into Gregor's world. This led me to consider that there are (at least) two readings of the reality principle, (at least) two different ways to apply it, calling in subtle distinctions between *mandates*, *prescriptions* and *invitations* to imagine which are defined in [Friend 2017b]. Nabokov and Smith_3, in order to justify their readings of *the Metamorphosis*, both apply a reality principle but in two different ways, i.e. when *invited* to imagine whether Gregor is a beetle or a cockroach, they apply the reality principle in different ways. Nabokov calls in entomology, while Smith distances himself from entomology. They both have good reasons to do so, given their prior commitments.

This situation consists in resolving a tension between an explicit prescription (explicit fictional propositions) and prior commitments (the reality principle). In reality, things are interrelated. Therefore, imagining that something is different in

a world which is the same is bound to create such tensions. The same kind of problem exists in Quine's holistic epistemology (and probably in Goodman's too). Observation and experiment are like prescriptions to include some new information into one's model of reality: it creates a tension. There are different ways of resolving such tensions. When it comes to epistemology, Quine favours a form of conservatism which consists in keeping one's core knowledge as fixed as possible. When it comes to counterfactual thinking and other imaginative scenarii, however, the priority really depends on what one wants to do. Different priorities yield different resolutions. This is why I termed this kind of indeterminacy "pragmatic".

In the last chapter devoted to the GBD, I will try to make explicit these conditions of possibility of great debates. Given the pragmatic indeterminacy of fictional events, it will be useful to use a comparative method in order to objectify the necessary ingredients of this faultless debate. I will thus construct a minimal pair which will reveal how the different features of the GBD that we have analysed get together so as to make this debate possible and great.

Chapter 13

Conditions of possibility of great debates

> These bits of information about the actual world would not be decisive, of course, but they would be relevant since they tell me more about the non-actual situation that I have selected.
>
> [Stalnaker 1968].

Is the great beetle debate an isolated case? Are there other debates over the elucidation of a fictional detail? Can we systematically construct great debates about any detail of a work of fiction? In this chapter, I will try to make explicit what the conditions of possibility of great debates are. Eventually, we should have a general recipe for constructing great debates.

On the face of it, the conditions of possibility of the GBD are straightforward. It depends on the possibility of applying a natural science to a fictional situation. In such conditions, one can import positive knowledge about reality into a fictional setting. This is, in essence, what a principle of reality aims at describing. Concerning Gregor, there is a natural science, namely entomology, which is arguably relevant for the question of Gregor's nature. Entomology is indeed relevant, because one is committed to imagining that it is fictional that Gregor is an *insect*; and entomology is a natural science whose objects, by definition, are the *insects*. Had Gregor turned into a primate, primatology would have been relevant and not entomology. But is there always a relevant natural science corresponding to the imagining of a fictional situation?

Of course not. Take, for instance, your favourite story involving a dragon. There are no branch of zoology whose objects contain dragons. Given what is usually

fictional about dragons, the branch of zoology specialised in the study of reptiles should be relevant.[1] But of course, no real reptile can breath fire, so there is no positive knowledge coming from zoology which would be relevant in analysing the mechanism of fire-breathing. Yet, we *know* a lot about dragons: if you read a story containing a dragon, you know pretty much what you should imagine: it breathes fire, it has scales, probably wings, etc. The reason why this is so is because there are quite precise *conventions* according to which dragons are such and such.[2] Maybe there is a natural science whose object is to study these kinds of conventions, like anthropology or ethnology. But I think it is pretty intuitive to set apart conventions from natural science descriptions. Hence, one can also rely on conventions when no natural science is available. But is there always a relevant natural science or convention corresponding to the imagining of a fictional situation?[3]

The answer is no. I will provide an example taken from Gogol's short story *the Nose* in which no natural science nor any convention can ground the imagining of a fictional detail. Consequently, I will argue that *the Nose* does not ground any analogous debate about elucidation. Interestingly, Gogol's and Kafka's stories are very similar so that they form a minimal pair. So I will show how one can define a pseudo-debate which is formally identical to the GBD, albeit an empty shell, so to speak.

[1] In [Borges 1967], p. 152, "The Western Dragon":

> A tall-standing, heavy serpent with claws and wings is perhaps the description that best fits the Dragon. It may be black, but it is essential that it also be shining; equally essential is that it belch forth fire and smoke. The above description refers, of course, to its present image; the Greeks seem to have applied the name Dragon to any considerable reptile.

[2] In [Borges 1967], one can find at least three different kinds of Dragons coming with different conventions: the "Chinese Dragon", the "Eastern Dragon" and the "Western Dragon". However, in the preface to the 1957 edition, Borges writes:

> We are as ignorant of the meaning of the dragon as we are of the meaning of the universe, but there is something in the dragon's image that appeals to the human imagination, and so we find the dragon in quite distinct places and times. It is, so to speak, a necessary monster, not an ephemeral or accidental one, such as the three-headed chimera or the catoblepas.

[3] In [Lewis 1978]'s idiom, one can use the term "carry-over" to denote either a relevant natural science or a convention.

13.1 On the non-existence of a great Nose debate

13.1.1 The fictional detail

The Nose is a short story written by Gogol and published in 1836. The main character of the story is the Collegiate Assessor Kovalyov who, one morning, wakes up to find his nose missing. He becomes literally nose-less. Alarmed, looking for his nose all around the city of St Petersburg, Major Kovalyov runs into his nose in the street, dressed in the uniform of a higher-ranking official than himself. Suddenly, the nose enters a church. Stunned Major Kovalyov follows him in. Inside the church, the two characters exchange a few words. Finally, the nose sends Kovalyov packing using his higher-ranking authority. Major Kovalyov then tries to start legal proceedings against his nose without success. The story unfolds with other interesting twists and turns.

Let us focus on a detail of the story. One of the fictional event is the following: nose-less Major Kovalyov is walking in the street and runs into his nose walking in the same street, dressed in the uniform of a high-ranking official. So the reader is explicitly asked to imagine the Nose *walking* down a street. But how would it do that? Does the Nose have legs? There are good reasons to think that it does not have legs, because it is a nose: but at the same time, if it is *walking* down a street, it probably has legs. I suspect there are different natural ways of getting around this difficulty, now I have raised the question: one would imagine the Nose, dressed, levitating ahead; or one would imagine a big nose bumping forward; or even a nose with long thin legs walking like a crane; or a very human-like creature whose face consists only of a nose.[4] There are probably other ways of realizing this invitation to imagine the Nose walking down the street.

13.1.2 Pairing minimally Gregor and Kovalyov's nose

It seems that the two stories are very similar in many respect, so that the comparison is not so artificial. The situation is, as I will show, similar to the point that we can construct a formal rendering of what a "great nose debate" would be.

Both fictional events under consideration are *detail* in their stories. If the question did not arise, most readers would probably never have thought it was an interest-

[4]There are numerous illustrations and adaptations of Gogol's story for visual media. I noticed that this last option is very widespread (see for instance this animated short film made by Alexandre Alexeieff and Claire Parker in 1963: https://www.youtube.com/watch?v=h9jHm455fuE). But the imagination of illustrators knows little limit, as a quick search on the internet reveals!

ing question while reading the stories. So in both cases, it is an elucidation of an *invitation* to imagine: there is no *mandate* involved.

Both fictional events yield several incompatible correct elucidations. Indeed, it is easy to see that we can consider several readers of *the Nose* such that some elucidate the fictional detail in a correct manner and some in an incorrect manner. That is to say, one can easily have different "Smiths" reading *the Nose* so as to get a distinction between correct readings and incorrect ones. The correct elucidations I consider in the previous subsection as "natural" ways of getting around the problem will do. Therefore, one can run the same argument as in the case of *the Metamorphosis* leading to the conclusion that there is a faultless disagreement. This leads to the same defense of pluralism about elucidation I developed above.

Both fictional events are just as impossible and implausible as can be. One may have the feeling that the two fictional events differ in their plausibility or possibility. But this is an irrational feeling. Both turning into a monstrous insect or losing one's nose while sleeping are completely impossible events. So any theory of plausibility or possibility which makes a significant difference between Gregor's fate and Kovalyov's lot will be deeply counter-intuitive to say the least.

Both stories are of the T-fantastic genre.[5] Both stories require imagining that a supernatural event occurred in a world where the "laws of reality" are as we experience them otherwise. Indeed, in Kovalyov's world, everything seems to be normal except for a nose on the loose. For instance, when Kovalyov tries to start legal procedures against his nose, the office clerk follows the same procedures as in reality and comes since there is no possibility of charging one's body part in reality, he cannot follow on from Kovalyov's demand. This prompts an incident in the office, since Kovalyov insists and gets angry at the office clerk.

In [Erlich 1956], the two stories are systematically compared and Erlich summarises the similarities in the following manner (his claim is in keeping with both stories being T-fantastic):

> Clearly, Gogol's nonsense narrative lacks the quality of an existential disaster. Yet it shares with the grimmer story of Kafka the discrepancy between its "realistic" mode of presentation and the utterly incredible

[5]Not all literary critics agree on this. For instance, [Hardy Jr and Stanton 2002] takes *the Nose* to fall under "magic realism". On the other hand, [Shukman 1989] and [Whitehead 2017] take *the Nose* to be a paradigmatic instance of the fantastic *in Todorov's sense*. Interestingly, for Todorov, *the Metamorphosis* and *the Nose* both raise the same issues for his original definition, and he treats them on a par. See [Todorov 1970], p. 177-180 and my discussion of this problem in subsection 9.2.2, footnote 15.

central event.[6]

These striking similarities justify my calling them minimal pairs: they differ only insofar as one grounds a debate and not the other.

13.1.3 A debate about Gogol's nose?

So it is easy to construct a debate about the Nose's walk in the form of a Q&A as I did for Gregor. Here is such a formal equivalent:

Q1: Is the Nose capable of moving on its own?

A1: Yes.

Indeed, it is explicitly said in the text that Kovalyov walks past his nose *walking* down a Street in St Petersburg.

Q2: Does the Nose have legs?

A2: Well, it depends own you imagine the scene.

There is one scenario in which the Nose has legs, and it literally *walks* past Kovalyov. Other scenarii involve other ways of progressing, like levitating ahead or bumping forward.

If we introduce two readers, we can artificially create a faultless disagreement based on the two available readings suggested here.

13.1.4 Aside on the explication and thematic interpretation of *the Nose*

Nabokov strikes again

There is a very large literature on the possible explication and thematic interpretations of Gogol's short story.[7]

[6] [Erlich 1956], p. 102.

Interestingly, the two uncanny, supernatural events are surprisingly similar even in the way they fictionally happen: both characters wake up with a body change, without any idea of what happened in the preceding night; in both cases, the other fictional characters do not seem very surprised or alarmed as the main characters naturally are.

[7] See the first two pages of [Shukman 1989] for an interesting review of the critical literature on Gogol's text.

Interestingly, Nabokov has commented on *the Nose*, too. He finds a "nasal leitmotiv" in the work of Gogol, a "nose-consciousness [which] resulted at last in the writing of a story, *The Nose*, which is verily a hymn to that organ". He remarks that the Russian language, quite remarkably, contains "hundreds of proverbs and sayings that revolve around the nose". Finally, he displays his own thematic interpretation of the story linking it to an episode of Gogol's life:

> The fact remains that Gogol's long, sensitive nose had discovered new smells in literature (which led to a new "frisson"). As a Russian saying goes "the man with the longest nose sees further"; and Gogol saw with his nostrils. The organ which in his juvenile work was but a carnival character borrowed from the cheap shop of ready-made clothes called "folklore", turned out to be at the climax of his genius his most important ally. When he destroyed his own genius by trying to become a preacher, he lost his nose just as Kovalev lost his.[8]

Again, Nabokov's interpretation is full of irony and wit and I will leave it to literary critics to appreciate this biographical interpretation of the short story.

Ephraim Sicher adds a very relevant piece of information, taken from the literary critic Vinogradov. At the time when Gogol writes the story, "nose jobs" were beginning to be a social phenomenon in Russia:

> Vinogradov (1929: 7-88) proposed a number of sources for *The Nose* and his excursion into the "nosology" of the 1820s-1830s is essential to an understanding of the ontological and compositional context of the story, as well as its reception in Russia and abroad. His point is that the motivation or demotivation of the plot can only be understood in the context of developments in nose-grafting or "plastic surgery" reported in the Russian press, the comic plot of the loss of the nose and punning on nasal figures of speech.[9]

One can see how inquiring about the nature of Gogol's Nose traces back to real phenomena, here. Although I doubt that this "nosology" would provide an answer to our debated elucidation...

Linguistically oriented criticism

Among many other thematic interpretations, one is particularly relevant to my purposes. It is what Ann Shukman calls "linguistically oriented criticism". In particular,

[8] [Nabokov 1959], p. 11.
[9] [Sicher 1990], p. 212.

she translates and comments on a passage from Roman Jakobson thus:

> Besides, the central and compounded "absurdity" of the tale [...] not only defies logic and probability, but is also impossible to envisage in the mind's eye; it is a play with words, a paronomasia, a textual joke, as the more linguistically oriented critics have noted. For Roman Jakobson, for instance, the story demonstrated the purely verbal nature of the "realised oxymoron", a figure of speech which has meaning but no object: "Such is Gogol's 'Nose' which Kovalyov recognises as a nose even though it shrugs its shoulders, wears full uniform and so on."[10]

How can a nose "shrug his shoulder"? Here is another fictional detail on which one could ground a debate. Ann Shukman, in [Shukman 1989], pursues this idea that *the Nose* is a textual joke, an extended play on words, which was made to defy literary analysis altogether.

13.1.5 The difference, and a pseudo-great debate

So the debate about the Nose's walking is formally identical to the GBD. But there is also a major difference. In the case of *the Nose*, there is no underlying natural science to justify one's readings. Indeed, one can perform the following *reductio* argument.

The relevant natural science for the studying of walks and gaits is a branch of anatomy called functional anatomy. Functional anatomy is the natural science which studies the relationship between anatomy and movement, answering such questions as: what are the constraints of the skeleton on possible movements? How do the muscles of the body cooperate to produce a given movement? How does the balance of the body coincide with movement? etc.[11] Unfortunately, functional anatomy says nothing about individual organs. A nose is not in any straightforward sense a possible object of functional anatomy.

Now, suppose there is a real natural science whose object of study is the walks, or rather the way of progression of organs separated from the body. Let us call this science *schismatic functional anatomy*. This natural science would describe the way livers move on their own when separated from their body of origin, the locomotion of lungs, single arms, brains and pairs of eyes, and so on. Naturally, schismatic functional anatomy would have something to say about the walks of noses;

[10] [Shukman 1989], p. 64.

[11] This is a very difficult and fascinating natural science which is currently challenged in its results by the advance in robotics. Making robots which can *walk* is a surprisingly difficult task; especially if one wants to design robots which can walk *like human beings*.

the different walks available for separated noses would probably depend on their size: snub noses do not move in the same way as big noses do; probably it would depend on the species they are separated from: dog noses tend to be quadruped whereas human noses tend to be bipeds (these are only statistical facts); and many other factors. As in all natural sciences, controversies and discoveries are part and parcel of the positive knowledge it delivers; the history of schismatic functional anatomy is also quite a thing, since most of the scientists were born Russian, a coincidence which is still an open area of research.

Given schismatic functional anatomy, one can construct an argument which supports one of the readings and contradicts others. So there would be a "great Nose debate" just like there is a great beetle debate: two readers disagreeing about a detail of a work of fiction, one producing an argument to favor their own reading with the help of a relevant natural science (namely schismatic functional anatomy), the other resisting the argument in the manner of Smith_3, calling upon the fact that one can imagine as one wants (within certain limits).

Now, schismatic functional anatomy is, as far as I know, merely a pseudo-natural science; hence, this reconstruction is merely a pseudo-"great Nose debate".

The other possibility of having a debate would be to have a convention. Just like we know that dragons breath fire, because there is a convention which tells us that it is fictional (without mention to the contrary) that dragons breath fire, there could be a convention which tells us how noses progress when severed from their body of origin. Unfortunately, I have never heard of any such convention.

Therefore, despite obvious formal similarities, there is no "great Nose debate". The closest we can get is to invent a cumbersome pseudo-"great Nose debate".[12]

13.1.6 Is there even a disagreement about *the Nose*?

We have shown that there is no way of justifying rationally one's imagining about the detail of the Nose's progression. No natural science can be called for and no convention can be used. In other words, there is no prior commitment to any imagining how a nose is supposed to progress. Therefore, in a sense, there is no *debate* (or disagreement) for there can be no argument.

Recall: Kölbel's general definition of a disagreement requires:

(i) entertaining incompatible propositions;

(ii) having some justification for it.

[12]If Gogol's story is really a joke as Shukman suggests, then maybe this pseudo-debate can be the final punchline.

The distinction between faulty and faultless is relevant when one considers *who* is justified: if both are equally justified, then we have a faultless disagreement. When philosophers argue against Kölbel, they agree on what a disagreement is. However, they typically question the *possibility* of both parties having *equal* justification.[13] So everyone seems to agree on what a disagreement requires to take place.

In the case of *the Nose*, it seems that condition (ii) is not met. Indeed, the readers of *the Nose* who are "disagreeing" at the level of A2 cannot provide any justification their preferred elucidation. Therefore, it seems that there is no *disagreement*. In other words, it seems that we have a pure case of imagining past each other: one can wonder why readers imagine the way they imagine, but one cannot wonder what is fictionally the case.

This suggest that fictional disagreements presuppose there is a real ground for disagreeing. Indeed, what is lacking in the case of *the Nose* is a relevant natural science or a convention. These are real things.

Consequently, our initial distinction between real and fictional disagreements is not so justified anymore.[14] Although it was heuristically useful to use this distinction, I think we have reached good reasons to criticise it. This, in itself, is a nice result of the minimal pairing I have been proposing in this section.

13.2 Further considerations about fictional imagination

13.2.1 Two kinds of fictional objects

Given the above discussion, one might be tempted to generalise. The comparison between Gregor and Kovalyov's Nose invites us to distinguish between two kinds of imagination.

Intuitively, one kind of imagination is constrained by something like a convention or some positive knowledge originating in a natural science; the other is free. Gregor as a fictional object originates in the exercise of a constrained imagination while Kovalyov's Nose comes from free imagination. Of course, it is not black and white, since I have remarked in the preceding section that Gregor himself, being fundamentally indeterminate, should be freely imagined when it comes to the detail. Therefore, in practice, fictional object are surely a complex mix of constrained and free imagination. But I think the distinction can be made rigorously all the same.

[13] See [Stojanovic 2017].
[14] As summed up in Table 9.1.

I will call *impure imagination* the imaginative activity which incorporates some information originating in positive knowledge or convention (for simplicity, I will henceforth omit mention to conventions, but everything I will say should equally apply to it *mutatis mutandis*). By contrast, *pure imagination* will denote an imaginative activity which incorporates no such external information.

This terminology is adapted from a passage in Proust *Sur la lecture* (I translate):

> I so much longed for a continuation of the book, and, if it ever was possible, for more information about the characters, so as to learn about their lives now, to live in such a way that the love that they had inspired was not so foreign to our own life [...].
>
> [Footnote: We can try to do this, using a detour so to speak, with books which are not of pure imagination and where there is a historical substratum. Balzac, for instance, whose somehow work is mixed with reality and wit that has not been transformed enough, lends itself particularly well to this kind of reading.][15]

As remarked above, the practice of fictional imagination is a very sophisticated combination of both kinds of imagination. Consequently, if we consider fictional objects, that is the result of fictional imagination, they will almost always be both pure and impure, depending on *how* one considers them.

Take, for instance, Pegasus. If one considers Pegasus on the background of zoology, it clearly is an object of pure imagination. Indeed, we should imagine that Pegasus is a winged horse, so it has both characteristic features of a mammal and characteristic features of a bird. But, according to zoology, no animal can be both a mammal and a bird: these categories are mutually exclusive. *Prima facie*, then, it seems that Pegasus is *neither* a mammal *nor* a bird. So imagining Pegasus is not constrained by zoology.[16] Of course, if one had further information about Pegasus

[15] Here is the original in French [Proust 1906]:

> On aurait tant voulu que le livre continuât, et, si c'atit possiblen avoir d'autres renseignements sur tous ces personnages, apprendre maintenant quelque chose de leur vie, employer la nôtre à des choses qui ne fussent pas tout à fait étrangères à l'amour qu'ils nous avaients inspirés [...]
>
> [Note de bas de page: On peut l'essayer, par une sorte de détour, pour les livres qui ne sont pas d'imagination pure et où il y a un substratum historique. Balzac, par exemple, dont l'œuvre en quelque sorte impure est mêlée d'esprit et de réalité trop peu transformée, se prête parfois singulièrement à ce genre de lectures.]

[16] Although most I think would agree that Pegasus is a vertebrate. So maybe I should say it is not constrained by *vertebrate* zoology.

(Does Pegasus lay eggs? What is the most recent common ancestor with horses?), then maybe the question could be settled.

By contrast, if we consider Pegasus in the background of the theory of gravitation, it clearly is an object of impure imagination. Indeed, Pegasus has a weight which is the result of its interaction with the earth in a way that can be predicted by Newton's theory of gravitation (or Einstein's, for that matter). This prompts the imagining of Pegasus very large wings, so that they can carry it in the air; the imagining that it takes a great effort for Pegasus to take off, and so on.

This line of reasoning generalises to all fictional objects. From Nabokov's viewpoint, Gregor is an impure object of imagination if we consider him against the background of entomology. But from Haldane's viewpoint (for whom size matters!), Gregor is a pure object of imagination given entomology. Debating fictional events boils down to determining the degree of pure and impure imagination necessary for what counts as a correct imagining.

One can then in principle define a fictional object as the set of constraints on imagination it requires. Everything which is not constrained is by definition left for pure imagination. As stressed above, what goes into the set of constraints is pragmatically indeterminate.

13.2.2 Literary debates and thought experiments

At this point, we have an interesting picture of fictional imagination as a mix of pure and impure imagination. Debates like the GBD are controversies about the amount of impure imagination involved in the elucidation of a fictional story. I will now show how this connects with what is usually called a "thought experiment".

The set of constraints which are definitional of fictional characters have, so to speak, a world-to-fiction direction of fit. That is, the constraints are a normative mechanism whose normativity comes from the real world, the positive knowledge we have about the world in the case under discussion, and they mandate imagining. Debating the detail of a story amounts to holding fixed the constraints and adjusting what the fictional object fictionally by adding or removing some constraints.

So in principle, one can wonder what would happen if one held fixed the fictional object as it is imagined in one's mind and trying to adjust the set of constraints accordingly.[17] In this case, one starts from a fictional character and tries to find what is the set of constraints on the real world which could define such a fictional object. In this situation, the direction of fit is reversed: it is fiction-to-world. I think

[17]This *having* the fictional object in one's mind is guaranteed in principle by the definition of pure imagination (one freely chooses to imagine such and such).

this situation corresponds to an empirical case where we imagine something and then wonder what the world would have to be like for such a thing to be the case. This is what is usually called a thought experiment.

Here is a concrete example of what I claim to be the dual character of literary debates and thought experiment.

Take La Fontaine's first fable, inspired by Aesope, *The Grasshopper and the Ant*. This fable mandates the imagining of two talking insects, a grasshopper and an ant. But, of course, it says nothing about the detail of such a marvelous event. It is thus easy to construct a debate analogous to the GBD. For in order to speak out loud, one has to have a phonatory device adapted for articulate language (says macro-anatomy). What are the specifics of the phonatory devices of talking grasshoppers and talking ants? Here we go again.

Instead of taking this now familiar naturalistic stance, one can easily say that the talking insects under discussion are pure objects of imagination when it comes to macro-anatomy. Hence one can freely imagine La Fontaine's nice dialogue.

At this point, there is this fictional this fictional situation in which a spendthrift singer unsuccessfully begs for food a stingy summer-worker. One can now wonder what kind of positive knowledge one would need to import such a situation into the fiction. In particular, one would need some psychological positive knowledge about singers and summer-workers. If psychology implied that all stingy people do not lend what they earned, then the fictional situation imagined above would be a case of impure imagination. Read in this way, "The Grasshopper and the Ant" matches the psychological theory according to which stingy people do not lend what they earned. This is the familiar reading of a fable as a thought experiment, containing a moral. The direction of fit is indeed fiction-to-world in this case.

At this point, one can go back in the other direction. There is now an explicit constraint for the imagination, so it should be possible to debate it. Indeed, one might question whether this begging for food is really in keeping with what psychological theory of singers says. For instance, one might hold that all real artists are proud. Singers are artists. So the grasshopper should be proud. But the grasshopper fictionally has no pride, for it begs for food. At this point, one can raise a debate about the correct elucidation of the fable. One would argue that the grasshopper is not a *genuine* artist, for it would be too proud to beg for food and would happily die when the time comes.[18] This debate has a world-to-fiction direction of fit.

At this point, I suppose one can go back in the other direction by refining one's psychological theory about artists including subtle considerations about artists with and without pride. And so on. This going back and forth, I contend, consists essen-

[18]See for instance [Suits 1978] for a nice interpretation of Aesope's fable along these lines.

tially in changing the direction of fit of our imaginative scenario. Extracting a moral from a fiction is a very easy task: it is as if fiction was made for it. The fine-tuning of the fictionality conditions that I presented here nicely shows how debates about elucidations and thought experiments can be seen as dual activities involving imaginative scenarii, thus explaining how come the shifting of direction of fit suggested here is so easy.

Conclusions

In this chapter, I have provided an analysis of the great beetle debate. I have shown that it is a faultless disagreement. This result should be contrasted with the three other possibilities I considered.

The GBD is not a case of imagining past each other. Although it was interesting to follow this idea for a while in order to make a fruitful connection between the GBD and the phenomenon of imaginative resistance. Much of the interest of the GBD lies in the fact that the two readers have a conflict of elucidation, which is the specific kind of interpretation focusing on determining what is fictionally the case.

The GBD is not a faulty disagreement either. Indeed, I constructed a relative definition of correctness for imagination so that we have a clear idea of what counts as a correct reading and what counts as an incorrect reading of *the Metamorphosis*. This was sufficient to apply quite strictly Kölbel's definition of a faultless disagreement to the GBD. I concluded from this result that one should aim for a thesis I called pluralism about elucidation, according to which there are some fictional events which are indeterminate. This indeterminacy takes the form of a dependence on the actual reader's idiosyncratic way of imagining.

This led me to wonder about the pervasiveness and the origin of this crucial notion of indeterminacy of fictional characters. I showed that fictional indeterminacy does not originate in semantic vagueness, nor is it a kind of ontological indeterminacy. I showed that this indeterminacy is most probably pragmatic in nature, for it depends on how one applies the reality principle and this really depends on the reason why one reads fiction in the first place. Literary analysis is one reason to read fiction and yields debates which would not arise in more mundane activities. I made a tentative conceptual link with Quine's idiosyncratic notion of indeterminacy, providing new insights on the question.

Finally, I made explicit what the conditions of possibility of the GBD and debates of such nature are. In order to do this, I used the method of minimal pairs which is a basic method in formal linguistics today. I argued that one can construct a minimal pair by putting side by side *the Metamorphosis* by Kafka and *the Nose* by Gogol. It

showed that the crucial necessary conditions for debates like the great beetle debate is the availability of a relevant natural science or of a convention.

This result suggests an interesting follow-up: one can distinguish between pure and impure imagination which corresponds to whether such a condition is met or not. I take Gregor to be a result of impure imagination and Kovalyov's nose to be a result of pure imagination. This should be immediately qualified, because in practice, every fictional object is both pure and impure depending on the background against which it is considered. This qualification echoes the previous result about the fundamental indeterminacy of fictional characters. This dual structure of fictional objects provides an elegant unification of two phenomena: the possibility of debates like the GBD and the possibility of thought-experiments using fictional characters. These are the two sides of the same fictional coin.

In this part, I was interested in the fine-tuning of the concept of fictionality, as defended in part 1. Philosophers generally consider fictionality from a very general point of view. For this reason, some literary scholars have been disappointed or sceptical about these philosophical approaches to "truth" in fiction, not without reason. On the other hand, I think the reason why philosophers did not get into the detail is because it involves a lot of sophisticated literary criticism which is very daunting and difficult to systematise.

I showed that with the help of the concept of disagreement, taken from debates at the interface of epistemology and semantics, one can transfer to philosophy many ideas and insights which flourish in literary criticism. I hope that literary scholars will find in this work good reasons to get interested in contemporary debates in the philosophy of fiction, so that they eventually challenge philosophers with richer and livelier examples than those philosophers usually consider.

Part III

A problem about fictional names

Chapter 14

Definitions and scope of the problem

In this chapter, I will focus on a particular problem about fictional reference. Referring expressions, when used in a work of fiction, do not need to pick out an actual individual. Interestingly, though, referring expressions originating in a work of fiction can also be used in non-fictional contexts. When so used, they seem to pick up an actual individual. In giving a theory of fictional reference in general, one is thus faced with this problem: fictional uses and metafictional uses of fictional referring expressions arguably do not abide by the same constraints.

I will focus on names only, although everything I say should generalise to other referring expressions *mutatis mutandis*. I will largely assume that one version of the historico-causal theory of names is correct, in the wake of Kripke, Donnellan and Evans particularly, because nowhere is the problem between the fictional and the metafictional more clearly expressed. I will make these assumptions explicit only when the arguments I will consider require them.

As suggested as a conclusion of Evans's argument against Lewis's counterfactual analysis of fictional statements in subsection 4.3.3, functionalism and causal theory of reference should go hand in hand. I will show how pretence semantics can be used in a non trivial manner to solve the problem about fictional reference. Consequently, I will be defending a version of anti-realism.

But first, some definitions.

14.1 Two orthogonal distinctions

14.1.1 Names in fiction

Works of (linguistic) fiction typically contain proper nouns. There are two kinds of names which one finds in a fiction: those which originate in a fiction and those which do not.[1] Historical novels, typically, uses both kinds of names. For instance, in *War and Peace* by Tolstoi, the name "Pierre Bezukhov" originates in the fiction whereas the name "Napoléon Bonaparte" does not. I call *fictional names* the first kind and *real names* the second kind. In principle, a fiction could contain only fictional names, or else only real names.

When it comes to names, there is also an important distinction between *empty* and *full* names. Indeed, either a name refers to an individual, or it does not. When it refers to an individual, the name is full. When it does not, the name is empty. A typical example of a full name is the name "François Récanati" which refers to the actual philosopher who bears that name. A typical example of an empty name is the name "Vulcan" which does not refer to anything.[2]

So we have two orthogonal distinctions: names are either real or fictional, either full or empty. Real names can be full or empty. But fictional names, by definition, are empty. That is: names originating in a fiction do not refer to any real-world individual.

14.1.2 Fictional names and empty names

But fictional names are a particular sort of empty names. Indeed, one can distinguish between two kinds of empty names: those with which one attempts to refer to one real-world individual but fails to do so, and those with which one does not attempt to refer to any real-world individual and thus do not refer. To introduce technical terms, when one fails to refer we say that the name is *de facto* empty, whereas when one does not even attempt to refer, we say that the name is *de jure* empty.[3]

Fictional names like "Pierre Bezukhov" are *de jure* empty: using them, one does not attempt to refer to any real-world individual, rather the name names a fictional

[1] Originating in a fiction means that the first occurrence of the name was in a fiction. Not originating in a fiction means that the first occurrence of the name was in non-fictional discourse.

[2] "Vulcan" was introduced as a name for an hypothetical planet which causes Mercury's perihelion, by Urbain Le Verrier. It turns out that there is no such planet. Mercury's perihelion is explained by Einstein's general theory of relativity.

[3] This terminology maps onto Walton's basic distinction between imagination and fictionality, as presented in subsection 5.2.1.

character. "Ordinary" empty names like "Vulcan" are *de facto* empty.[4] Hence, fictional names are empty names, although there is no *failure of reference*, properly speaking. By contrast, ordinary empty names simply fail to refer.

Some have been puzzled by this last distinction. They ask: how is it possible for a name not to refer if they do not fail to refer? For if it does not fail to refer, shouldn't it be referring? One can see how these questions are play on words. But it points to a *specific* problem about fictional names, which is the central problem of this part.

To sum up, the logical space thus intuitively open is the following. There are full names like "François Récanati" and "Napoleon Bonaparte" whose function is to name and refer to some actual individual. There are ordinary empty names like "Vulcan" which aim but fail at referring to some actual individual, they are *de facto* empty. There are fictional names like "Pierre Bezukhov" which do not refer to any actual individual, although they are used to name and refer to some fictional character within the scope of a fiction, they are *de jure* empty.. All these names can be found in fictional texts. But fictional names are dependent on fictional texts in a way the other three kinds are not: they, and they only, originate in fictional discourse.

A general theory of names should take into account all these kinds of names. However, it seems that theorizing about fictional names in particular is challenging in its own sake. Fictional names face a problem which is revealed by the linguistic data I am about to present. I will discuss how specific to fictional names the problem is. It seems that a similar problem also affects a certain kind of empty names. If that is true, then the theories about fictional names which I will discuss in the following chapters should generalise to these empty names.

14.2 Fictional and metafictional uses of fictional names

Although fictional names *originate* in a fictional text, they can easily be later used in non-fictional discourse. Hence the distinction between *fictional* and *metafictional* uses of fictional names. Here are paradigmatic examples of such uses:

[4]The question of *mythical* names, like "Pegasus", is a vexed question. It depends on whether one holds that myth is a kind of fictional discourse or not, which is a controversial issue. I will follow [Salmon 1998] who argues that myth is non-fictional discourse. See also [Schaeffer 1999] (pp. 149-153) for a similar view. Consequently, "Pegasus" will be on a par with "Vulcan". Nothing in what I will say hinges on this, though.

(16) Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, seemed to unite some of the best blessings of existence; and had lived nearly twenty-one years in the world with very little to distress or vex her.

(17) Emma Woodhouse is a fictional character and main protagonist of Jane Austen's novel *Emma* first published in 1815.

(16) is the opening sentence of [Austen 1815] and first occurrence of the name "Emma Woodhouse". It is the paradigmatic fictional use of the name "Emma Woodhouse", where it is used to introduce a fictional character.

By contrast, (17) is a metafictional sentence. It does not and could not appear in Jane Austen's novel.[5] Such a sentence can typically be found in an encyclopedia entry. It is called "metafictional" because one typically produces such a sentence while talking *about* Jane Austen's fictional story.

The main contrast between the two sentences is that (17) is true whereas (16) is fictional. Indeed, fictional sentences come with fictionality-conditions.[6] Therefore, (16) has fictionality-conditions. By contrast, (17) is not a fictional sentence and seems to express truth-conditions *simpliciter*. Thus, if pretence semantics is useful for sentences like (16), it seems that it cannot handle sentences like (17). Pretence semantics is thus severely limited when it comes to giving a theory of fictional names.[7]

From this observation, indeed, one can reason in the following manner. If metafictional sentences like (17) have truth-conditions, it follows that in such context the name "Emma Woodhouse" refers. Indeed, given a widely accepted principle of compositionality, the truth-conditions of a sentence is a function of the semantic contribution of its parts. So the name "Emma Woodhouse" should contribute to the truth of (17). Given its place and function in the sentence as well as the fact that it is a proper noun, it must be that the semantic contribution of "Emma Woodhouse" consists in providing a referent. So "Emma Woodhouse" refers in (17). Consequently, fictional names, in metafictional contexts, are not *empty* names.[8]

[5]Arguably, it *could* appear in Jane Austen's fiction so that it would be fictional that Emma Woodhouse is a fictional character. But Jane Austen's novel would thus be a metafiction akin to Pirandello's *Six Characters in Search of an Author*. It clearly is not.

[6]See chapters 6 and 7 of this dissertation for such a view.

[7]Noting this fact, many have argued that metafictional uses of fictional names sound the death knell of pretence semantics (and of functionalism in general). This is why the functionalist should defend its ground on the problem of fictional reference.

[8]For simplicity, I will not consider gappy propositions here, since it does not change the main debate I am considering in this chapter. For defence of gappy propositions, see [Braun 1993]. For a critical discussion of gappy propositions involving fictional names, see [Everett 2003].

The conclusion of this argument contradicts the intuitive definition of fictional names. *Prima facie*, one would think that fictional names are a special kind of empty names: they do not refer to actual individuals. But this simple argument forces us to conclude that fictional names, despite appearances, are not empty names. This situation led to a debate in the philosophical literature.

Philosophers who are convinced by a form of the argument given above are said to hold a form of *realism* about fictional names. Realists hold that fictional names are not a kind of empty names. Philosophers who want to resist the argument and stick to the intuitive view are said to hold a form of *anti-realism* about fictional names. They argue that fictional names are a kind of empty names.[9]

The aim of this part is to defend a form of anti-realism which is compatible with and articulated to pretence semantics as argued for in part 1. Before I can do this, I will discuss the specificity of fictional names with respect to "meta" sentences. We will see in the next section that real names and empty names are also to be found in "meta" sentences. In trying to distinguish between metafictional sentences and the other "meta" sentences, we will see that some interesting complications will lead to a revision of the logical space given above.

14.3 How unique are metafictional uses?

It might be thought that the problematic metafictional uses are not unique to fictional names. After all, some sentences involving real names and ordinary empty names are also "meta". Let us call them "quasi-metafictional sentences". Here are some examples:

(18) François Récanati is a real person.

(19) Vulcan is a hypothetical planet.

How should we distinguish between metafictional and quasi-metafictional statements?

14.3.1 On a special feature of metafictional statements

One of the crucial distinctions between (16) and (17) concerns the kinds of predicates which are true of "Emma Woodhouse". In (16), Emma Woodhouse is presented as

[9]Stacie Friend, who is also an anti-realist, calls herself an "irrealist". I do not have any philosophical reason to prefer my label over hers. However, I find "anti-realism" far easier to pronounce over and over again than "irrealism". This is a non-native English speaker's opinion.

a flesh and blood individual with typical human attributes. Indeed, my neighbour could be like Emma in that she is handsome, clever and rich. By contrast, the kinds of predicates in (17) could not be shared by my neighbour. No real person could truly be said to be a fictional character, created by an author. In the context of (17), Emma Woodhouse is presented as an "individual of paper", so to speak.[10] What is true of this individual of paper suggests that we deal with an abstract thing which has been created by a human author.[11]

In order to distinguish between these two entities I will use the expressions "flesh and blood individual" and "individual of paper". These expressions are meant to be ontologically neutral. They simply capture the intuition about the different kinds of predicate that one typically finds in fictional and metafictional uses. Individuals of paper, especially, have predicates which are strange to flesh and blood individuals in the following sense: if one applies a typical "metafictional" predicate to a flesh and blood individual, it yields a category mistake.[12] Though nothing substantial hides under these terms, they will prove useful for expression purposes in the following discussions.

Interestingly, some of what is true of the individual of paper should count as a reason to deny that the flesh and blood individual exists. In other words, the truth of (17) entails the truth of:

(20) Emma Woodhouse does not exist.

Indeed, if Emma existed, then she would not be a *fictional* character. So the truth of (20) is inferred *a priori*, simply relying on what (17) means.

Hence, (20) has truth-conditions and is thus also a metafictional statement. But its status is somewhat particular since it does not say much, it is not very informative. It merely says that the name "Emma Woodhouse" does not refer. From this viewpoint, one can see why (20) is entailed by other metafictional statements. Indeed, fictional names are *de jure* empty, so the mere understanding of the name "Emma Woodhouse" in (17) triggers the corresponding existential (20). In other words, it is no use trying to look into the world whether Emma Woodhouse exists or not. It is part of the meaning of fictional terms as used in metafictional statements to be empty.[13]

[10]The expression is adapted from Plascencia's metafiction *The People of Paper*. It corresponds to what [Van Inwagen 1977] generically calls a "creature of fiction".

[11]For this reason, one might want to say that it is an "abstract artefact", following [Thomasson 1999]. I will present and discuss Thomasson's view in due time.

[12]In the sense given in [Ryle 1949].

[13]I will have some more to say about this inference at the end of chapter 18.

By contrast, (20) does not entail any other metafictional statement. Indeed, it could be, from the truth of (20) alone, that "Emma Woodhouse" is an ordinary empty name. So (20) is asymmetrically entailed by some metafictional statements.

It seems a good place to ground the distinction between metafictional and quasi-metafictional statements. Indeed, one would expect that the distinctive feature of fictional names, namely that they are *de jure* empty, gets transmitted to metafictional uses so that it grounds its specificity. Let us try out the idea.

Fictional names vs real names

The above inference seems to distinguish metafictional statements with quasi-metafictional statements involving real names like (18). Indeed, no such true sentence entail their corresponding negative existential:

(21) François Récanati does not exist.

On the contrary, *any* true sentence containing "François Récanati" will be a good reason to believe in the denial of (21). This is because the name "François Récanati" is a real name, hence no true sentence involving that name can entail that the name does not refer.

Fictional names vs ordinary empty names

As for quasi-metafictional statements involving an ordinary empty name like (19), it seems that one can follow almost the same line of reasoning.

Indeed, it seems that "being a hypothetical planet" entails "not existing". Hence, (19) seems to entail its corresponding negative existential:

(22) Vulcan does not exist.

But this inference is not *a priori*. For many "hypothesised" things were eventually found to exist. "Vulcan" is *de facto* empty, therefore one must check in reality whether there is such a thing as Vulcan before one can evaluate the truth of (22).

In other words, Vulcan has not introduced the name "Vulcan" to denote a *nonexistent* planet. When Le Verrier introduced the name "Vulcan" it was indeed a hypothetical planet, but this actually gave him good reasons to think that Vulcan existed. As a reminder: Le Verrier introduced the name "Neptune" in the same manner, as a mathematical hypothesis, few years before and, as it turned out, Neptune indeed exists. He was thus reasonably confident in Newton's theory of gravitation applied to celestial mechanics. It so happened that Newton's theory of gravitation applied to

celestial mechanics is inaccurate. In particular, in the vicinity of the Sun, Einstein's general theory of relativity is much more accurate than Newton's theory.[14]

Hence, it is not the case that the predicate "hypothetical" *a priori* entails "not existing". Rather, the reasoning from (19) from (22) is done either by empirical investigations or by showing the inaccuracy of Newton's theory of gravitation in the vicinity of the Sun. In other words, Le Verrier was not making a logical mistake when he was naming a hypothetical planet "Vulcan" *and* believing that this planet existed.

One can see how "Vulcan"'s situation is crucially different from that of "Emma". This is crucially the difference between *de jure* nonexistence and *de facto* nonexistence.

14.3.2 Complications

Necessary nonexistence

One can complicate matters when considering the following true sentence:

(23) Vulcan is a nonexistent planet.

This sentence *a priori* entails (22) given the meaning of "nonexistent". So it seems that we have found a quasi-metafictional sentence entailing its corresponding negative existential after all.

This example might sound controversial, for, as made explicit above, Vulcan was not introduced as a *nonexistent* planet. But one can find ordinary empty names which are introduced precisely to name a nonexistent. For instance Salmon thus introduces the name "Nappy":

> I hereby introduce the name "Nappy" to refer to the new emperor of France, whoever that might be, if there is one, and to refer to nothing otherwise. Take note: I do not introduce "Nappy" as a name for a particular fictional character that I just created. I am not storytelling and I am not pretending to use "Nappy" as a name of a person. Nor do I subscribe to any theory to the effect that France now has an emperor. Rather I introduce "Nappy" as a name for the actual present emperor of France, provided contrary to my every expectation that there presently is an emperor of France. Barring a fairly radical skepticism, we know that

[14] This is because the Sun's mass bends the space-time in a way that visibly affects its gravitational attraction of Mercury.

> there is no such person as Nappy. Nappy is not a fictional character, not a mythical character, not a fabrication, not a flight of fancy.[15]

Following this introduction, the following negative existential seems to be entailed *a priori* by Salmon's last two sentences:

(24) Nappy does not exist.

Here is Williamson's even more radical example of a necessary nonexistent individual:

> Let Nemo* be the possible person who would have resulted from the union of sperm cell S and some ovum O* other than O. Nemo and Nemo* seem individually possible but not jointly compossible; they could not both have existed. If the existence of a set requires the existence of its members, the set {Nemo, Nemo*} could not have existed.[16]

Given the fact that {Nemo, Nemo*} is *necessarily* nonexistent, it follows that the following negative existential is *a priori* entailed by Williamson's last sentence:

(25) {Nemo, Nemo*} does not exist.

These examples suggest that the logical space given above should be revised so as to distinguish between *contingently* empty names, *necessarily* empty names and *fictional* names. It is not clear how this three-fold distinction maps onto the previous *de facto* vs *de jure* distinction, though it is clearly more fined-grained. Since it seems that necessarily empty names are clearly not fictional names, it follows that the above criterion to distinguish between metafictional and quasi-metafictional statements is wrong.

Fictionalised names

When it comes to (23), though, I suggest the following story which opens to a side issue about the characterisation of empty names. "Vulcan", I claim, has been surreptitiously turned into a fictional name here. Indeed, in using the name "Vulcan" in (23), one does not use the name "Vulcan" in keeping with Le Verrier's introduction of the name. In other words, one does not try to refer to a real planet and fail to do so.

What is really happening with (23) is that one tells Le Verrier's historical feat in a fictional tone of voice. Hence "Vulcan" has been "fictionalised", so to speak.

[15] [Salmon 1998], p. 305-6.
[16] [Williamson 1990], p. 168.

So (23) is a metafictional sentence and it unsurprisingly entails its corresponding negative existential (22).

This observation suggests that ordinary empty names can easily *become* fictional names. It seems that when an ordinary empty name is known to be empty, there is a strong tendency to "fictionalise" it. A very straightforward fiction can be set up at the expense of the people who are mistaken about the term. Those who believe the empty name is full.

For instance, since we now know better than Le Verrier, it is easy to conceive of the Solar system as Le Verrier believed it to be, while knowing that it is not thus. The result is a fiction which can be shared overtly and is used to represent mistaken beliefs.

On the other way round, one might argue that it is difficult to conceive, in a thought experiment, of a use of an empty name which would resist such "fictionalisation". Indeed, each uncontroversial instance of empty names we have to consider are names which are known to be empty. In other words, unspotted empty names are just phenomenologically identical to real names.

This is why, I think, ordinary empty names get assimilated to fictional names so easily.

Dynamics of names

The dynamics of names seem to go even further than this. Indeed, it seems that one can also "fictionalise" real names. The paradigmatic case being the use of "Napoléon Bonaparte" in Tolstoi's *War and Peace* as mentioned in the beginning. From Tolstoi's story, one can thus produce true metafictional sentences involving the real name "Napoléon Bonaparte" like the following:

(26) Napoléon Bonaparte, as presented by Tolstoi in *War and Peace*, is purely fictional.

I think the same remarks as for (23) apply: the name "Napoléon Bonaparte" has been surreptitiously turned into a fictional name here.

Interestingly, my intuition is that (26) entails the following negative existential:

(27) Napoléon Bonaparte, as presented by Tolstoi in *War and Peace*, does not exist.

Though not the following:

(28) Napoléon Bonaparte does not exist.

These intuitions, I think, should be explained by this interesting dynamic "fictionalisation" of names suggested above. The detail of such a theory is not the concern us here, for it does not affect the fictional vs metafictional data at issue.

Conclusions

In trying to find distinguishing features of metafictional statements as opposed to quasi-metafictional statements, we failed to find necessary conditions. Some empty names yield quasi-metafictional statements which *a priori* entail their corresponding negative existentials without being fictional names. Consequently, everything I will say about fictional names should generalise to these necessary empty names.

I will focus on fictional names only in the following chapters. I think they are a good place to start, because they seem to be rather pervasive once we acknowledge that there is some dynamics of names. In my arguments, I will use the fact that fictional names are *de jure* empty. I do not know how such a premise would affect the generalisation to all kinds of empty names. I will thus remain cautious on the possibility of such a generalisation, for I do not know which parts of the following arguments should be modified and which parts are unaffected by my hidden assumption.[17]

14.4 Conclusions

The problem about fictional names is thus the following: how should one account for the two uses of fictional names? With a fictional name, one can construct fictional sentences with fictionality-conditions *and* metafictional sentences with truth-conditions. Pretence semantics is designed to handle the first kind of sentences but is helpless with the second kind.

If it is sufficient to have a fictional name to meet this problem, I also showed that it is not necessary. This suggested a more sophisticated logical space for empty names than the distinction between *de facto* and *de jure* empty names. Nothing will hinge on this sophistication in what I will say, for I will henceforth only talk about fictional names. This merely suggests that the problem I am working on is more pervasive than one wight have primarily thought. This suggestion is reinforced by the fact that (empty) names have dynamic features: they can get fictionalised.

[17]Especially because Salmon distinguishes necessary empty names with both fictional names and contingently empty names at the end of [Salmon 1998] from an ontological perspective. I do not know how his argument affects the linguistic problem, though.

The problem is a linguistic one, but it also points to an ontological problem. The linguistic problem consists in providing a theory of fictional reference which can account for the fictional vs. metafictional data. The ontological problem consists in explaining what is the ontological status of fictional characters. But the two kinds of statements suggest that there are two very different kinds of "fictional characters".

If one focuses on fictional statements only, one is led to conclude that fictional characters are nonexistent flesh and blood individuals which very much resemble real individuals. Emma Woodhouse, from this fictional perspective, is indeed "handsome, clever, rich" as anybody could be. By contrast, if one focuses on metafictional statements only, one is led to conclude that fictional characters are existent individual of papers which are something like abstract artefacts. Emma Woodhouse, from this metafictional perspective, is a literary figure which has been very influential in the history of English literature from its creation onward.

Still, we usually talk of fictional characters as *one*. So it seems that, at best, they have contradictory facets: one is the flesh and blood individual and the other the corresponding individual of paper. The ontological problem can be seen as trying to give an adequate description of the essential interaction between the two facets. It is a thorny issue. In this part, I will try to stick as much as possible to the linguistic problem in the hope that the ontological problem eventually gets clearer. However, ongoing debates mix the two problems very easily, as we will see.

I will provide a solution to the linguistic problem in this part, by providing a theory of fictional reference which is in keeping with anti-realism, i.e. the claim that fictional names never refer. The ontological consequence of such a view is that fictional characters do not exist in any sense.

This part will proceed in two steps. Chapter 15 and 16 should be seen as a critical part. In chapter 15, I will describe the ongoing debate between realism and anti-realism about fictional names, showing that they each champion one half of the data and both have trouble to extend their analysis to the other part of the data. In other words, realism is tailor-made for handling metafictional sentences and anti-realism is designed for fictional sentences but none efficiently solve their problematic data. At this point, one would naturally want to combine the strengths of both views into a single account. They are the so-called "hybrid views" on fictional names. In chapter 16, I will discuss two hybrid views and show that they are bound to fail. This will lead me to reconsider the linguistic data from the start.

By contrast, chapter 17 and 18 should be seen as a constructive part. In chapter 17, I will proceed to a thorough analysis of the distinction between the fictional and the metafictional. I will show that the characterisation in terms of truth-conditions and fictionality conditions is not enough to account for the data. Instead, I will

show that we need a notion of metaphysical perspective. I will define the notion by contrasting it with other notions of perspective one can find in narratology and in the visual arts. After applying the definition of a metaphysical perspective to linguistic fictions, I will show what the anti-realist needs to account for the whole data. Chapter 18 will then proceed in introducing free logic and combine a version of positive free semantics with pretence semantics. This combination yields a general account of fictional reference which is in keeping with anti-realism.

I should add that my aims in this part are really modest and local. I simply want to try out anti-realism and see how far one can go. Anti-realism holds that fictional characters do not exist, which is, I take it, a very reasonable view. But anti-realism about fictional names meets difficult objections so that it is not considered plausible in certain philosophical circles anymore. There are two challenges for the anti-realist which I try to address. The first one is to answer the "reluctant realists", those who started from anti-realism and were convinced by the realist's argument.[18] This is the critical part which aims not at destroying their position, but casting some doubts on their conclusion. The second one is internal to anti-realism. As we will see, it is quite ironic that anti-realists stumble over negative existentials (among other metafictional statements), for the core intuition of anti-realism is that *fictional characters do not exist*! In providing a semantic analysis of negative existentials involving fictional names which is in keeping with pretence semantics I show that, even if anti-realism might not be preferred at the end of the day, it can be made self-contained and coherent. This is what I called the constructive part, for lack of a better term.

[18]The expression is adapted from Putnam calling Quine a "reluctant platonist" about mathematical entities. For Quine started off as a champion of nominalism before accepting into his ontology numbers because of his indispensability argument.

Chapter 15

The debate between realism and anti-realism: a dead-end

Realism about fictional names is the view that fictional names refer. Realism thus claims that individuals of paper exist, and that fictional names refer to them. There are several versions of realism, according to which the precise ontological status of individuals of paper is made explicit.[1] But they all share the central tenet here given.

Anti-realism, by contrast, is the denial of realism. Anti-realism thus claims that fictional characters do not exist in any sense. In particular, they deny that individuals of paper exist. There are also different forms of anti-realism. I favour that of Everett which is combined with a functionalist account of fictional "truth" according to which talk of fictional character is a kind of pretence.

In this chapter, I will show how each position is tailor-made for each kind of uses of fictional names. As a result, they both have good grounds for development. But it also seems that they both fail to extend so as to account for the kind of use they are not "made for". Hence, both views are severely limited. It looks like this head-on debate is a dead-end.

One natural way out of this contradictory debate consists in trying to combine good insights from both views. This strategy is at the core of the so-called "hybrid views". I will discuss them separately in the following chapter.

First I will discuss realism, showing how they easily account for metafictional statements and extend to fictional sentences with difficulty. Then, I will discuss

[1]Neo-meinongian usually think of fictional characters as abstract objects akin to numbers, for instance [Parsons 1980] and [Zalta 1983]. Artefactualists, on the other hand, construe them as abstract artefacts, for instance [Van Inwagen 1977] and [Thomasson 1999]. Possibilists view them as concrete unactualised possibilia, for instance [Lewis 2015].

anti-realism, showing how they easily account for fictional statements and extend to metafictional sentences with difficulty. This picture of the debate should motivate one to send the two views back to back and consider the hybrid views as relevant alternative views.

15.1 Realism about fictional names

15.1.1 Realism and metafictional sentences

The argument in favour of realism

Realism about fictional names comes from the observation of metafictional sentences. Realists are convinced by the validity of the following simple argument:

- Metafictional statements are truth-evaluable statements with a fictional name in the subject place.
- The principle of compositionality requires that a name in the subject place of a truth-conditional statement refers.
- Therefore, fictional names do refer.

Since there are metafictional statements and they have truth-conditions, the argument, according to realists, goes through.[2]

Note on the metafictional data

Note that there are many kinds of metafictional statements. (17) is one kind of metafictional statements. We have also seen negative existentials which are also metafictional statements, for they are truth-evaluable statements with a fictional names in the subject position.[3] There are also other kinds of metafictional statements which have received much attention in the fictional literature, like comparatives and intentional statements:

[2]Note that this argument is one version of Quine's argument using the notion of "ontological commitment". [Thomasson 2003a] thus argues that, given literary theory is an empirical science, the Quinean ontological criterion is relevant to the debate between realism and anti-realism about fictional names. See also [Récanati forthcoming] (p. 2) for a more specific argument leading to artefactualism.

[3]Some people deny that negative existentials are statements of the subject-predicate form. I will ignore this complication for the moment. See chapter 18 for a critical exposition of these views.

(29) Emma Woodhouse is much happier than Emma Bovary.

(30) My neighbour is in love with Emma Woodhouse.[4]

There is no systematic, exhaustive typology for metafictional statements.[5] John Woods is, to my knowledge, the only one to have tried to provide one in [Woods 2018]. It will be useful to rely on his typology for further discussion. He distinguishes between three types of metafictional statements, with examples:

> *External sentences*: Sentences expressing the observations and speculations by *readers* about the goings-on in a story. ("Watson was the narrator of the Holmes stories" or "Almost certainly Holmes and Watson weren't lovers.")
>
> *Intensional sentences*: Sentences reporting relations in which *we ourselves* stand to the objects and events reported by [fictional] sentences. ("Agatha Christie admired Holmes more than any other detective.")
>
> *Cross-over sentences*: Sentences registering *cross-story comparisons*. ("Holmes was certainly more intelligent than Li'l Abner." "Superman was stronger than Mighty Mouse and might well have been stronger than Batman.")[6]

So, in principle, any kind of metafictional statement can be used to run the argument in favour of realism.[7] It suffices that one recognise at least one kind of metafictional statement to run the above argument. Since sentences like (17) are the least controversial kind of metafictional statements, it will be convenient to focus on them.

[4]This statement is meant to be akin to statements like "X pities Anna Karanina" or what Oscar Wilde puts into the mouth of Vivian in *The Decay of Lying*:

> One of the greatest tragedies of my life is the death of Lucien de Rubempré.

[5]Though some non-exhaustive typologies have been very influential. Especially that of [Van Inwagen 1977] and [Currie 1990].

[6][Woods 2018], p. 74.

[7]This point is made in [Everett 2013] §6.1 (pp. 120-138). He thus distinguishes between "three forms of argument for fictional realism" which are versions of the above argument relative to each kind of metafictional statements.

Realist account of metafictional statements

The semantic analysis of metafictional statements, according to realists, is thus straightforward. They are subject-predicate statements. The fictional name in the subject position refers to an individual of paper. The statement is true iff the individual of paper referred to falls into the extension of the predicate.

For instance, (17) is true iff the individual of paper the name "Emma Woodhouse" denotes indeed falls into the extension of "being a fictional character and main protagonist of Jane Austen's *Emma*". The same analysis applies, *mutatis mutandis*, to (29) and (30).

The main question raised by metafictional sentences for the realist is an ontological one. How should one think of these "individual of paper"? There are different ways of answering this question, yielding different forms of realism about fictional names.[8] I think Amie Thomasson's view is the most appealing and dominant view among realists today.

Thomasson introduced the ontological category of "abstract cultural artefacts". There are many kinds of abstract artefacts, among which are fictional characters. Here is, in her own words, a description of this view:

> I have argued in [Thomasson 1999] (chapter 1) and [Thomasson 2003a] that the best view of what fictional characters are (if there are any, as I think there are), is that fictional characters are abstract cultural artefacts, relevantly similar to other social and cultural entities including particular laws of state (the U.S. Constitution, the Miranda Laws), works of music (Nielsen's Symphony No. 4, Op. 29, "The Inextinguishable"), and the works of literature in which fictional characters appear (Tolstoy's *War and Peace*). These things are all abstract in the sense that they lack any particular spatio-temporal location, but unlike the Platonist's abstract entities, they are artefactual – created (not discovered) at a certain time, e.g. through the author's activities in writing a work of fiction, and are contingent (not necessary) entities that might have never been created.[9]

In [Thomasson 1999], she gives a book-length argument in favour of this realist view which is now called "artefactualism".

This view is particularly appealing in that it avoids the suspicion stemming from alternative neo-Meinongian views which accept "nonexistent objects" into one's on-

[8] Everett thus distinguishes between two brad categories in [Everett 2013], p. 139. His distinction is between the realists who take fictional characters to be nonexistent concrete objects, and those who take them to be existent abstract objects. I need not enter the detail here.

[9] [Thomasson 2003b], pp. 219-220.

tology.[10] Furthermore, it fits well our pre-theoretical intuitions about the fact that fictional characters, though not concrete things that one can handle and manipulate, are created by authors who are, in a very intuitive sense, craftspersons of cultural material.

I will keep using the expression "individual of paper" in order to avoid any kind of ontological commitment in the following. My presenting the difficulties of realism should thus apply to any version of realism.

15.1.2 Realism and fictional sentences

Driven by this success, realists have tried to extend their view to account for fictional sentences. They have thus tried to give a realist semantics for fictional discourse. That is a semantic analysis of fictional sentences which take for granted that fictional names refer.[11]

Distinction between two predication-like relations

The basic strategy is to distinguish between two predication-like relations. The first is called *exemplifying* and the second *encoding*.[12]

An individual is said to *exemplify* a property when the individual falls into the extension of the property, i.e. when the property is true of the individual in the usual sense. By contrast, An individual is said to *encode* a property when the property is part of what characterises the individual, i.e. what makes the individual this particular individual. But the property does not really apply to it. It is as if the property accompanied the individual, waiting to apply.

Reviewing the origin of this distinction will be useful to understand the distinction. It was first introduced to meet Russell's objections to Meinong's theory of object. In [Meinong 1904], Meinong defended the view that every set of properties should denote an individual. Consequently, any well-formed definite description should denote an individual. Individuals then fall into two broad categories: the existent ones and the nonexistent one, depending on how reality is.

[10] See [Parsons 1980] and [Zalta 1983] for such views.

[11] This view is thus contradicting the functional account given in part 1.

[12] This terminology comes from [Zalta 1983]. There are alternative terminologies: [Parsons 1980] distinguishes between "having" and "including" a property; [Mally 1914] distinguishes between "determining" and "satisfying"; [Castañeda 1975] between "internal" and "external" predication; [Rapaport 1978] between "constituency" and "exemplification"; [Fine 1982] between "implicit" and "explicit" predication; [Van Inwagen 1977] between "having" and "ascribing".

In [Russell 1905], Russell objected by considering the following definite description: "the existent golden mountain". The corresponding object should at least be, by Meinong's standard, existent, golden and a mountain. But as it so happens that there us no golden mountain in this world, it should be empirically true that the individual denoted by "the existent golden mountain" does not exist. This amounts to saying the following statement is true:

(31) The existent golden mountain does not exist.

But this is a contradiction. Therefore, Russel argues, Meinong's theory is false

But there is a way of saving Meinong from Russell's contradiction. One can distinguish between the two predication-like relations which can link an individual to a property. In particular, "the existent golden mountain" does not *exemplify* existence, though it *encodes* it. Indeed, the individual at issue here encodes three properties: "being existent", "being golden" and "being a mountain"; it exemplifies many properties, like "being nonexistent", "being Russell's example", etc.

Application to fictional characters

The distinction can arguably be applied to fictional characters, and thus to account for fictional and metafictional uses of fictional names. Metafictional statements like (17) are a clear case of exemplification. Hence the semantic analysis of metafictional statements like (17) is very simple and in keeping with the surface structure of the sentence.

Intensional and cross-over statements like (30) and (29) are analyzed in the same way as Russell's example given above. So Emma Woodhouse should encode what is fictionally "true" of her whereas she exemplifies "being happier that Emma Bovary" and "being my neighbour's favourite character". This naturally leads to treating fictional statements like (16) as pure cases of encoding. Fictional sentences are thus comparable to long definite descriptions which constitute the individual of paper called "Emma Woodhouse".

This solution is in keeping with realism, since the name "Emma Woodhouse" is here a real name which refer to an individual of paper. The only requirement is that an individual of paper be the sort of thing which can both encode and exemplify properties, which is a matter of definition.

Difficulties

The realist's account of fictional statements is thus utterly counter-intuitive. Indeed, the realist holds that an author, when they write, writes about an individual of paper,

i.e. an abstract object or a cultural artefact. It seems that Jane Austen describes a young woman, but, the realist holds, this is not the case. What Jane Austen *really* does is to construct an abstract artefact dwelling in this real world.

Moreover, the realist comes up with a very similar implausible story about what the reader does when they read. One would think that the reader imagines about a young woman. But, in fact, they are reading about an abstract artefact which encodes the fictional propositions the reader imagines. Consequently, if the reader is annoyed, or angry, or moved by a particular feat of Emma Woodhouse, they feel emotions, according to the realist, towards a real abstract entity of some kind. This is utterly implausible.

Though realist argue that they can account for both metafictional and fictional statements, their semantic analysis of fictional statements has some very implausible consequences one might want to avoid.

Aside on a technical difficulty: A specific difficulty comes from the semantics of cross-over statements. (29) seems to be a metafictional sentence, so the kind of predication involved should be that of exemplifying. Emma Woodhouse should then exemplify the property of "being happier than Emma Bovary". But how can one construct a property from encoded information?

Abstraction is the way one can extract a property from exemplified information. For instance, given by degree of happiness, one can abstract the predicate "being happier than Louis Rouillé". Using lambda notation, it suffices to define it in the following manner: $\lambda x.$"x is happier than Louis Rouillé".

One should have a story about how lambda abstraction works for encoded information though so as to get "being happier that Emma Bovary" from the encoded information, say, "Emma Bovary commits suicide". This is a technical difficulty which has been discussed at length among realists.[13] Of course, one cannot simply generalise lambda abstraction at the risk of facing Russell-like counterexamples like that of the "existing golden mountain". Existence, here, is encoded, but it should not be possible to abstract it so that it then becomes exemplified.

I think the difficulty generalises to all kind of reasoning combining information which is both encoded and exemplified. The distinction between exemplifying and encoding precisely consists in distinguishing between types of information so as to prevent mixing them together and having Russell-like counterexamples. But it also appears that we do mix the two types of information when we reason about fictional characters. For example, here is a piece of inference which sound valid:

[13]See for instance [Parsons 1980] (§6.4, p. 167-170) for a seminal discussion of this.

- It is fictional that a has the lower body of a horse and it is fictional that a has the upper body of a man.

$\therefore$ Therefore, a is not a realistic character.[14]

15.1.3 Conclusion about realism

I showed how realism about fictional names is tailor-made to account for metafictional sentences. But it appears that the realist account of fictional sentences leaves a lot to be desired. In particular, the anti-realist account of fictional sentences is much more natural and straightforward than the realist one.

At this stage, the anti-realist can call on simplicity and intuition to back up their position, while the realists might want to hold on the firm conviction that their account of fictional names, since they handle so easily metafictional statements, is worth a little sophistication. I will argue that looking at the debate from the anti-realist viewpoint, one ends in a similar dead-end.

15.2 Anti-realism about fictional names

15.2.1 Anti-realism and fictional sentences

The pretence-semantic account of fictional sentences

Fictional uses typically produce fictional statements. Pretence semantics, as developed in chapters 6 and 7, is tailor-made for providing the semantic analysis of such statements. Such an analysis involving fictionality-conditions instead of truth-conditions naturally extends to a theory of fictional reference.

Indeed, pretence semantics has it that the author, or producer, of the fictional statement does not describe facts. Instead, they pretend to describe facts. This idea of pretend description of facts yields a characteristic two-fold logical structure which we have discussed and defended. In short, the author pretends to be a narrator which either is linked to the fictional events or pretends to be linked to the fictional events.

This idea applies to fictional reference: fictional reference is pretend reference. As Récanati remarks:

> When the protagonists of the fiction are themselves fictional characters, the pretend assertion rests on an ancilliary act of *pretend reference*. One of the merits of the pretence approach is that it is compatible with

[14] This example is adapted from [Proudfoot 2006], p. 14.

> a demanding notion of reference, such as the view that genuine reference requires some form of acquaintance with the reference ([Récanati 2012]). On this view one cannot refer to something that does not exist, or of which one has only descriptive knowledge; but one can pretend to refer to such things, and that is what happens in fiction.[15]

For instance, here is how pretence semantics accounts for (16). Jane Austen pretends to be a narrator reporting on some fictional events they have come to know about (either because they have witnessed them or heard reliable testimony of the events, or because they are pretending to witnessing or reliably testifying about them). The narrator is thus using the name in such a way that they are linked by a relevant causal chain of reference. Fictionally, whatever reason they have to know about the events will suffice to make sure that there is such a chain. The opening sentence of *Emma* thus introduces a fictional name as one could introduce a real name in a non-fictional context, presupposing the fictional reference.[16]

Aside on the scope of fictional sentences

Following the same line of reasoning, Frege called fictional names "mock proper names":

> Names that fail to fulfill the usual role of a proper name, which is to name something, may be called mock proper names. Although the tale of William Tell is a legend and not history, and the name "William Tell" is a mock proper name, we cannot deny it a sense. But the sense of the sentence "William Tell shot an apple off his son's head" is no more true than the sentence "William Tell did not shoot an apple off his son's head". I do not say that this sense in false either, but I characterise it as fictitious.[17]

Frege's examples are interesting in that they are not direct quotations of a fictional text. Indeed, neither of Frege's sentences are to be found in the original legend of William Tell. But they correspond to what is usually called implicit fictional

[15] [Récanati forthcoming], p. 4.

[16] Note that pretence semantics very naturally explains why real names and empty names can be used on a par in a fictional context. In *War and Peace* by Tolstoi, it is fictionally the case that "Napoléon Bonaparte" and "Pierre Bezukhov" are names. Given the reality principle, within the fiction, "Napoléon Bonaparte" refers to the historical figure imported into the fiction while "Pierre Bezukhov" refers to a fictional character.

[17] [Frege 1980].

"truths", as defined in chapter 1. Given the fictionality-conditions extracted from the legend of William Tell, it follows that Frege's first sentence is fictional and the second is not.

This observation suggests that one should extend the class of fictional uses so as to encompass any sentence expressing a fictional proposition. Consequently, fictional uses are not restricted to which are direct quotations of fictional texts, like (16). Indeed, when one tells or re-tells the story, one is fictionally using the fictional name. The fictional use is deferred to the initial fictional telling produced by the narrator in the fictional text. This deferred use is naturally explained within pretence semantics: one extends the original pretence by continuing it after the reading is done.

Strawson makes this point very nicely:

> [Uses of fictional names] are themselves many and various; but we can say that many of them cluster round the central notion of telling, or re-telling, a story. The linguistic forms of these uses of language are taken over, complete, from the fact-stating uses; and the linguistic *functions* of the fact-stating uses are, as it were, reproduced, story-wise, within the story-telling uses. One of the functions there reproduced is that of identifying reference, often, though not only, performed by the use of names; only, of course, the reference is to characters in the story, not to people in the world.[18]

Acts of re-telling a story include *inter alia* summarizing a fiction and translating a fiction. Here is one example of a summarizing act based on Jane Austen's *Emma*: (16) is a fictional use of "Emma Woodhouse", here is an object-fictional use of it:

(32) Emma Woodhouse is an independent, wealthy woman who lives with her father in the English countryside near the village of Highbury.

15.2.2 Anti-realism and metafictional sentences

Driven by this success, anti-realist have tried to extend their view to account for metafictional sentences. They have thus tried to give an anti-realist semantics for metafictional discourse.[19] That is a semantic analysis of metafictional sentences which take for granted that fictional names do not refer.

[18] [Strawson 1967], p. 6.
[19] See especially [Walton 1990] (ch. 10), [Evans 1982] (ch. 10) and [Everett 2013].

Extended pretence

The basic idea is to introduce the notion of an "extended" pretence.[20] An extended pretence is a pretence which is parasitic on another pretence, called the "base" pretence. The extended pretence is a pretence, and thus comes with a set of principles of generation which in turn define "extended" fictionality-conditions.

To understand how this works, let us look at cross-over statements which are taken to be the most successfully accounted for using this idea of an extended pretence.[21] In order to understand (29), one has to "merge", so to speak, the two underlying, relevant fictions by Jane Austen and Gustave Flaubert. The merging intuitively consists in having the two Emmas meet in imagination and then compare which is happier, in this imaginative scenario.

This can be formalised in pretence semantics by opening a pretence in which both Emmas are characters. This pretence should have principles of generation according to which both Emmas are identical to what they are in their fiction of origin. Given a principle of reality straightforwardly applied, one understands what "happy" means in this pretence. In particular, since it is generally accepted as true in the real world that one who decides to get married is happier than one who commits suicide, it is fictionally so in the extended pretence. Emma Woodhouse, at the end of *Emma*, finally decides to get married while Emma Bovary, at the end of *Madame Bovary*, commits suicide. Hence, it will be fictional in the extended pretence that Emma Woodhouse is happier than Emma Bovary. Therefore, (29) is fictional in the extended pretence.

This strategy can handle all kinds of metafictional statements, so the anti-realist claims. Pretence semantics can thus yield fictionality-conditions for all metafictional statements using extended pretences.

Difficulties

But the anti-realist is only half-way through, so to speak. For they have not given *truth-conditions*, but merely fictionality-conditions in the relevant extended pretence. When it comes to cross-over statements like (29), this maybe heard as a good result. Indeed, I suppose many people will join the anti-realist intuition that Emma Woodhouse is not *really* happier than Emma Bovary: it is merely fictional, albeit neither *Emma*-fictional nor *Madame Bovary*-fictional.

[20] The term comes from [Everett 2013] (§3.3).

[21] See in particular [Crimmins 1998] for a seminal pretence-semantic account of cross-over statements.

But this consequence of anti-realism sounds utterly implausible when it comes to metafictional statements like (17). Indeed, such statement sounds true: Emma Woodhouse is *really* a fictional character. This is not merely fictional. As for negative existentials like (20), the situation for the anti-realist is quite ironical. On the one hand, the anti-realist is committed to the view that fictional characters do not exist in any sense so they should recognise as *true* that Emma does not exist, since she is a fictional character. On the other hand, given the treatment of metafictional statements as taking place in an extended pretence, negative existentials involving fictional names end up having no truth-conditions but only extended-fictionality-conditions. This strong tension clearly casts some doubts on the anti-realist's account of metafictional sentences.

At this point, there are two kinds of anti-realists. Those who bite the bullet and hold that metafictional statements do not have truth-conditions, despite appearances. Récanati calls them the "radical simulationists", quoting [Evans 1982] and [McDowell 1977] (I add [Everett 2013]). Those who take the worry seriously and try to fix it (he quotes himself and [Sainsbury 2018], I add [Walton 1990]). They have to give a story about how to extract truth-conditions from extended-fictionality-conditions.

Biting the bullet: Here is a quote from [Everett 2013] so as to illustrate what radical simulationism amounts to. Everett informally explains what the underlying extended pretence of metafictional statements like (17) looks like, before going on to formalizing the core principles of generation underling such pretences:

> I suggest that we see the deployment of such expressions as "fictional" as taking place within the scope of an extended pretence. [...] There is the distinction between things that really exist and those which only exist within the scope of the make-believe or fiction. [...] In the simplest cases when we want to articulate [this] distinction, we will engage in an extended pretence in which we pretend that our domain of discourse contains all those entities which occur within some fiction and that those entities are as they are characterised by that fiction. [...] Within this extended pretence, those entities which genuinely exist will count as having the property of being real and those which do not will count as having the property of being fictional.[22]

As noted above, although it delivers fictionality-conditions in keeping with intuition, it implies the implausible consequence that there is no metafictional statement which is true *simpliciter*.

[22][Everett 2013], p. 65-6.

Piggybacking: The second strategy consists in providing a mechanism to systematically convert fictionality-conditions into truth-conditions and show that this mechanism applies in this case. This mechanism has been called "piggybacking".[23]

Among others, Walton argued for such a strategy in [Walton 1993]. He came up with very insightful analogies with the conversational use of metaphors and other figurative expressions. It so happens that one often does not notice that such and such expression is not *literally* true, especially when using so-called "frozen metaphors" which are so well entrenched that they have become sayings and lose their literal meaning altogether. Examples of frozen metaphors are: "traffic jam" and "waves of immigrants", etc.

In these cases, Walton claims, there is an underlying pretence. But it is not used to pursue fictional "truths". Rather, the speaker uses the underlying pretence so as to interpret the sentence as if it described reality. As Walton puts it:

> The metaphorical statement (in this context) implies or suggests or introduces or calls to mind a (possible) game of make-believe. The utterance may be an act of verbal participation in the implied game, or it may be merely the utterance of a sentence that *could* be used in participating in the game. In saying what she does, the speaker describes things that are or would be props in the implied game. It may be possible in favorable cases to paraphrase what she says about them with reasonable fidelity. Typically, the paraphrase will specify features of the props by virtue of which it would be fictional in the implied game that the speaker speaks truly, if her utterance is an act of verbal participation in it.[24]

The crucial idea Walton suggests is to use principles of generation of the underlying pretence in the reverse direction, so to speak. It consists in imagining what the real world prop would have to be for the fictional proposition to be true *simpliciter*.

I think I have given a nice example of how this can work in practice at the end of my discussion about the great beetle debate, in subsection 13.2.2. The challenge is to give a specific account of what I called the "direction of fit" of imaginative scenarii: a world-to-fiction vs a fiction-to-world direction of fit. I think it is conceptually possible to theorise about this phenomenon. However, going down this route involves complexities one might want to avoid if possible.

To anticipate a little, I will propose a third route for the anti-realist in chapter 17 and 18. Drawing on this powerful idea of extended pretence, I will show how it can

[23]The terminology comes from [Richard 2000]. For seminal accounts of the same sort, see also [Evans 1982] (pp. 263-8), [Walton 1990] (chapter 10).

[24][Walton 1993], p. 46.

be used to give an analysis of metafictional statements up to the problematic ones. I will then claim that the problematic ones can be handle using a version of free logic and that this technical apparatus combines nicely with that of pretence semantics. Crucially, since negative existentials create an internal tension in the anti-realist's theory, I will naturally focus on them. The result of this should be that anti-realism is still a live, coherent and relatively elegant option.

15.3 Conclusions

Realism and anti-realism about fictional names are contradictory views about the nature of fictional names. Realists and anti-realists are thus bound to disagree and enter a rational debate. Since both views have been grounded on steady ground, the debate has been going on for a long time.

I showed how both views champion their favourite linguistic data. Realism gives a straightforward, elegant semantic analysis of metafictional statements. Anti-realism handles nicely fictional statements. However, when it comes to extending one account to cover the data they are not tailor-made for, I showed that things can get very complicated very quickly, for both views imply implausible consequences.

Consequently, both views stare stonily at each other. Each time one makes a move, the other is arguing against it on the ground of simplicity. But at the same time, given the internal coherence of each position, each camp is convinced that the core intuition which backs up their account is worth a little sophistication.

At this point, one might want to compromise and benefit from both the realist and the anti-realist's insights. Indeed, it seems that if one could articulate both views, then one would be able to account for both fictional and metafictional statements without engaging into any unwanted sophistications. This diplomatic line of reasoning was echoed by several philosophers in the literature. They put forward what can be called "hybrid accounts" of fictional reference. In the following chapter, I will present such views and express my doubts about their conclusions.

Chapter 16

Hybrid accounts of fictional names

Hybrid views on fictional names are views which combine insights from both realism and anti-realism about fictional names. Typically, one wants to defend that fictional names do not refer in fictional contexts so that the pretence-semantic account of fictional sentences can apply as it is. In metafictional contexts, by contrast, fictional names do refer to the individual of paper so that a straightforward semantic analysis of metafictional statements be available. The main problem for hybrid accounts is to explain how the facets of fictional names are to be articulated into a single theory.

I shall distinguish between two hybrid accounts. First the *ambiguity view*, according to which fictional names are ambiguous and the ambiguity is resolved in context. The second is the *polysemy view*, according to which fictional names are polysemous and one sense is selected in context. The first view is easily argued against and paves the way for the second more sophisticated view. The second view is more difficult to refute. I will offer an argument which will cast some doubts on the detail of the polysemy view. This will lead me to keep its core insight so as to revitalise anti-realism in the following chapters.

16.1 Against the ambiguity view

16.1.1 Presentation of the view

The idea is to think of a fictional name as two separate names, i.e. two separate lexical entries. The two names are thus homonyms, for they have the same phonological realisation. But, so the ambiguity view goes, they are accidentally ambiguous.

The ambiguity at issue here is thus lexical ambiguity. It is a well-known natural language phenomenon. Many words are thus ambiguous, i.e. accidentally homony-

mous. For instance, in English, "bat" is ambiguous: it can either denote the animal or the sports accessory. This leads to distinguish between two lexical entries for "bat".

According to the ambiguity view, when it comes to fictional names, one of these entries is that of an empty name and the other is that of a real name. In the context of a pretence, by definition, one is forced to resolve the ambiguity by opting for the empty name. By contrast, in a serious context, one opts for the real name. Since fictional statements are associated with a pretence context and metafictional statements with serious contexts, this view nicely unifies accounts for the linguistic data.

Here is how [Salmon 1998] presents the view:

> On this account, the name "Sherlock Holmes" is ambiguous. In its original use as a name for a human being – its use by Conan Doyle in writing the fiction, and presumably by the reader reading the fiction – it merely pretends to name someone and actually names nothing at all. But in its non-pretend use as a name for the fictional character thereby created by Conan Doyle, it genuinely refers to that particular artefact. In effect, there are two names. Though spelled the same, they would be better spelled differently, as "Holmes_1" for the man and "Holmes_2" for the fictional character. Neither names a real man. The latter names an abstract artefact, the former nothing at all.[1]

There are two linguistic arguments against such a view, which I will present now.

16.1.2 The argument from renaming

Suppose the ambiguity view is correct. So there are two names "Emma Woodhouse_1" and "Emma Woodhouse_2" which happen to be homonymous.

There are many examples of ambiguous real names. For instance, the real name "Aristotle" is ambiguous. Indeed it can be used in different contexts either to refer to the teacher of Alexander the Great or to the Greek shipping magnate.

In such typical cases of ambiguity, it is always possible to disambiguate by renaming the two people with non homonymous names. One would thus call the teacher of Alexander the great "Aristotle of Stagira" and the Greek shipping magnate "Aristotle Onassis". Once the renaming is done, one can substitute "Aristotle" in context in order to disambiguate. For instance, it is true that *Aristotle Onassis* had an affair with Maria Callas, but it is not so true of *Aristotle of Stagira*.

[1] [Salmon 1998], §4, p. 294.

What this renaming shows is that the two individuals are not systematically connected to each other. In particular, one could exist without the other. It is an *accident* that the name "Aristotle" has two distinct referents.

If "Emma Woodhouse" is ambiguous, the same strategy should be available.

Let us, for instance, rename the fictional flesh and blood individual "Emma 21 Woodhouse" and the individual of paper "Emma Paper Woodhouse" One can now substitute "Emma Paper" for "Emma Woodhouse" in metafictional statements. Consequently, one should say that Emma Paper is the main character in *Emma*, and not Emma 21. But this is really odd, for it seems that Emma 21 is the main character in *Emma*. Indeed, the main character in *Emma* is a young woman, handsome, clever and rich. Such renaming forces one to accept that the main character in *Emma* is an abstract object. This is absurd.

What is odd is that, contrary to the Aristotle case, we do have the intuition that Emma 21 and Emma Paper are systematically connected. In particular, it is not the case that one could "exist" without the other. By contrast, we have the intuition that Emma Paper emerges from metafictional considerations about Emma 21. This is one reason why the ambiguity view is wrong. The intuition hereby expressed drives the polysemy view which we will shortly discuss.

16.1.3 The argument from anaphora and co-predication

There are statements which involve both fictional and metafictional predication and only one name. These come in two categories: one can use an anaphoric pronoun, or directly co-predicate of the fictional name. Here are examples:

(33) Emma Woodhouse$_i$ is a fictional character created by Jane Austen. She$_i$ is handsome, clever, and rich, with a comfortable home and a happy disposition.

(34) Emma Woodhouse, handsome, clever, and rich, with a comfortable home and happy disposition, is a fictional character created by Jane Austen.[2]

Genuine lexical ambiguity typically forbids such complex constructions, as shown by the following ungrammatical sentences:[3]

[2]If you doubt that my examples are "natural", here is the first paragraph of the Wikipedia entry "Emma Woodhouse":

> Emma Woodhouse is the 21-year-old protagonist of Jane Austen's novel *Emma*. She is described in the novel's opening sentence as "handsome, clever, and rich, with a comfortable home and a happy disposition." Jane Austen, while writing the novel, called Emma, "a heroine whom no-one but myself will much like."

[3]These tests originate in [Zwicky and Sadock 1973].

(35) * The bat$_i$ was sleeping upside down and Bob put it$_i$ in his bag to go play baseball.

(36) * Bats are mammals used to play baseball.

Hence, the fact that (33) or (34) are grammatical sentences shows that "Emma Woodhouse" is not lexically ambiguous as the ambiguity view predicts.

16.1.4 Towards the polysemy view

The above linguistic observations suggest that fictional names do not get associated with two unrelated lexical entries. By contrast, one should hold that the different uses of a fictional name are systematically related to each other. This leads to the idea of polysemy.

Here is how François Récanati comments on this linguistic data in [Récanati 2018]:

> For anaphora to be possible, it is enough if the referent of the name and that of the pronoun are so closely related that application of the same lexical item to both is acceptable. [...] [T]he fictional character referred to by its name is closely related to the flesh and blood individual portrayed in Doyle's stories – it is that individual *qua* fictionally portrayed so [...] the same name "Sherlock Holmes" can be used for both. That is arguably also what explains the possibility of anaphora [...], despite the difference between the fictional character targeted by the metafictional utterance and the flesh and blood individual targeted by its [fictional] continuation.[4]

16.2 In arguing against the polysemy view

16.2.1 Basic idea

The polysemy view consists in saying that both uses of the fictional name are systematically related. Intuitively, it seems that the metafictional use "comes after" or "derives" from the fictional use, in the sense that a fictional name, by definition, originates in a fiction and can *later* be used to produce metafictional statements. From an ontological viewpoint, it is as if the individual of paper's coming into existence supervened on the fictional flesh and blood individual. As Récanati points out in

[4] [Récanati 2018], §5.

the above quotation, once the pretence is shared and several people have imagined the flesh and blood individual in the way the fiction required them to, then they can refer back to the individual of paper which is the fictional flesh and blood individual "*qua* fictionally portrayed so".

This view was first defended by Kripke in his second John Locke Lectures, published as [Kripke 1973/2013]. Peter van Inwagen also defended such a view on independent grounds in [Van Inwagen 1977].[5]

The view is described in detail in [Salmon 1998]:

> According to Kripke, as the name "Sherlock Holmes" was originally introduced and used by Conan Doyle, it has no referent whatsoever. It is a name in the make-believe world of storytelling, part of an elaborate pretense. By Kripke's lights, our language licenses a certain kind of metaphysical move. It postulates an abstract artefact, the fictional character, as a product of this pretense. But the name "Sherlock Holmes" does not thereby refer to the character thereby postulated, nor for that matter to anything else, and the sentences involving the name "Sherlock Holmes" that were written in creating the fiction express no propositions, about the fictional character or anything else. They are all part of the pretense, like the actors' lines in the performance of a play. It is only at a later stage when discussing the fictional character from the standpoint outside the fiction, speaking about the pretense and not within it, that the language makes a second move, this one semantical rather than metaphysical, giving the name a new, non-pretend use as a name for the fictional character. The language allows a grammatical transformation, says Kripke, of a fictional name for a person into a name of a fictional person. Similarly, van Inwagen writes, "we have embodied in our rules for talking about fiction a convention that says that a creature of fiction *may* be referred to by what is (loosely speaking) 'the name it has in the story'" (cf. [Van Inwagen 1977]).[6]

According to the polysemy view, shared (linguistic) pretences essentially involve two steps. There is a "metaphysical move", in which an individual of paper is "postulated". Later on, there is a "semantic move" which enriches the meaning

[5] Quotes from [Searle 1975] and [Schiffer 2003] in [Récanati forthcoming] show that these authors take the polysemy view for granted. Note that the view is called the "ecumenical view" which he defines as "accepting there are these two types of use of fictional names, and considers the fictional use as basic".

[6] [Salmon 1998], p. 294.

of the fictional name in such a way that the fictional name can thereby be used as a real name to refer to the individual of paper. Eventually, the fictional name gets two closely connected senses, hence the *polysemy* view.

This view is hybrid between anti-realism and realism, performing a sort of *Aufhebung*. It is precisely because one pretends to refer with a fictional name (the basic tenet of anti-realism) that one later succeeds in referring using the same name name (the basic tenet of realism). There is dynamic. From an ontological viewpoint, something was brought into existence; from a semantic viewpoint, the name metamorphosed.

16.2.2 Detail of the view

In this subsection, I will present one version of the polysemy view which is more detailed.[7] Applying the notion of polysemy to fictional names is done in a two-step fashion.

These two steps do not correspond to the description of the view given above. First, I will define a notion of polysemy which can be applied to fictional names, calling in linguistic works on lexical polysemy. Second, I will say how this can apply to fictional names so has to yield the polysemy view.

Polysemy and dot-objects

First, one should observe that the linguistic phenomenon of polysemy very commonly challenges metaphysical categories. Indeed, one lexical entry can be used to denote several things which usually fall into separate ontological categories. For instance the word "book" can be used to denote a physical object in certain contexts, and some information in another context. Similarly, the word "lunch" can be used to denote some food in certain contexts, and an event in another context.

It would be a mistake to think that the words "book" and "lunch" are lexically ambiguous so as to postulate *two* lexical entries for these words. Indeed, anaphoric sentences and co-predication sentences with such words are perfectly grammatical:

(37) The lunch$_i$ was delicious but it$_i$ took for ever.

(38) The lunch was delicious and took for ever.

(39) This book$_i$ is boring and it$_i$ doesn't even fit in my bag.

(40) This book is boring and doesn't even fit in my bag.

[7] The view presented here has been defended from different perspectives in [Rouillé 2015], [Récanati 2018] and [Terrone 2017].

This linguistic data shows that these words are not ambiguous, but polysemous. They have interconnected meanings which cannot be separated into different lexical entries.[8]

To account for this pervasive kind of lexical polysemy, [Pustejovsky 1998] introduced the notion of a "dot-object".[9] A dot-object is a complex entity involving different facets which can fall into distinct metaphysical categories. For instance, it is usually acknowledged that physical things and informational content are very different kinds of objects. Indeed, the mereological structure of physical objects and informational content is very different. *Books*, however, are complex objects which contain or articulate *both* physical things and informational content. As for *lunches*, they are complex objects containing or articulating *both* physical things (the food eaten) and a social event.

These complex objects are called "dot-objects" for Pustejovsy introduced an operation to compose simple kinds of object symbolised by a dot. The lexical entry associated with, say, "book" thus contains as a type: $physical\ object \bullet informational\ content$. These types should be seen as syntactical constraints, akin to theta-roles in generative grammar, at the lexical level. Hence the name of the discipline in which dot-objects are discussed: the generative lexicon. The introduction of dot-objects explains the grammaticality of the above sentences.

Dot-objects were introduced for many other purposes. For instance, one can individuate complex objects in different manners. Suppose that there are four copies of Plato's *Sophist* on a shelf. One can either say that there is only one book on the shelf if one focuses on the informational facet of the dot-object "book", or say that there are four books on the shelf if one focuses on the physical facet of the dot-object "book".

Here is the rendering of the notion of dot-objects in [Récanati 2018]:

> According to [Pustejovsky 1998], when a polysemous word applies to two closely related entities, as in the "lunch" example, the two senses which the word can take can also combine so as to yield a third sense, which encompasses the first two. In such cases, "the concept referred to by a lexical item... is the cluster of the two individual types along with the dotted type" ([Pustejovsky 1998], p. 94).[10]

[8] This presupposes that all words involved in anaphoric and co-predication statements are polysemous. However, it does not follow that all polysemous words give rise to such statements. I simply focus on one kind of polysemy here, for it is the one that matters for fictional names.

[9] He was later followed by [Chomsky et al. 2000], [Asher 2011], [Cooper 2007], [Luo 2012], [Gotham 2016].

[10] [Récanati 2018], p. 11.

Now, given that anaphoric and co-predication sentences using fictional names like (33) and (34) are possible, this suggests that "fictional characters are natural candidates for the status of dot-objects."[11] Indeed, given what has been said above, fictional characters look like a dot-object containing at least two facets, one facet corresponding to the fictional flesh and blood individual and one facet corresponding to the individual of paper.

Fictional characters as dot-concepts

However, one should observe that there is a crucial difference between fictional characters and other dot-objects. As Récanati puts it:

> Of course there is something peculiar about fictional characters construed as dot-objects: the flesh and blood individual Sherlock Holmes, who is supposed to be a facet of the fictional character Sherlock Holmes (the other facet being the cultural artefact), does not exist![12]

At this point, one could engage in arguments on the ontological nature of dot-objects. But Récanati, following Chomsky and others, rather takes a cognitive turn which both avoids the ontological discussion and seems to be particularly well-suited to fictional characters:

> [Here is] Chomsky's cognitive approach: dot-objects have no metaphysical reality; it is the mind that puts distinct things together under a single heading for purposes that have nothing to do with metaphysics or science. From this standpoint the proper object of study should be dot-*concepts* (e.g. the concept of lunch, [...] or the concept of fictional character) rather than dot-objects; and the study of dot-concepts should be carried out without presupposing that there is some one thing in the world these concepts are concepts of.[13]

Récanati then builds on previous work done in [Terrone 2017] and gives a detailed description of what dot-concepts are in the mental-file framework. Dot-concepts should be viewed as two separate mental files, the *fiction file* and the *source file*, which store information respectively about the fictional flesh and blood individual and the individual of paper.[14] The two files are essentially linked to each other, since the source file, by definition, contains a pointer to the fiction file.

[11] [Récanati 2018], p. 11.
[12] [Récanati 2018], p. 11.
[13] [Récanati 2018], p. 13.
[14] This is Terrone's "Two-fileness hypothesis".

16.2.3 Argument against the polysemy view

It seems to me that the difference between fictional names and other polysemous words like "book" and "lunch" associated with dot-concepts is worse than Récanati suggests. In other words, even dot-*concepts* are challenged by fictional names.

Dot-concepts are introduced to describe a particular cognitive phenomenon. It is a result of "the mind that puts distinct things under a single heading for [idiosyncratic purposes]". These idiosyncratic purposes challenge do not abide by the constraints of science or rational metaphysics. In particular, they do not presuppose that "there is some one thing in the world these concepts are concepts of." Let us call this *cognitive polysemy*.

But the kind of polysemy we are looking for when looking at fictional names is not cognitive polysemy. Indeed, a fictional name is supposed to be polysemous between an empty name and a real name. It is not about how we conceptualise the world, but how we categorise words. For this reason, I think we should call it a metalinguistic polysemy.

It resembles, for instance, the polysemy of a word which can be both a noun and a verb, depending on the context. For instance the word "phrase", in English, can be used as a noun and a verb: it either means a unit of speech or to express oneself using words. Of course, these two senses are systematically related. Depending on the syntactic environment, the word "phrase" will be either used as a noun or a verb. This is an instance of metalinguistic polysemy which is not straightforwardly cognitive polysemy.

So a fictional name like "Emma Woodhouse" should be akin to "phrase". According to the ambiguity view, it should either be an empty name (in fictional contexts) or a real name (in metafictional contexts). But, appealing to the context does not help in this case. For the context cannot change the fact that a name *actually* is empty or real. In a pretence context, one can make as if the name was real although it really is empty. But it does not follow that the name is actually real. On the other way round, one could make as if a real name was empty. But it does not follow from this kind of pretence that the name is actually empty.

This does not amount to saying that names are, in essence, univocal. For it may very well be that one holds the view that bodies and mind are distinct entities, but one can use the same name to denote either the mind or the body of a person. It means that a name can be either real or empty, but not both, which is a matter of course.

This shows that the dot-concepts are not on the right track. They miss the metalinguistic feature of the debate between realism and anti-realism about names and thus try to combine contradictory features into one single name. Part of the

debate between realism and anti-realism is linguistic. And no sophistication of one's concepts can resolve the fact that the debate between realism and anti-realism is (partly) a contradictory debate about whether fictional names are *actually* real or empty.

16.2.4 Counterproposal

I think it is useful to go back to the two "moves" of the polysemy view (as described in subsection 16.2.1). The first move is "metaphysical": it is the emergence of the individual of paper over and above the flesh and blood one. The second is "semantic": it is the "grammatical transformation, says Kripke, of a fictional name for a person into a name for a fictional person".

I think the metaphysical move corresponds to a familiar experience.[15] As such, it clearly points towards interesting phenomenological data. Probably, I would say that Terrone and Récanati's mental-file story tackles this phenomenological data.

However, I do not think there are compelling reasons to think that the move is metaphysical in the strong sense. Anti-realism is precisely the view that no such emergence takes place. The reason that is usually advocated for such a metaphysical move is that one should make possible the second "semantic move".

I do not think that names can undergo grammatical metamorphoses. I think there is a much more reasonable story to tell here. What happens is that a new empty name is introduced into (ordinary) language. First, a fictional name like "Emma Woodhouse" is introduced within the pretence and everyone pretends it refers to the flesh and blood individual in the fiction. Second, one recognises that "Emma Woodhouse" is an empty name (because it originates in a fiction) and then use it as such in non-fictional discourse. "Emma Woodhouse" is empty all along.

This counterproposal is in keeping with anti-realism about fictional names. What is true about the polysemy view is that there are two uses of fictional names and that there is dynamic. The metafictional use "comes after" the fictional use. But what is false is that the dynamic is a metamorphosis of an empty name into a real one.

[15] Calvino makes a very similar observation from a completely different viewpoint in [Calvino 1988] ("Quickness"), describing the phenomenon as "magic":

> I would say that the moment an object appears in a narrative, it is charged with a special force and becomes like the pole of a magnetic field, a knot in the network of invisible relationships. [...] We might even say that in a narrative any object is always magic.

What the counterproposal needs is a theory which says how one can use an empty name like "Emma Woodhouse" in a subject-predicate statement and get a true proposition. "Reluctant realists" say this is impossible. I will show in chapter 18 that it is possible, provided one uses free logic.

This counterproposal is a way of taking the good insights from the polysemy view so as to revitalise anti-realism.

16.3 Conclusions

In this chapter, I focused on what I called the hybrid accounts of fictional names which duly note that the debate between realism and anti-realism produces dead-ends. The core motivation of hybrid accounts consists in trying to combine good insights from both realists and anti-realists. The challenge for such views is to explain how these insights should combine into a unified theory of fictional names.

The ambiguity view says the combination should appeal to a notion of lexical ambiguity. Against the ambiguity view, I argued that positing two homonymous fictional names is utterly implausible because (i) the usual renaming method is not available and (ii) grammatical anaphoric and co-predication sentences involving fictional names are available.

By contrast, the polysemy view in an improvement over the ambiguity view. It starts from the linguistic data involving anaphora and co-predication, and claims that the underlying phenomenon has to do with polysemy. This idea is both appealing and suggestive. In order to make things more precise, Récanati and others used some interesting work on lexical polysemy to argue that fictional names should be associated with dot-concepts, i.e. complex mental representations articulating mainly two facets. The two facets are mental representations of the two (pseudo)referents of fictional names.

I cast some doubts about this picture. I remarked that the kind of polysemy at stake when it comes to fictional names is special in that it contains a metalinguistic component. Indeed, the two alleged senses of a fictional name correspond to both tenets of realism and anti-realism: either the name is a real name or it is empty. This cannot be captured using dot-concepts, so it seems that the polysemy view went wrong.

However, I acknowledged that the polysemy delivers a good insight, namely that there is dynamic involved. A fictional name which is *first* introduced in fiction, *later* can be used outside fiction. This insight can be used to revitalise anti-realism. It suffices to say that when the fictional name is *later* used, it is empty (contra the

polysemy view). This revitalisation of anti-realism is the subject of the next two chapters.

In chapter 17, I will go back to the data and give a proper analysis of it, given this dynamic feature of fictional names. I will show that behind the distinction between the fictional and the metafictional hides a three-fold distinction between an internal, external and mixed perspective. This will circumscribe the really problematic statements for anti-realism: I will emphasise the role of negative existentials in this light. In chapter 18, I will give a semantic analysis of negative existentials which is in keeping with anti-realism, using free logic. Finally, I will show how one can combine pretence semantics with a version of positive free semantics. This will show that one can give a fully anti-realist account of all kinds of metafictional statements.

Chapter 17

Internal, external and mixed perspectives

Let us go then, back to the distinction between the fictional and the metafictional.

The central idea of this chapter is that the distinction which is revealed by the fictional vs metafictional linguistic data has been hidden behind the distinction between fictionality and truth (as discussed in part 1 of this dissertation). Consequently, philosophers have tried to reduce the fictional and metafictional to the distinction between pseudo-assertion and assertion.[1] In the first part of this chapter, I will rehearse this reduction.

Although I think the distinction between pseudo-assertion and assertion is real and important to account for, I will argue that one needs more fine-grained distinctions. I will show that the distinction between assertion and pseudo-assertion does not square well with the distinction between reference and pseudo-reference, which is at the core of an anti-realist's theory of fictional reference. This will lead me to a fine-grained distinction between three kinds of statements involving fictional names: the internal, external and mixed perspective statements.

In order to theorise about perspective, I will step away from the linguistic data and consider the visual arts for a moment. I will show how one can define a notion of perspective which can then subdivide into the three categories we need consider. I will finally claim that the notion of perspective taken from the visual arts can be applied to linguistic fiction without problem. This will conclude the analysis of the

[1] I will use the term "pseudo-assertion" throughout. I take expressions like "pretend-assertions", "mock assertions", "make-believe assertions", "fictional assertions", "quasi-assertions" to be synonymous here. I will use the expression "serious assertion" as meaning the same as "assertion". The former will be used to mark the contrast with pseudo-assertion in context.

distinction between the fictional and the metafictional. I will show how anti-realism has problems with only a few kinds of metafictional statements. These will require a special treatment which is the subject of the following chapter.

17.1 Pseudo-assertions and assertions

17.1.1 The Fregean origin

Frege's thesis is that there is a crucial difference between genuine assertions and what he calls "mock assertions". He introduces them in the following passage:

> Instead of speaking about fiction we could speak of "mock thoughts". Thus, if a sense of an assertoric sentence is not true, it is either false or fictitious, and it will generally be the latter if it contains a mock proper name. [Footnote: We have an exception where a mock proper name occurs within a clause of indirect speech.] Assertion in fiction are not to be taken seriously, they are only mock assertions. [...] A work of fiction is not meant to be taken seriously at all: it's all play. [...]
>
> The logician does not have to bother with mock thoughts, just as a physicist, who sets out to investigate thunder, will not pay any attention to stage-thunder.[2]

According to Frege, the semantic analysis of assertions requires a truth-conditional apparatus, while that of pseudo-assertions falls outside the scope of the truth-conditional apparatus. In other words, assertions express propositions which are either true or false, depending on the worldly facts they aim at representing. Pseudo-assertions, on the other hand, express propositions which do not aim at representing worldly facts.[3] But, as I have shown in detail in chapter 5 and 6, it does not mean that pseudo-assertions are completely disconnected from worldly facts. They are tied up to reality in a convoluted manner, through the functioning of props within a shared pretence. This distant correspondence to real facts grounds the possibility of giving systematic fictionality-conditions for fictional sentences. While assertions come with truth-conditions, pseudo-assertions come with fictionality-conditions.

This "distance" comes from pretence. As Frege puts it, if pseudo-assertions do not have truth-conditions, it is because "it's all play". Pseudo-assertions are not

[2] [Frege 1980].

[3] The word "proposition" should be understood broadly here. The caveat made in chapter 1 about this is in order.

meant to be "serious" assertions. By contrast, they are the rehearsing of a serious assertion, i.e. the *staging* of a serious assertion. Frege uses a very striking analogy: pseudo-assertions are to assertions what stage thunder is to real thunder. Another comparison would be: what all-in wrestling is to boxing. Only the latter really hurts, despite appearances.

Consequently, according to Frege's insight, a theory of pseudo-assertion is the combining of two independent theories. One needs a theory of assertion and a theory of pretence. Frege makes it clear in the quote that he aims at giving the former only (just like the physicist studies real thunder without taking care of stage thunder). But, pretence is also a natural, albeit mostly social, phenomenon. So one can give and independent general theory of pretence.[4]

17.1.2 Theorizing pseudo-assertion

Searle's view

A next step in this Fregean direction comes from [Searle 1975]. His interpretation of the distinction between pseudo-assertion and assertion is very clear. It is expressed in his theory of speech acts.

Assertions, according to Searle, are rule governed-actions. Pseudo-assertions are derivative assertive speech acts which essentially consist in overtly breaking the rules governing assertions. This "breaking the rule" means cancelling the *semantic* rules of assertion by a *pragmatic* mechanism. Consequently, pseudo-assertions are indeed speech acts which derive from assertions: they are essentially dependent on the possibility of asserting.[5]

Here is how Searle presents this general picture:

> I find it useful to think of [the rules for assertion] as rules correlating words (or sentences) to the world. Think of them as vertical rules that establish connections between language and reality. Now what makes fiction possible, I suggest, is a set of extralinguistic, non semantic conventions that break the connection between words and the world established by the rules [for assertion]. Think of the conventions of fictional discourse as a set of horizontal conventions that break the connections established by the vertical rules. They suspend the normal requirements established

[4]Walton's theory of *mimesis* is a case in point. From this perspective, one can see that Walton's work is in line with Frege's thesis.

[5]This is what is called the "asymmetric dependency thesis about fiction" in [Récanati forthcoming] (p. 4), which Récanati defends against Matravers attack in [Matravers 2014].

> by these rules. Such horizontal conventions are not meaning rules; they are not part of the speaker's semantic competence. Accordingly, they do not alter or change the meaning of any of the words or other elements of the language. What they do rather is enable the speaker to use words with their literal meanings without undertaking the commitments that are normally required by those meanings. My [...] conclusion then is this: the pretend illocutions which constitute a work of fiction are made possible by the existence of a set of conventions which suspend the normal operation of the rules relating illocutionary acts and the world.[6]

Pretence, for Searle, is really the "suspension" of semantic rules for assertion. This view is explicitly inspired by Coleridge's famous phrase: "that willing suspension of disbelief which constitutes poetic faith".[7]

As for the rules for assertion, Searle refers back to his [Searle 1969]. With a little simplification, one can say that, according to Searle, assertions are governed by the following rules:

1. *The essential rule*: the utterer of an assertion commits themselves to the truth of the expressed proposition.

2. *The preparatory rules*: the utterer must be in position to provide evidence or reasons for the truth of the expressed proposition.

3. *The non-triviality rule*: the expressed proposition must not be obviously true to both the speaker and the hearer in the context of utterance.

4. *The sincerity rule*: the utterer commits themselves to a belief in the truth of the expressed proposition.

These correspond to what Searle calls the "vertical rules". Producing a pseudo-assertion consists in overtly violating these four rules.

Generalizing Searle's view

Following Frege's insight, Searle's view of assertion does not rely on any *specific* theory of assertion. It only requires that asserting is a *rule-governed* activity, or that there are constitutive norms for assertions. Such a requirement is a very common

[6] [Searle 1975], p. 326.

[7] The notion of a "play" is also present in Searle's article, just like in Frege. Searle, however, points towards Wittgenstein's notion of "language game" to express this point which may not be what Frege had in mind.

assumption in contemporary debates about assertions.[8] But the rules for assertion need not be Searle's.

In more recent debates, other norms for assertions have come to the attention. For instance, [Williamson 2002] argues that asserting is governed by a "knowledge rule" according to which:

> (K-A) One must: assert p only if one knows p.

This view has undergone many refinements. For instance, [García-Carpintero 2004] argues for the following rule which insists that the knowledge at issue should be *transmitted* in assertive speech acts:

> (TK-A) One must: assert p only if one's audience comes thereby to be in the position to know that p.

Alternatively, others argue that knowledge is too strong a requirement, and prefer to use the notion of belief. For instance, [Bach 2008] defends the norm:

> (B-A) Assert only what you believe.

This view has also been refined, for many authors have argued that a relevant notion of *justification* should be added to that of belief. The belief norm becomes:

> (JB-A) Assert only what you believe with justification.

Any of these rules (among others) can be "broken" in Searle's sense. Let us call *assertive norm* one's favourite constitutive norms for assertion. Let us call *assertive force* the being in force of an assertive norm. A *pseudo-assertion*, in general, is a speech act whose assertive force is cancelled by a pragmatic mechanism.

There are several relevant force-cancelling mechanisms which yield pseudo-assertions. Fiction, by definition, is such a mechanism. Among other candidates, supposition, irony, quotations are generally taken to be such mechanisms too. Here is how Récanati efficiently puts it:

> Pretend assertions (such as those made by the author of a fiction) mimic genuine assertion. Whichever norms regulate genuine assertions are tacitly invoked in pretend assertions: whoever pretends to assert pretends that the normative conditions of genuine assertions are satisfied.[9]

[8] See for instance Pagin, Peter, "Assertion", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/assertion/.

[9] [Récanati forthcoming], p. 4.

Pseudo-reference

There is a similar distinction at the level of reference. In order to make the parallel plain, one can think of reference as a rule-governed activity. For instance, one can opt for a causal theory reference norm according to which there must be a causal chain which relates the act of reference (for instance the uttering of a name) to the referent. One can thus make explicit the following rule of reference, according to causal theories of reference (CR):

> (C-R) One must: refer to i using the name "i" only if there is a causal chain which relates i to one's utterance of "i".

Having one's preferred constitutive rule for reference, one can define what pseudo-reference in a Searlean manner. It is the overt breaking of the rule by a pragmatic mechanism. Among other contexts, fiction, by definition, should count as a relevant pragmatic mechanism for the cancellation of reference.[10]

As Frege notes, it seems that pseudo-reference can appear in a serious assertion in "a clause of indirect speech". This also arguably happen in some metafictional contexts. This is the reason why one cannot conflate the two distinctions into one. I will show this in the following section.

17.1.3 Shortcomings of the reduction

The reduction

At this point, one can be tempted to identify the two distinctions. We thus have the following reduction:

Truth-conditions (serious assertion)	**Fictionality-conditions** (pseudo-assertion)
Serious reference (to the individual of paper)	**Pseudo-reference** (to the flesh and blood individual)
Metafictional discourse	*Fictional discourse*
(17) Emma Woodhouse is a fictional character and main protagonist of *Emma*.	(16) Emma Woodhouse handsome, clever, and rich seemed to unite some of the blessing of existence.

Table 17.1: Reduction of the distinction

[10] See [Donnellan 1975] for a detailed view of this cancellation mechanism at the level of reference using something like (C-R) for serious reference. He calls this mechanism "ending in a block".

More fine-grained distinctions

However, it can be argued that the above reduction too rigidly ties up assertion and reference together so that they must both shift to pseudo-acts when the context is fictional and both stay serious when the context is metafictional. Indeed, it seems that some metafictional statements involve pseudo-reference within an overall serious, truth-evaluable assertion.

For instance, intensional statements seem to be a case in point. Take the following example:

(30) My neighbour is in love with Emma Woodhouse.

In this case, it seems that "Emma Woodhouse" cannot refer to the individual of paper. Indeed, what my neighbour likes about Emma is not primarily related to the fact that she does not exist, that she has been created by Jane Austen and first appeared in print in 1815. On the contrary, my neighbour loves the flesh and blood individual: the handsome and clever young woman, the way she thinks, talks, etc. At the same time, it seems that (30) is truth-evaluable, since my neighbour is either really in love or not in love: this is, in principle, a fact in reality. If this description is correct, then one needs to make room for a serious assertion which contains an act of pseudo-reference.

There are arguably many other types of metafictional statements which involve this mixture of pretence and seriousness. For instance, Récanati made a strong case for parafictional statements in several places which is worth considering. Parafictional statements are fictional statements prefixed by a sentential operator of the form "In the relevant fiction / According to the relevant fiction". This prefixing seems to yield a truth-evaluable statement about the flesh and blood individual: Here is how parafictional statements are described in [Récanati forthcoming]:

> Parafictional discourse turns out to be intermediate between fictional discourse and metafictional discourse. When we say "In Conan Doyle's novels, Sherlock Holmes is a detective", we explicitly refer to the fiction, by means of the phrase "Conan Doyle's novels"; to that extent, we do take a "metafictional" perspective. This is different from being immersed in a fiction and not reflectively thinking about it. At the same time, parafictional discourse discloses the *content* of the fiction, what it represents; and this is does, in continuity with fictional discourse, by giving to imagine what the fiction prescribes its practitioners to imagine. That is what the audience of a parafictional utterance does: she imagines a fic-

tional state of affairs while simultaneously "tagging" the imagined state of affairs as one that is depicted in the fiction.[11]

New distinctions

Consequently, one is led to a more fine-grained view that the reduction given above. At this point, it is useful to introduce an informal notion of perspective to give a name to the three kinds of statements, going back to the initial intuition that metafictional statements are *about* the fiction whereas fictional statements are *within* the fiction.

At one end, there are internal perspective statements which involve pseudo-reference and pseudo-assertion. At the other, there are external perspective statements which involve serious reference and serious assertion. In the middle, there are mixed-perspective statements which involve pseudo-reference and serious assertion. Here is an update of the table with this data in mind:

	Serious assertion	**Pseudo-assertion**
Serious reference (to the individual of paper)	**Pseudo-reference** (to the flesh and blood individual)	
External perspective	*Mixed perspective*	*Internal perspective*
(17) Emma is a fictional character.	(30) Emma is my neighbour's favourite character. (29) Emma Woodhouse is happier than Emma Bovary	(16) Emma is handsome, clever and rich.

Table 17.2: Mixed perspectives and pseudo-reference

Intuitive notion of perspective

It seems that the correct distinction looming behind the fictional and the metafictional is that of shifting *perspectives* on the fictional events. First one can have an internal perspective in the sense that one is fully "immersed" in the pretence while producing of interpreting sentences involving a fictional name.[12] Second one can have an external perspective in the sense that one "steps outside" of the pretence

[11] [Récanati forthcoming], p. 20.

[12] Jean-Marie Schaeffer thus defines a notion of fictional "immersion" in [Schaeffer 1999] (§3.4), pp. 165-179.

and talk *about* it from the real world perspective. Some statements like (30), however, requires that one has one foot in the pretence and one foot outside it, so to speak.

Consequently, it is not so clear how this three-fold distinction using an intuitive notion of perspective maps on to the original distinction between the fictional and the metafictional. It is really a finer-grained description of the data. The problematic cases are the mixed perspective statements, for the internal and external ones clearly require fictionality-conditions and truth-conditions respectively. We hare thus left with the following table:

Truth-conditions	**???**	**Fictionality-conditions**
Serious reference	**Pseudo-reference**	
External perspective	*Mixed perspective*	*Internal perspective*
(17) Emma is a fictional character.	(30) Emma is my neighbour's favourite character. (29) Emma Woodhouse is happier than Emma Bovary	(16) Emma is handsome, clever and rich.

Table 17.3: Mixed perspectives: true or fictional?

In order to decide what semantic analysis one should give to mixed-perspective statements, I think we should get a better understanding of what a *perspective* is. Theorizing about perspective is a difficult task for it is bound to cover a huge variety of phenomena. I think what is needed here is a notion of *metaphysical* perspective, in the sense that one needs to talk about different referring environments, different "worlds".

17.2 Looking for a metaphysical perspective

> The exact definition of what is meant by a perspective is not quite easy. So long as we confine ourselves to visible objects or to objects of touch we might define the perspective of a given particular as "all particulars which have a simple (direct) spatial relation to the given particular."
>
> [Russell 1915], p. 413.

17.2.1 Narratological points of view

First and foremost, I need to dispel a possible confusion: the notion of metaphysical perspective I am looking for is not the same as the narratological notion of a *point of view*.

Narratologists often use the term "perspective" and "point of view" interchangeably to make subtle distinctions among possible narrators.[13] To use Genette's terminology, one should distinguish between a *homo-diegetic* narrator whose point of view is intrinsic to the fiction in the sense that it is an eye-witness of the fictional events and a *hetero-diegetic* narrator whose point of view is extrinsic to the fiction is the sense that it is not anchored in any specific character in the story.[14]

Classic examples of homo-diegetic narrators are Fitzgerald's *The Great Gatsby* or Nabokov's *Lolita*. These first-person narrators are characters in the story, eye-witness of the fictional events. Consequently, they are partial, biased and potentially unreliable. Classic examples of hetero-diegetic narrators are Austen's *Emma* or Balzac's *A Harlot High and Low*.[15] These third-person narrators are omniscient entities who should not be counted among the protagonists. They are typically thought of as unbiased and reliable.[16]

The main reason why narratologists introduced this distinction between different points of view is to talk about "objectivity" within the fiction. What "objectively" happened to Lolita? According to Humbert Humbert, she fell in love with him and they eloped. However, once the reader acknowledges the unreliability of Humbert as a narrator, one will say that, rather, she is a victim of paedophilia. This is what "objectively" happened. "Objectivity" in narratology is a subtle notion which is not grounded on facts, but on narratological cues.

However, both kinds of narrator have an internal perspective on the fictional events. In other words, both kinds of narrator are part and parcel of the double-structured pretence which underlies the fiction, as described in chapter 7. Indeed, they each correspond to one interpretation of the report principle which is a fundamental principle of fictional discourse. As a result, both kinds of narrators produce

[13] See for example [Booth 1983] or [Genette 1991].

[14] In order not to mix up terminology, I am systematically using "point of view" qualified by "extrinsic" and "intrinsic" when I talk about narratology. These terms comes from [Lamarque and Olsen 1994]. By contrast, I use the term "perspective" qualified by "internal" and "external" to denote the metaphysical perspective I am after.

[15] The French title is *Splendeurs et misères des courtisanes*.

[16] There are obviously all sorts of subtleties, complications and mixed points of view which are interesting for the narratologist to consider.

fictional statements. And neither produce metafictional statements.[17]

Here is how the contrast between a narratological point of view and a metaphysical perspective is put in [Lamarque and Olsen 1994]:

> What are these dual perspectives, the internal and the external point of view on fictional content? Although it is helpful heuristically to speak of what goes on inside and outside "worlds" the two perspectives that we seek to identify do not require the idiom of "worlds". They can be captured either as modes of speaking or as modes of responding.
>
> For example, the distinction shows itself systematically in the different kinds of answers we are inclined to give to questions about the content of works of fiction. Who created Frankenstein's monster? One answer, from the internal perspective, is of course: Frankenstein. Only from the external point of view must the reply be: Mary Shelley. The author is not part of what is imagined under the fictive stance.[18]

In other words, a metaphysical perspective consists in keeping track of where one stands with regard to the fictional events. It tells on which side of the border one stands: inside or outside the pretence, in fiction or in reality.[19]

17.2.2 Perspective in the visual arts

Metaphysical perspective is not the well-known linear perspective from drawing either. Although, as we will see, a precise characterisation of what linear perspective is will put us on the right track.

Projective geometry has a long and fascinating history. Many Renaissance architects and engineers pioneered in projective geometry, developing for the first time the mathematical technique of "linear perspective".[20] This technique enables one to draw a three dimensional space on a two-dimensional space with an illusion of reality. Filippo Brunelleschi invented linear perspective in the early fifteenth century. The method was systematised and popularised by the humanist Leon Battista Alberti.

[17] Even in metafiction, as seen in subsection 7.2.3, the narrator produces *fictional* statements.

[18] [Lamarque and Olsen 1994], p. 144.

[19] See [Lavocat 2016] for this terminology. She argues that fiction essentially comes with such a *metaphysical* border. Of course, the "border" at issue here is *not* physical, so the use of the term is metaphorical. Some theoreticians argue against her view by suggesting that she meant to take the metaphor at face value. See for instance Nadia Yala Kisukidi in Le Procès de la Fiction (starting from 2:28:00 onward).

[20] This technique was very probably already known to ancient Greek painters, but this know-how was lost in history.

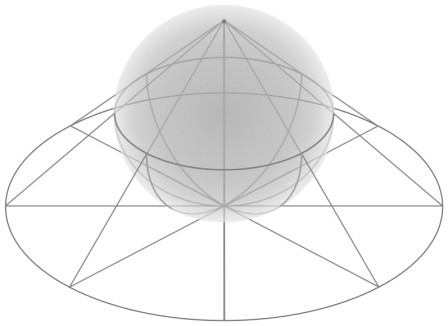


Figure 17.1: Stereographic projection of a sphere

Projective geometry

Projective geometry studies points, lines and plans and their projective properties, namely the properties preserved by projection. Projections are a set of transformations which can be rigorously defined. Intuitively, a projection transforms a geometrical figure into another figure, given a point which is called a perspective. The classic example is the rendering of parallel lines in the three dimensional space as two lines which intersect at the horizon in a two-dimensional space; the perspective being the point of view of the observer. In this context, a perspective is a point which, so to speak, organises the space one considers.

Here is a simple example of a projection.[21] Take a sphere. Take a plane tangent to the sphere. You can define the point which is both on the sphere and in the plane as the South pole. Now consider the North pole as a perspective. You can easily project every point on the sphere onto the plane relative to the North pole. The way of projecting each point on the sphere is very simple: draw a line passing through the North pole and the point to be projected. The projection is the intersection of this line with the tangent plane. In this way, every point on the sphere except for the north pole has a projection on the plane.

According to this projection, the equator of the sphere, for example, will be a big circle centered on the south pole on the plane. All the latitudinal lines will be projected onto concentric circles (bigger and bigger as the latitude gets closer to the north pole), all the longitudinal lines will be projected onto straight lines intersecting each other at the south pole.[22]

[21] This is called a "stereographic projection". See Figure 17.1 for an illustration.

[22] This projection can be used to draw a map of Earth although it breaks down the usual metric system used on a sphere. There are many different projections of sphere onto a plane which have all sorts of properties. For a long list of such projections, see https://en.wikipedia.org/wiki/

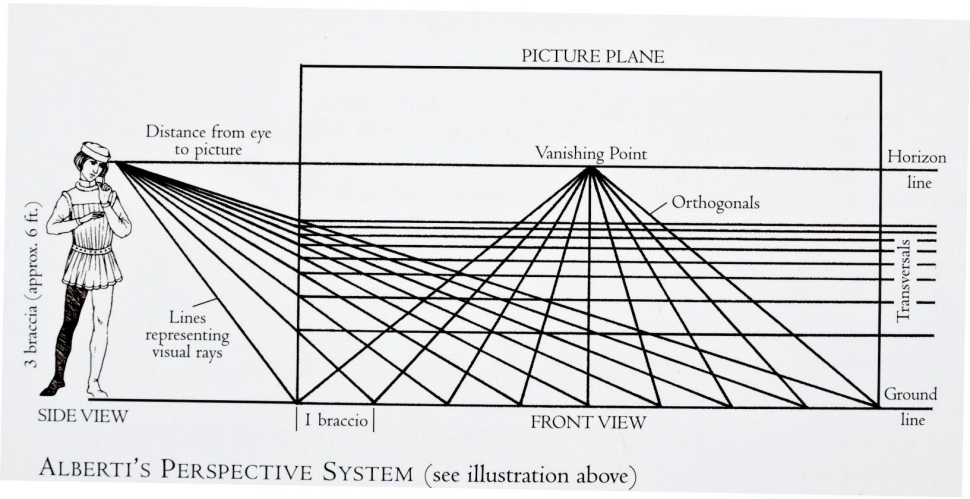


Figure 17.2: One-point linear perspective with horizon

Linear perspective

Linear perspective is a little more complicated. It consists in projecting parallel lines to lines converging into a so-called "vanishing point". Then one needs to have a way of retrieving orthogonal lines to the projected parallels. The usual method consists in choosing a line standing for the "horizon". (See for instance Figure 17.2)

The vanishing point and the horizon are usually chosen in such a way that the perspective is ideally that of the eye of the onlooker of the depicted three-dimensional space. This is how one can create the illusion of a three-dimensional space. But this need not be so, if one is not after illusion.

Indeed, in principle, one can define as many vanishing points as one wants. However, given that there are only six "natural" directions (the four cardinal directions, plus up and down) it is difficult to see what a seven- or eight-point perspective rendering of the three dimensional space would be. For this reason, there are well-known drawing techniques up to six-point perspective (with six vanishing point) but not more.[23]

List_of_map_projections.

[23]For a very short explanation and illustration of several-point perspective, see https://termespheres.com/6-point-perspective/.

Internal perspectives

At this point, one can see that the perspective point is not necessarily projected into the depicted space. The example with the sphere given above is a case in point: the perspective point, which is the north pole, does not get projected onto the tangent plan. The projection has created a so called "point at infinity", which is "outside" the depicted space. It is thus nowhere to be found in the depicted space. However, there is nothing mysterious about it. Indeed, one can easily infer, from the projection on the plane where this north pole is on the original sphere. It is also possible to define lines at infinity, using more complex projections.

Therefore, it is possible to mathematically distinguish between perspective points which inhabit the depicted space and those which do not. Typically, the case of linear projection creating an illusion of three-dimensional perception is a case where the perspective point inhabits the depicted space. The perspective point is usually that of the viewer. Hence, although it inhabits the depicted space, it is usually not represented in the painting. But it is possible to depict it. For example, Velásquez's *Las Meninas* used a very nice trick to depict the perspective point: he depicted a mirror which reflects the the royal couple that the painter is painting. (See Figure 17.3)

This distinction between perspective point at infinity and perspective point inhabiting the depicted space, however, does not correspond to the distinction between internal and external perspective. It rather corresponds to the narratological distinction between an intrinsic point of view and an extrinsic point of view discussed above. The analogy is robust.

17.2.3 Perspectives and conventions in the visual arts

Klee's drawing lesson

The distinction between the internal and external perspective is not given by mathematics. This observation is the core of Goodman's argument in [Goodman 1976] (§1.3 "Perspective", pp. 10-9).

Goodman argues against the idea that one-point linear perspective is the "natural" or "realistic" way of rendering the three-dimensional space into a two-dimensional medium. The proponent of this view (that Goodman considers) argues that one-point linear perspective is an application of the laws of optics and it thus represents what we actually see. But Goodman rightly points out that we have to use conventions in order to infer the perspective point in most cases, because projective geometry alone is indeterminate as to where the perspective point should be. As a



Figure 17.3: Velázquez 1656 *Las Meninas* (Museo del Prado, Madrid)

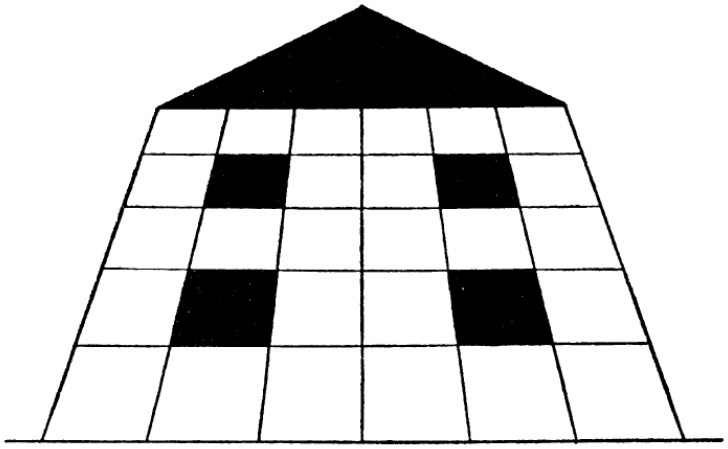


Figure 17.4: Drawing from Paul Klee's *Pädagogische Skizzenbuch* (1925)

result, Goodman shows that one point linear perspective really is an illusion applying the mathematical techniques of projective geometry *and* a great deal of conventions.

Paul Klee illustrates the point very forcefully (see Figure 17.4). Klee made a two dimensional drawing with one vanishing point and observes that "the drawing looks quite normal if taken as representing a floor but awry as representing a façade, even though in the two cases parallels in the object represented recede equally from the eye."[24] What Klee points out is that there is a convention according to which when there is only one vanishing point, it is a point at infinity *ahead*, and not *above* the perspective point. In other words, when we see a drawing with parallels meeting at infinity, we conventionally see the parallels as being horizontal and not vertical. But this is not given by mathematics, it is a convention.[25] Note that Paul Klee used this trick in several of his paintings to get a particular aesthetic effect. For instance in *Main Paths and Bypaths*, there is a sustained ambiguity about whether one should consider the vanishing point to be ahead or above (see Figure 17.5). So his drawing

[24]Cited in [Goodman 1976], p. 16.

[25]This is why we find algorithms which "correct" the image in numerical cameras in order to make sure that vertical lines will not converge. Such algorithms are presented, for instance, in https://www.college-de-france.fr/site/gerard-berry/course-2018-01-31-16h00.htm.



Figure 17.5: Paul Klee 1929 *Main Path and Bypaths* (Museum Ludwig, Köln) (fragment of the original)

lesson is not a mere thought experiment disconnected from his art. Goodman goes further and shows that these conventions have something to do with the way we usually look at pictures, i.e. the way we handle the *things* we call pictures. Indeed, in general, the picture of a huge facade will come with a perspective high and far enough so as to keep the vertical lines parallel. This perspective is not "realistic". The idea behind Klee's drawing is that Figure 17.4 could be the drawing a huge facade from our ordinary perspective, on the floor looking up. But it looks like a floor. Now, if we tilt the picture enough, Goodman remarks, we will suddenly accept the same image as the drawing of a huge facade.

Tilting the drawing requires an external perspective on the depicted space. Goodman's argument paves the way to the right metaphysical distinction I am after.

The external vs internal distinction in the visual arts

In visual arts, there are indeed external and internal perspectives on the depicted space. The external perspective is the visual perspective of the actual viewer. It depends on where they stand relative to the picture, if they are in front of the picture or looking at it from a funny angle. It also depends on whether there are people or opaque things in the way, as may happen in a crowded museum for instance, or in

an art studio where the pictures are not hanged high on a wall but put against the wall on the floor one covering another. By definition, from this external perspective, the viewer can see the frame of the picture. Also, by definition, if the picture is very large, or the viewer very close, they can always move back so they have "the big picture".[26]

The internal perspective, on the other hand, is the *inferred* one from the content of the picture. If the picture uses the linear perspective technique, then projective geometry can be used to make such an inference, other wise it is of no use. Most of the time, this inferring is not formalised and we do it intuitively. However, one can make explicit the inference using the methods of projective geometry to make a point. For instance, one can ask: "where is the point of perspective in this painting?"

If the linear perspective is done accurately, this can be inferred somewhat precisely, as in Leonardo's *The Last Supper* for instance where one can try to measure where the inner perspective is relative to Jesus, given plausible measures of the depicted scene.[27] If the linear perspective is inaccurate as for instance in Duccio's *The Annunciation*, such an inference of the internal perspective will not be possible. But we plausibly infer such an inner perspective either by correcting the mistakes or by postulating an inner perspective which does not inhabit the depicted space. (See the described contrast between the two paintings in Figure 17.6.) "Inaccurate", here, is not meant to be an aesthetic judgement.

From this internal perspective, by definition, the frame is not visible. Also, by definition, it is not possible to get closer to the depicted objects and individuals.

Clarification on the scope of the distinction

One might think that such a distinction between the external and the internal perspective is dependent on a painting representing a "space", in which, for instance, one can depict objects using linear perspective. What about "abstract" paintings? As, for instance, paintings from Jackson Pollock's "drip period". It does not seem that they project any space *in which* one can find an internal perspective.

I think the distinction holds in these cases, too. The internal perspective is trivially akin to a point at infinity, for there is no depicted space. The external perspective, by contrast, is that which makes it possible to see the painting. For

[26] Unfortunately, it is not always the case in museums like the Louvre, where some huge paintings are exposed... in a corridor.

[27] This has been done, using different methods in [Schwartz 1988] and [García-Salgado 2005], for instance.



(a) Duccio 1311 *The Annunciation* (The National Gallery, London)



(b) Leonardo Da Vinci 1490 *The Last Supper* (Convent of Santa Maria delle Grazie in Milan)

Figure 17.6: Duccio's inaccurate perspective compared to Leonoard's accurate one

instance, the lightning and the position of the painting in the room are features of the external perspective only.

One might want to argue that such a distinction of no (aesthetic) interest. Even

if this is true (which I very much doubt) this is besides the point for I have defined a notion of *metaphysical* perspective. So I claim that the distinction between internal and external perspective generalises to all kinds of visual representations.

17.3 Metaphysical perspective in linguistic fictions

The metaphysical notion of perspective here defined applies to linguistic fictions straightforwardly. To see this, one has to change the relation of "seeing" in the visual arts for the relation of "making reference" in linguistic fiction. This can be done without any problem, provided one endorses both a causal view of perception and reference. I have been assuming such views all along.

17.3.1 External and internal perspectives

In [Récanati 2012], the relation of reference is secured through what is called an "epistemically rewarding relation" (ER). The main ER relations are perception, memory and reliable testimony. ER relations correspond to what Lewis calls *relations of acquaintance* as defined in [Lewis 1999], p. 380:

> There are the relations that someone bears to me when I get a letter from him, or I watch the swerving of a car he is driving, or I read his biography, or I hear him mentioned by name, or I investigate the clues he has left at the scene of his crime. In each case, there are causal chains from him to me of a sort which would permit a flow of information. Perhaps I do get accurate information; perhaps I get misinformation, but still the channel is there. I shall call such relations as these *relations of acquaintance*.

In order to make singular reference to some individual, an ER relation (or a composition of ER relations) must hold between a cognitive subject and the object of reference. In such a framework, being able to make singular reference simply amounts to being ER-related. Conversely, if one cannot be ER-related to some individual, then it is impossible to make singular reference to it.

Fictional characters, by definition, are not inhabiting the same spatio-temporal framework as us. As such, it is impossible for a real cognitive agent to be ER-related to a fictional character. This is tantamount to saying that one cannot directly refer

to fictional characters, i.e. that fictional names are empty names. This corresponds to the external perspective on fiction.

Within a pretence, by contrast, one pretends to be ER-related to fictional characters. Consequently, one is able to directly refer using fictional names, within the pretence. This corresponds to the internal perspective on fiction.

The crucial difference between the two contexts is that one *pretends to be someone else*. Indeed, if one remained the same, then they would not be ER-related to fictional characters. So the cognitive subject involved in an external perspective is distinct from the cognitive involved in an internal perspective on the fictional content. In an external perspective, the cognitive subject involved must by the empirical self, with its history and abilities. In an internal perspective, let us call the cognitive subject involved a *pseudo-self*. In the case of a linguistic fiction, the pseudo-self thus generated is usually called the narratee.[28]

A pseudo-self, in general, need not have the same history or ability as the empirical self. However, it inherits some of the empirical self's abilities, as for example that of making singular reference. It seems also plausible to say that what the empirical self *cannot* do, the pseudo-self will not be able to do. The pseudo-self is a projection of the empirical self within the pretence, just like the internal perspective in the visual arts is a projection of the actual *viewer* (crucially endowed with a visual system) within the depicted space.[29]

So the distinction between reference and pseudo-reference boils down to a distinction between the kind of self involved in the utterance. We thus have a general definition of what a metaphysical perspective in a linguistic setting: a *metaphysical perspective* is a context in which the reference of "I" is fixed.

17.3.2 Mixed perspectives

The distinction between the empirical self and the pseudo-self, and what ER relations they hold, is the key to the mixed perspective. In mixed perspectives, given that one is making singular reference using a fictional name, it must be that such statements involve a pseudo-self. But it is also a given that the content of mixed-perspective statements is not determined by fictional statements, so the pseudo-self is not identical with the narratee. Consequently, mixed perspective statements come with a background pretence which generate its own pseudo-self. This is precisely

[28]See chapter 7 where I showed how the narratee is systematically generated in a double-structured pretence.

[29]See appendix C for a fruitful analogy between pseudo-selves and virtual machines in computer programming.

what the notion of extended perspective is modelling.

In particular, the underlying pretence puts constraints on the similarity that must hold between the empirical self and the generated pseudo-self. For instance, in the case of:

(30) My neighbour is in love with Emma Woodhouse.

the underlying pretence must be such that the pseudo-self is both ER-related to *Emma Woodhouse* and to the empirical self's neighbour. The easiest way of doing so is to devise an imaginative scenario in which the pseudo-self is essentially the empirical self, except that it entertains an ER relation with Emma Woodhouse on top of its already existing ER relations.[30] Within this pretence, the pseudo-self can make direct reference to both Emma and the neighbour.

By contrast, it seems that a sentence like:

(29) Emma Woodhouse is happier than Emma Bovary.

does not put any special constraint on the similarity that must hold between pseudo-self and the empirical self. Since the understanding of (29) consists in merging the two relevant fictions so as to have both Emmas "meet" in imagination, I suppose the easiest way to achieve this would be to generate the pseudo-self which entertains the union of all the ER relations associated with the two narratees of *Emma* and *Madame Bovary*. One would merge two internal perspectives, so to speak. But this need not involve a pseudo-self particularly similar to the empirical self: it may share none of the ER relations held by the empirical self.

This shows that the notion of pseudo-self allows for a continuum of cases within mixed perspectives. I think this fits well the linguistic data of mixed-perspective statements, for there seems to be many different cases. Consequently, one can easily account for our feeling that some mixed perspective statements are true or false (like (30)), whereas others feel more fictional (like (29)). This feeling is systematically correlated to the kind of constraint put on the similarity that must hold between the empirical self and the pseudo-self. The more similar they are in the underlying pretence, the more we are inclined to interpret the sentences as empirically true. (30) *feels* true, because it involves *me* (and my actual neighbour); (29) does not so. However, strictly speaking, both statement are fictional for they are "true" in an extended pretence. Putting constraints on the pseudo-self allows to *mimic* genuine

[30] Or maybe one should "merge" the ER relations of the empirical self and of the narratee generated by the reading of *Emma*.

truth-conditions.[31]

So we end up with the following table:

Truth-conditions	**Fictionality-conditions**			
External perspective	**Mixed perspective** (extended pretence)			**Internal perspective** (base pretence)
Empirical self	*Pseudo-self* (≈ empirical self)	...	*Pseudo-self* (no constraint)	*Pseudo-self* (= narratee)
(17) Emma is a fictional character.	(30) My neighbour is in love with Emma Woodhouse.	...	(29) Emma Woodhouse is happier than Emma Bovary.	(16) Emma is handsome, clever and rich.

Table 17.4: Final table of data: a continuum of mixed perspectives

The dots represent the continuity of cases. The further on the left of mixed perspective, the more empirically true the statement feels, the more the pseudo-self approximates the empirical self.

17.4 A proposal for anti-realism

17.4.1 Anti-realism and external perspective statements

Mixed perspective statements can thus be accounted in pretence semantics for given two extra notions. One is that of extended pretence, the other is that of a pseudo-self. It nicely explains how pseudo-reference to fictional characters can yield a continuum of cases where fictionality-conditions are used to mimic truth-conditions.

Here is how Everett discusses this simple extension of his framework:

> So my account can treat the sort of "mixed perspective" discourse that we have been considering in a very straightforward way. Such discourse will all take place from within an internal perspective, albeit one

[31] At the far end of this spectrum, one finds parafictional statements like: "In Jane Austen's *Emma*, Emma Woodhouse is handsome, clever and rich." These are bound to strongly feel empirically true, since the pseudo-self involved must be the same as the empirical self which is the actual reader, acquainted with Jane Austen's book. See appendix B for an account of parafictional statements along these lines, using a fruitful analogy with the notion of metadata as used in computer programming.

> pertaining to an extended pretence rather than the original base pretence. Where traditionally people have distinguished between fictional character discourse made from an internal fictional perspective and fictional character discourse made from an external real-world perspective, we should rather distinguish between discourse which takes place within the original base pretence and discourse which takes place within an extension of that base pretence, and which is used to convey information about it. Granted this view, it is hardly surprising that, when we engage in an extended pretence, we frequently ascribe both sorts of property to the things we talk about, that we "mix" our perspectives, as it were.[32]

However, pretence semantics cannot handle external perspective statements. External perspective statements have truth-conditions. Pretence semantics cannot deliver truth-conditions. The best it can do is to deliver fictionality-conditions which *feel* like truth-conditions. As shown in subsection 15.2.2, at this point, Everett bites the bullet and say that external perspective statements *feel* true although they are fictional in the relevant extended pretence.

In order to avoid this bothersome consequence, I propose that we use a version of positive free logic to provide truth-conditions for the external perspective statements. Free logic is a formal framework designed to handle both designating and non-designating terms extensionally.[33] It is thus in keeping with anti-realism for it takes for granted that, by definition, fictional names are empty names. Free logic has already been used in [Evans 1982] and [Sainsbury 2009], but in both works it was a version of *negative* free logic. I develop this proposal in the following chapter by presenting a positive free semantics and then combining it with pretence semantics.

17.4.2 Focus on negative existentials

Broadening the picture

True negative existentials are not restricted to those involving fictional names. Hence a semantic analysis of negative existentials is somewhat a side step in theorizing about fictional reference. I will thus rehearse the problem of negative existentials from scratch in the following chapter. It will be a good introduction to free logic and it will hopefully broaden the picture so as to make conceptual connections between reference and existence in general, not through the sole mediation of fiction. This broadening the picture will especially be welcome as an ending of this dissertation.

[32] [Everett 2013], p. 165-6.
[33] So the principle of compositionality is untouched.

However, I think there are also good reasons to focus on negative existentials given the aim of this part. In defending anti-realism, one is naturally led to negative existentials for two reasons.

A strange twist of fate

As already hinted at, I can now explain why negative existentials are like a thorn in the anti-realist's side.

Negative existentials like (20) are clearly external perspective statements. Indeed, they are not fictional. They are true precisely because one looks at the fiction from the outside.

Some may want to doubt this, on the ground that negative existential statements have a controversial logical form anyway. So they would simply dismiss negative existentials from external perspective statements until their logical structure is agreed upon. Indeed, if it turns out that negative existentials are existentially quantified statements involving no name at all, then they fall outside the data about the different uses of fictional names.[34]

But this dismissal is, I think, not available for the anti-realist. Indeed, the basic tenet of anti-realism is that *fictional characters do not exist* (equivalently that fictional names are empty names). So they are, to the bone, committed to the *truth* of this claim. And this claim entails singular negative existentials like:

(20) Emma Woodhouse does not exist.

Therefore, the anti-realist should say that (20) is true. As a result, (20) cannot an internal perspective statement (besides, the denial of (20) is fictional in *Emma*!) So the anti-realist has compelling reasons to hold that (20) is a good candidate for the external perspective.

On the other hand, the anti-realist cannot deliver truth-conditions for *any* external perspective statements. This what has been shown in the previous paragraph. The best they can do is to deliver mock-truth-conditions. Consequently, they are committed to the view that (20) is not, strictly speaking, truth-evaluable.

I think this is a strong internal tension on the anti-realist's side. Consequently, I take it that the priority is to show how one can give truth-conditions for negative existentials, before any other external perspective statements.

[34] See the beginning of chapter 18 for a critical review of this position and similar ones.

The logical status of negative existentials

A second reason to focus on negative existentials comes from the fact that they are entailed by all other external perspective statements. For instance, the truth of:

(17) Emma Woodhouse is a fictional character and main protagonist of Jane Austen's *Emma*.

entails the truth of (20). This was already hinted at in subsection 14.3.1. We can now prove it.

Given the meaning of "I", it must be that the following is true within each perspective:

(41) I exist.

Furthermore, ER relations are such that they preserve existence. Indeed:

- If a and b are ER-related and a exists, then b exists.
- If it is impossible for a and b to be ER-related, and a exists, the b does not exist.

External statements, I showed, force the reference of "I" to be the empirical self. By definition, the empirical self cannot be ER-related to any fictional character. Hence, given the truth of an external perspective statement, the corresponding negative existential must be true.

17.5 Conclusions

In this chapter, I have given a detailed analysis of distinction between the fictional and the metafictional .

I started with the naïve idea that the distinction could be reduced to the well-known distinction between serious assertion and pseudo-assertion. This latter distinction originates in the work of Frege and was later developed by Searle. It can be defined rigorously without presupposing a specific theory of assertion.

I showed that this reduction is inadequate, for it yields a too coarse-grained description of the data. In particular, what is interesting about the distinction between the fictional and the metafictional is that it points to an interesting, systematic taking apart of reference and assertion. We thus had to make room for so-called "mixed perspectives". These mixed perspectives do not square well with the distinction between serious and pseudo-assertion.

This more complex picture put me on the move for a general theory of perspective which can divide into three types: the external, the internal and the mixed perspectives. I termed this notion of perspective "metaphysical" for it consists in keeping track of where one stands, relative to the fiction. I showed that the notion of metaphysical perspective should not be conflated with that of a narratological point of view. I then showed that a proper analysis of linear perspective puts one on the right track. Thanks to an argument one can find in [Goodman 1976], I could define a distinction between the internal and the external perspective in the visual arts. I illustrated this section with paintings and drawings in order to give steady intuitions before turning to linguistic fictions again.

I explained why this notion of metaphysical perspective can easily be extended to linguistic fictions. Indeed, in [Récanati 2012]'s framework, the analogy between perception and direct reference is complete. In this framework, the shift in metaphysical perspective consists in pretending to be someone else. This grounds a distinction between the empirical self and what I termed a "pseudo-self". The internal perspective is thus when the pseudo-self is identical of the narratee (as defined by the principles of generation underlying each fiction). The external perspective, by contrast, involves the empirical self. As for mixed perspective, the making reference to a fictional character by name requires to have a pseudo-self. This gave a good argument for treating mixed perspective statements as involving an extended pretence, in keeping with the anti-realist's strategy.

The pseudo-self involved in mixed perspectives is generally not identical to the narratee. Interestingly, one can put constraints on the similarity that must hold between the pseudo-self generated by the underlying pretence and the empirical self. I showed how cross-over statements and intensional statement put different constraints on the pseudo-self they generate. This allows for a continuum of mixed perspectives. This, in turn, explains one's phenomenology about these statements: some *feel* empirically true whereas others do not. The correlation is straightforward: the more the constraint is that the pseudo-self should approximate the empirical self, the more empirically true (or false) the resulting mixed perspective statement feels.

Anti-realism can thus straightforwardly account for internal and mixed perspective statements, using pretence semantics. However, I showed that pretence semantics is of no use for external perspective statements thus defined. This puts the anti-realist in a delicate situation. I proposed that, now we have clearly defined the problematic cases for anti-realism, we use free logic to handle external perspective statements. I gave good reasons to focus on negative existentials, which is the topic of the next chapter.

Chapter 18

Freedom for Emma!

18.1 The problem with negative existentials

Negative existentials are sentences of the form: "So-and-so does not exist", "There is no such thing as so-and-so", etc. Intuitively, a negative existential is used to say that the referring expression "So-and-so" is empty, i.e. does not refer.

"So-and-so" here is supposed to be a singular term, i.e. a term purporting to make singular reference to an individual. Consequently, general terms like "horses" or "unicorns" are not here available substitutions for "so-and-so" and thus do not yield negative existentials in the sense discussed here. As indefinite descriptions of natural languages are usually identified with general terms, "a horse" or "a unicorn" are not available substitutions either. Since the sentence "Unicorns does not exist" is ungrammatical and the sentence "A unicorn does not exist" is borderline grammatical, it is useful to appeal to our syntactic intuitions of English here.[1]

Natural language candidates for substitution of "so-and-so" are thus names, indexicals, pronouns and definite descriptions. I will mainly focus on negative existentials involving names in this chapter, although the theories I discuss apply to all well-formed negative existentials thus defined.

The following is, according to anti-realism, a *true* negative existentials in this sense:

(20) Emma Woodhouse does not exist.

Indeed, "Emma Woodhouse" is a fictional name. According to anti-realism, by definition, a fictional name does not name any real individual. Hence, (20) is true by

[1]Negative existentials here defined correspond to what is called "singular existence statements" in [Quine 1939] to contrast them with "general existence statements" like: "Unicorns do not exist".

definition.

(20) being true by definition has been found problematic for different reasons: either because it conflicts with one's favorite theory of proper names (especially direct reference), or because one finds it difficult to understand how (20) can express a (true) proposition if the name "Emma Woodhouse" be empty. It is thus plausible that one wants to deny, after analysis, that (20) is true *by definition*. *Prima facie*, though, the anti-realist can give a good justification for the intuitive truth of (20).

This is part of the *problem* with negative existentials, which I am about to set up.

18.1.1 The Naïve Analysis

Here is a tentative observation about (20) that I will call the *naïve analysis* (NA). According to NA, the logical form of (20) is that of a negated predicative sentence, i.e. "a is not ϕ".

First, NA says that (20) is the negation of:

(42) Emma Woodhouse exists.

Given classical truth-conditions for negated sentences, (20) is true iff (42) is false. This part of NA is not controversial.

Second, NA claims that (42) is a predicative sentence of the form "a is ϕ". This is justified by the surface structure of (20). This second claim, by contrast, is very controversial in the philosophical literature as we will shortly see, although I take it that NA is *prima facie* very intuitive.

18.1.2 Classical Quantification Theory

Central tenet of CQT

The following is the central tenet of classical quantification theory (CQT):

(43) Everything exists.

There are good reasons to think that (43) is a tautology (or boils down to one). Indeed, given the definition of "everything" it seems that (43) must be true in all circumstances. Quine puts this idea in the opening of "On what there is":

> A curious thing about the ontological problem is its simplicity. It can be put in three Anglo-Saxon monosyllables: "What is there?" It can be answered, moreover, in a word – "Everything" – and everyone will accept

> this answer as true. However, this is merely to say that there is what there is.[2]

According to Quine, (43) boils down to saying that "there is what there is", which is a tautology. He offers an argument in favor of his view. The argument crucially relies on CQT.

Quine's argument

In first order logic, "everything" is part of the logical vocabulary. Consequently, its meaning is made explicit and is the same in all models. Recall:[3]

- $\forall x\phi$ is true in a model $M =< D, v >$ iff for all $d \in D$, $v(\phi[d/x]) = 1$.

In words: a predicate ϕ is true of "everything" just in case the sentence "a is ϕ" is true for all the possible "things" a might possibly denote, i.e. all the individuals in the domain. That is to say: each possible instantiations of the sentence yields a true sentence.

For Quine "to exist" famously means "to be the value of a variable". In other words, there exists all and only the individuals which are in the domain of the chosen model. This is merely a logical way of saying that all the existing individuals are the "things" the model recognises as individuals.

To sum up, "everything" (as defined in CQT) targets all the individuals of a given domain of individuals. Saying of these that they "exist" ensures that all these selected individuals are part of the domain. But this is true regardless of which domain one chooses to consider. Hence, (43) is true in all models by definition: it is a tautology.

This does not mean that everyone agrees on *what there is*. As Quine remarks: "There remains room for disagreements over cases". Indeed, philosophers can disagree about which model is the adequate model of reality. They thus disagree about which individuals actually or really exist. In particular, they disagree about what is the domain of the actual or real world model. What they agree about is that all the things in the real world are the existing things of the real world.

Despite Quine's forceful statement and his logically motivated argument, it is not the case that everyone accepts (43) as being true by definition. Some philosophers reject CQT and adopt instead a non-classical quantification theory which blocks the above reasoning.

---

[2] [Quine 1948], p. 21.

[3] v is an *interpretation* function of the language of quantification theory. I follow Priest's notation in [Priest 2008a].

18.1.3 The problem

These preliminary considerations about the meaning of "existence" showed that both NA and CQT seem to be good starting points. It is only natural to want them both in order to give the correct semantic analysis of negative existentials. But is it possible to hold both NA and CQT?

Suppose NA is correct. Then one can infer from the truth of (20) that there is something (namely Emma Woodhouse) which does not exist. This contradicts (43). Hence CQT cannot be true.

This argument is usually taken to force philosophers to take sides. The "logical orthodoxy" has denied NA to keep CQT; whereas the "logical heterodoxy" decided to drop CQT in favor of NA.[4] Unfortunately, I will show in the following that both doxies have defective theories for negative existentials due to their philosophical commitments. Fortunately, it is possible to hold both NA and CQT together. It suffices to block the above inference. This will open the way for free logic.

18.2 Orthodoxy and heterodoxy on negative existentials

18.2.1 Dropping NA: the logical orthodoxy

Logical orthodoxy is a venerable tradition in philosophy denying that existence is a genuine predicate. Kant is the most famous of them; Abelard can be seen as Kant's medieval forerunner[5], Frege, Russell and Quine as Kant's followers on this.[6]

If existence is not a genuine predicate as the logical orthodoxy argues, NA cannot be correct, despite its *prima facie* appeal. To the extend that NA is intuitive, rejecting NA can be seen as a problem. The logical orthodoxy's task is thus to give the logical form of negative existentials like (20) as a plausible alternative to NA.

Different kinds of negative existentials

The orthodox distinguish between different kinds of negative existentials, depending on the referring expression involved in the statement. Here are the different well-

[4]The terminology comes from [Lewis 1990]. Lewis considers himself a spokesman for the orthodoxy in this paper.

[5]Alain de Libera gives very interesting textual evidence for this claim in one of his lessons at Collège de France: https://www.college-de-france.fr/site/alain-de-libera/seminar-2014-06-26-11h30.htm (between 33'00 and 45'00).

[6]See [Salmon 1987] (pp. 60-1) for very relevant quotes from Frege and Russell.

formed negative existentials:

(44) This (pointing to something) does not exist.

(45) I do not exist.

(46) She does not exist.

(47) The present King of France does not exist.

One can see at least two sharply distinguishable classes of negative existentials here. There is a clear intuition that sentences like (44) and (45) are always false; while it is easy to imagine situations in which sentences like (46) and (47) are true.[7] I will put sentences like (20) aside for the moment.

An explanation for this fact is that (44)-(45) are *self-defeating statement*, to use Hintikka's terminology in [Hintikka 1962a], for each time someone (successfully) utters (44)-(45), one contradicts oneself. Indeed, in using "this" (pointing to something) or "I", one contextually refers to something in the vicinity or to oneself. On the other hand claiming that these do not exist amounts to saying that one cannot successfully refer to them, which contradicts what one just did. Hence, there is no context in which one can succeed in saying (44)-(45) truly.

There is a hidden premise in this line of argument which should be made explicit. It is generally acknowledged that the following is the principle linking reference and existence:[8]

(REF) Successful reference presupposes existence.

By contrast, it is very clear that (46)-(47) are not self-defeating statements at all. If you want to explain to your daughter that Santa Claus is not going to come on December 25, you are likely to say that "He does not exist" truthfully. As for (47), it is empirically true.

Following (REF), the natural move for the logical orthodoxy is to deny that successful reference is taking place in these cases. This is why (44)-(45) are crucially different from (46)-(47). The task of the logical orthodoxy is to paraphrase (46)-(47) so as to show that, contrary to appearances, pronouns and definite descriptions are not genuine referring expressions.

The paraphrasing of definite descriptions has famously been done by Bertrand Russell in [Russell 1905] and in [Russell 1919] (ch. 16: "Descriptions"). (47) is paraphrased into:

[7] You might worry about (44) and (46). I merely need a contrast here. Please focus on the stark contrast between (45) and (47) and disregard (44) and (46) if they confuse you.

[8] I think this is the principle driving the logical orthodoxy into thinking that one does not need to have an explicit predicate for singular existence: all we need is a good theory of reference.

(48) There is no unique individual such that it is the present King of France.[9]

The logical form of such a paraphrase is that of an existentially quantified formula in first order logic: there is no singular term involved. Pronouns can be seen as contextual definite descriptions and be paraphrased away after the same method.

Therefore, the logical orthodoxy has two methods for dealing with negative existentials: either the negative existential involves a genuine referring term and thus the negative existential is always false, or it involves a pseudo-referring term which should be paraphrased away. It is thus committed to the view that (20) and (44)-(47) do not have the same logical form. However, there is no argument giving good reasons to think that the above distinctions should have an effect on the logical form of negative existentials. The orthodox view on negative existentials in general is thus bound to be very inelegant: it postulates, without argument, that there is no unified logical form for all negative existentials. This inelegance is a consequence of rejecting NA.

Now, what does the logical orthodoxy tell us about negative existentials like (20), involving names? There are two main proposals in the literature, depending on one's preferred theory of proper names. Descriptivism about proper names puts (20) on a par with (46)-(47); direct reference about proper names puts it with (44)-(45). I will discuss them in turn.

Descriptivism and negative existentials

We started with the idea that (20) is true. Hence, it is not a self-defeating statement. So, there is a strong inclination to think that the name "Emma Woodhouse" has the same logical form as a definite description. By Russell's standard, then, it is not a device of genuine reference. This claim is the tenet of descriptivism about proper names.

One can think of many different descriptions which could fit the name "Emma Woodhouse". For instance, the name "Emma Woodhouse" could be an abbreviation for the long definite description listing all the predicates which are fictionally "true" of Emma Woodhouse in *Emma*: "The handsome, clever, rich woman ..." By contrast, "Emma Woodhouse" could be turned into "the woman named 'Emma Woodhouse' in Jane Austen's *Emma*". So, there is room for debate over which kind of definite description hides under the name "Emma Woodhouse".

[9]This is not Russell's paraphrase but a simpler equivalent one. For the record, Russell's paraphrase in [Russell 1919] is the following twofold statement: "(1) 'x is the present king of France' is not always false and (2) 'if x and y are present kings of France, then x and y are identical' is always true."

These debates, however, need not concern us here. Indeed, Quine has devised a general method for turning a proper name into a definite description.[10] One can introduce an *ad hoc* predicate which stands for one's preferred definite description. "Emma Woodhouse" thus generically becomes: "The emma-woodhousiser". (20) can then be paraphrased into:

(49) There is no unique individual such that it is the emma-woodhousiser.

But such paraphrase does not seem to have the same truth-conditions as (20). (49) is predicted to be false when the extension of the predicate "emma-woodhousiser" is exactly one. By contrast, it does not seem that (20) should be false in this case. Suppose for instance, that there was a woman in the real world which matches the description hidden under the name "Emma Woodhouse" perfectly. One could argue that (20) would still be true, for the fictional name is *de jure* empty.[11]

Furthermore, (49) is predicted to be true iff either the extension of the predicate is null or more than one. But this second condition sounds really odd, when it comes to (20). One could try and fix this oddity by taking the uniqueness condition in the description to be a presupposition, rather than a plain assertion, following the strategy given in [Strawson 1950]. In which case (20) would be predicted to be truth-valueless if the definite description targets more than one individual. I am not sure how this fits our intuitions about (20). I think these case-base line of reasoning is an artifice of the theory designed to yield predictions fitting our intuitions.

Descriptivism thus have problems with negative existentials. Maybe these can be solved but at the cost of inelegant sophistications. Moreover, there are some independent reasons to reject descriptivism about proper names and to adopt direct reference instead.

Direct reference and negative existentials

There are very compelling arguments in favor of direct reference for names which can be found in [Kripke 1972]. These arguments are independent from the semantic analysis of negative existentials. Direct reference is the view that proper names *are* devices of genuine reference. Hence, when it comes to negative existentials like (20), the descriptive strategy cannot work.

Consequently, negative existentials are a problem for direct reference as Kripke acknowledges in [Kripke 1973/2013]. Indeed, given the above general principle (REF)

[10] In [Quine 1948], p. 27.
[11] There is such an argument in [Kripke 1972].

and direct reference, it follows that negative existentials like (20) are always false. [12] This is a bad prediction, for (20) is clearly true.

One easy way out of the problem would be to deny that "Emma Woodhouse" is a genuine name, because it is a fictional name. Instead, it could be thought of as a descriptive name for which direct reference does not apply and the descriptivist paraphrase should work fine. This is arguably a possible line of response for the direct reference theorist.[13]

I think this response begs the question. What is the distinctive feature of a fake name like "Emma Woodhouse" if not the fact that a negative existential like (20) is true? Indeed, there is no syntactical difference between "Emma Woodhouse" and, say, "Saul Kripke". So I don't see why there should be a difference in the logical form between (20) and:

(50) Saul Kripke does not exist.

Of course, (20) and (50) differ with respect to their truth-value, for (50) is clearly false. But explaining this difference in truth-value by postulating a difference in logical form between (20) and (50) is very dubious. If one goes on to say that the right paraphrase one wants to give to these sentences depends on their truth-value, then the argument is circular. One needs good reasons for distinguishing between genuine and fake proper names which are independent of one's preferred semantic analysis of negative existentials. These are lacking.

A more plausible solution to the problem of negative existentials for direct reference has been given in [Donnellan 1975]. It consists in strengthening the general principle (REF) into:

(REF+) Successful reference is equivalent to existence.

According (REF+), denying existence amounts to denying that one can *successfully* refer using the referential expression at issue. That is to say: negative existentials affirm the failure of reference of the involved referring expression.

Donnellan then gives the detail of a theory of successful reference, in keeping with direct reference. By definition, successful reference depends on the existence of a historico-causal chain linking the utterance of the referring expression and the object of reference. In the case of a proper name, the historico-causal chain starts with a baptism of the object of reference; all the subsequent successful uses of the name should either directly point to this initial baptism or to another successful

[12] The reasoning is the same as before with negative existentials involving demonstratives and indexicals (44)-(45).

[13] See in particular [Currie 1990] (pp. 128-135) for such a line of reasoning.

use of the name.[14] For instance, "Saul Kripke" was so introduced sometime after November 13 1940 somewhere in New York. My present successful reference to Saul Kripke by using the name "Saul Kripke" rests on my learning his name from my philosophy professors; my philosophy professors successfully used the name either because they met Kripke or because they learnt it from other philosophers; Kripke's successful use of his name very plausibly comes from his parents or any other relevant people who were present at his baptism. All the successful uses of "Saul Kripke" thus stem from the initial baptism of Saul Kripke about 80 years ago. This network of successful reference shall continue in the same manner long after Krikpe's death.

Consequently, a proper name does not successfully refer iff there is no such network originating in a baptism. To use Donnellan's terminology: when the historico-causal chain of reference "ends in a block". A block is a non-conventional or unsuccessful dubbing event. Fictional names are paradigmatic examples of non-conventional dubbing events. Indeed, there is clearly a network of successful utterance of the name "Emma Woodhouse" which traces back to 1815, in England. But the first time the name "Emma Woodhouse" was introduced was not in the context of a baptism of a real person or object. Rather, it was in Jane Austen's novel *Emma*. So the real world object which is at the root of the use of "Emma Woodhouse" is a book, not a woman. Such a category mistake is the paradigmatic example of a "block" in Donnellan's sense.

Donnellan's paraphrase of (20) is thus the following:

(51) "Emma Woodhouse" does not successfully refer, i.e. the history of uses of the name "Emma Woodhouse" ends in a block.

Interestingly, (51) is a negated predicative sentence. So this view seems in line with NA. So, in principle, it seems that Donellan avoids the structural problem mentioned earlier: all kinds of negative existentials are paraphrased in the same manner. Of course, the networks involved for, say, indexicals do not obey the same constraints as those for names.[15]

However, it is at the cost of substantively altering NA, for the predication involved in (51) is metalinguistic. It is predicated *of the name* "Emma Woodhouse", that it does not successfully refer. NA claims on the other hand that the predication is a standard objectual predication. In other words, NA says that (20) says something *of Emma Woodhouse* (namely that she does not exist) and not of "Emma Woodhouse".

[14]By "baptism" I mean an overt dubbing event whereby the name is introduced. It is not necessarily a religious event.

[15]The structural features of networks associated with demonstratives and indexicals should explain the fact that (44)-(45) are self-defeating.

So I doubt this metalinguistic feature of negative existentials predicted by Donellan is as intuitive as NA.[16] In principle, one might argue that negative existentials are metalinguistic statements. But one would need good, independent reasons to think so.

Moreover, one would also need an independent argument for the strengthening of (REF) into (REF+), which is lacking. These are not knock-down arguments of the theory, but, I suggest, a real drawback one might want to avoid.

18.2.2 Dropping CQT: the logical heterodoxy

The logical heterodoxy is also a venerable tradition philosophy taking singular existence to be a genuine predicate so that NA is correct. Alexius Meinong is a famous philosopher of this tradition, when he defended the idea that one can truly assert:

(52) There are things which do not exist.

Emma Woodhouse, for instance, is one of the inhabitants of Meinong's famous "jungle", as the truth of (20) shows.[17] (52) contradicts the central tenet of classical quantification theory according to which (43) is a tautology (or boils down to one). Meinong and his followers should thus criticise the "classical" theory of quantification, providing instead a non-classical, or heterodox, or Meinongian theory of quantification.[18] Non-classical quantification theory thus interprets "everything" (and its dual "there is") so that (43) is false.

A Meinongian model is triplet $< D, E, v >$ with D a non-empty set and E a subset of D called the "inner domain".[19] The inner domain E contains all the existing

[16] I follow Evans and Kripke in this doubt. As Evans put it in [Evans 1982], p. 349:

> Kripke emphasises the requirement that (in the cases we are interested in) singular terms must be seen as *used*, in both negative and affirmative existential statements. As I have said, I think this is absolutely correct; indeed, it is partly constitutive of the problem.

[17] The expression "Meinong's jungle" comes from [Routley 1980].

[18] I do not claim to make justice to Meinong's rich and complex philosophical system here. My calling "Meinongian" the logical heterodoxy should be interpreted as a tribute to Meinong rather than an adequate interpretation of his view on existence.

[19] Another possible way of defining a Meinongian model is to add to a classical domain an "outer domain" O aside the classical domain D such that $O \cap D = \emptyset$. The two strategies are equivalent. To my knowledge, the "outer domain" strategy was first put forward in [MacColl 1905]. The "inner domain" strategy, I think, is closer to Meinong's spirit in [Meinong 1904].

Historically, the first formalisation of these ideas come from free logicians. See [Cocchiarella 1966] for the inner-domain strategy and [Leblanc and Thomason 1968] for the outer-domain strategy, both

individuals. The nonexistent ones are in $D \setminus E$.[20] The definition of "everything" is as usual ranging over all the individuals in D:

- $\forall x\phi$ is true in $< D, E, v >$ iff for all $d \in D$, $v(\phi[d/x]) = 1$.[21]

According to this definition, it is easy to construct a Meinongian model in which (43) is false and (52) true.

Meinongians argue that the real world model is a Meinongian model. One key moment in their argumentation is the fact that negative existentials like (20) *are* true, and they accept NA.

There are well known metaphysical objections to this Meinongian strategy: the dual domain idea was charged of "ontological extravagance" by the logical orthodoxy.[22] These objections can arguably be met by the so-called "neo-meinongians", as in [Parsons 1980] and his followers. I do not have any interesting, new things to say about this debate here. My criticism of the logical heterodoxy will rather focus on two little-known technical problems which I take to be serious problems. I will discuss them in turn.

On the multiplication of quantifiers

Given a Meinongian model, one can define several interconnected pairs of quantifiers. Indeed, one "universal" quantifier can be defined to range over D as seen above, but one can decide to define another ranging over E and another ranging over $D \setminus E$. Here are the formal definitions (with useful names and symbols for discussion):

1. (Unrestricted universal) $v(\Pi x\phi) = 1$ iff for all $d \in D$, $v(\phi[d/x]) = 1$.
    - (Unrestricted existential) $\Sigma x\phi \equiv \neg\Pi x\neg\phi$
2. (Restricted universal) $v(\forall x\phi) = 1$ iff for all $d \in E$, $v(\phi[d/x]) = 1$.

stem from [Church 1965]. See [Bencivenga 2002] §8 for historical detail. For this reason, there is a debate over whether "Meinongian" quantification theory should count free quantification theory or not, see [Paśniczek 2001]. In my view, it does not. Indeed, I define free quantification theory as validating the central tenet of CQT (43), as made explicit below.

[20] Another way of achieving the same result is to distinguish between different kinds of variables and develop a so-called "many-sorted" logic as in [Wang 1952]. One might think this is more interesting in that this strategy consists in applying a general theory to the theory of quantification. The two approaches are equivalent though: see [Rescher 1960] for a proof.

[21] The same definition can be found in [Jacquette 1991] and [Priest 2008a].

[22] The expression comes from [Nolt 2007]. Classic objections are in [Russell 1905] and [Quine 1948].

- (Restricted existential) $\exists x\phi \equiv \neg\forall x\neg\phi$

3. (Dual restricted universal) $v(\pi x\phi) = 1$ iff for all $d \in D \setminus E$, $v(\phi[d/x]) = 1$.

   - (Dual restricted existential) $\sigma x\phi \equiv \neg\pi x\neg\phi$

Given these definitions, if one introduces a predicate of existence $\mathfrak{E}$ whose extension is E, we have the following validities:

$$\models \forall x\mathfrak{E}x$$

$$\models \pi x\neg\mathfrak{E}x$$

$$\models \neg\Pi x\mathfrak{E}x$$

In words, these validities say that the restricted universal $\forall$ is a quantifier with existential import. The unrestricted universal Π is the Meinongian quantifier without existential import. The dual unrestricted universal π has what could be called "non-existential" import.[23]

Furthermore, the three pairs of quantifiers are inter-definable in the following sense:

- $\forall x\phi \equiv \Pi x(\mathfrak{E}x \to \phi)$
- $\pi x\phi \equiv \Pi x(\neg\mathfrak{E}x \to \phi)$
- $\Pi x\phi \equiv \forall x\phi \wedge \pi x\phi$

The first problem for Meinongian theory of quantification thus consists in interpreting these different quantifiers one can formally define. In particular, when it comes to applying the theory to negative existentials, one needs to argue that the natural language "everything" and "there are" which appears in (43) and in (52) is best construed as the *unrestricted* quantifiers Π and Σ.[24]

Let us grant, for the sake of argument, that the distinction between $\forall$ and Π corresponds to the distinction between the natural language "everything" and "all";

[23]This dual unrestricted universal is, to my knowledge, nowhere to be found in the literature although perfectly well defined here.

[24]Interestingly, Priest reviews the history of logic to show that the restricted universal $\forall$ somehow became the "classical" definition of "everything" with the work of Frege and Russell whose views on negative existentials are everything but neutral. See [Priest 2008b] for textual evidence in favour of this claim. So maybe the burden of proof is on the logical orthodoxy rather than on the logical heterodoxy here.

and that $\exists$ and Σ correspond to "there exists" and "there are".[25] What about π and σ, then? A natural language correlate would be a pair of quantifiers which range only over the set of nonexistent things. As noted above, it should be true by definition that "$\pi x \neg \mathfrak{E} x$". In natural language, this would mean that "every-non-thing does not exist" is something like a tautology, just like (43) is. This "every-non-thing" really sounds like a logical artifice.

A Meinongian theory of quantification can thus be seen as a logical apparatus completely disconnected from natural language. As such, one ban be sceptical about this move away from CQT. Indeed, quantification theory aims at theorizing about the natural language quantifiers. In order to mollify the sceptic, one needs to have an account of how the non-classical theory of quantification squares with natural language and why this weird multiplication of quantifiers should not bother us. This account should be motivated independently of the analysis of negative existentials.

On the multiplication of domains

Another worry stems from the arbitrariness of the double domain of Meinongian models. The main reason why Meinong himself wanted to distinguish between two domains of individuals is because of true negative existentials (assuming that NA is correct). But given a double domain model, one can construct problematic negative-existential-like sentences at the second level. For instance:

(53) X is neither existent nor nonexistent.

If "X" is a singular term, and "being existent" and "being nonexistent" two predicates whose extensions are respectively E and $D \setminus E$, then it seems that "X" should name an individual which is outside of D. At this point, there are two ways of proceeding.

One is to follow the spirit of Meinong's first move into a double-domain model. One would add a domain D_1 which includes D (which includes E). Of course, one can then construct a third-level negative existential which says of X that it is neither in E, nor in D, nor in D_1. One would thus add a domain D_2 which includes D_1 (which includes D, which includes E). And so on... The resulting Meinongian model contains an infinite well-founded set of domains: $< \{D\}_\omega, E, v >$. But this really sounds implausible, or utterly *ad hoc* if it is motivated only by the acceptance of NA and the truth of (20).[26]

[25] As proposed in [Priest 2008b].

[26] Consider, to boot, that one would be able to define as many inter-definable pairs of quantifiers. Such infinitely large models have been explored in [Rescher 1960]. As far as I can tell, Rescher has not defined any dual restricted quantifier in the above sense, though. However, I could not read the whole of Rescher's paper which is not freely accessible.

The other is to argue that (53) does not mean anything, for there is nothing which is neither existent nor nonexistent. One would thus try to paraphrase away the problematic second-level negative existentials, as logical orthodox do with first-level negative existentials. But why should one accept first-level negative existentials and stop there? The disagreement between the logical orthodoxy and heterodoxy becomes purely verbal if one follows this second strategy. One can translate the logical orthodox idiom "exist" into the heterodox "is existent or nonexistent" and both views agree on everything.

18.2.3 Down with the doxies!

The logical orthodoxy tried to hold on to CQT by dropping NA. Consequently, the kind of semantic analysis it provides for negative existentials relies on paraphrase. There are many available paraphrases in the literature. One can arguably devise a paraphrase to fit to our intuitions, but, as I have shown, this is bound to be very convoluted and sound utterly *ad hoc*. Negative existentials are recalcitrant statements for the logical orthodox, because they are committed to the denial of NA.[27]

By contrast, the logical heterodoxy was prompt to abandon CQT in order to stick to NA. On top of the many metaphysical objections the logical heterodoxy has to meet at some point, I added two technical worries which are not discussed in the literature. On the one hand it seems that the logical heterodox should handle a multiplication of quantifiers and a multiplication of domains of quantification. This is not a formal problem, for one can develop formal languages with multiple quantifiers and domains. But, as it appears from the somewhat local problem of negative existentials, it seems that dropping CQT amounts to opening Pandora's box. One might try to avoid doing this and stick to a manageable theory of quantification which nicely connects quantification and existence at a conceptual level. In particular, I insisted that these formal sophistications dangerously drive one away from natural language. Since negative existentials are a natural language phenomenon, this route can be thought of as wrong-headed.

At this point, one should be very much inclined to explore the possibility of keeping both CQT and NA together. This can be done, as free logicians have shown.

[27]Moreover, the method of paraphrase cannot handle quantified negative existentials *at all*. So even if one handles negative existentials in keeping with logical orthodoxy, the solution will not generalise. For a discussion of this point, see [Kroon 2003].

18.3 Free existence

> To understand the concept you should think of free as in free speech, not as in free beer.
>
> Richard Stallman

18.3.1 The core idea of free logics

Here is, in one sentence, Lambert's definition of a "logic free of existence assumptions with respect to its terms, general or singular" (FL):

> It is a logic in which the quantificational phrases "every" and "some", and their stylistic variants, have their classical interpretations and there are no statements that are logically true only if it is true that **G** exists for all general terms **G**, or it is true that **s** exists for all singular terms **s**.[28]

FL is thus in line with CQT. But at the same time, in FL one cannot prove a theorem using the assumption that s exists simply because s is a singular term. This is tantamount to saying that FL is a language in which one can use both denoting and non-denoting terms. Here is Nolt's way of putting it: "free logics are logics whose singular terms need not denote members of the quantificational domain."[29]

Axiomatizing FL should thus include two axioms which meet these requirements. One needs to introduce a singular existence predicate $E!$[30] to formulate these axioms:

(54) $\forall x E!x$

(55) $\forall x Ax \rightarrow (E!t \rightarrow A[t/x])$

(54) is the formal rendering of (43). (55) is a weakening of the of Universal Instantiation (UI) which says, roughly, that one can substitute a singular term for a variable in a universally quantified formula only under the express condition that the singular term denotes.

Given that both quantifiers are inter-definable, this weakening also affects Existential Generalisation (EG). Accordingly, FL is designed to block the inference which yields the incompatibility between CQT and NA. Let us formalise the argument given in subsection 18.1.3.

Following NA, (20) gets translated into:

[28] [Lambert 2003], p. 124.
[29] [Nolt 2007], p. 1057.
[30] It reads "ee shriek" or "ee bang"!

(56) $\neg E!(ew)$

The next step of the argument is an instance of Existential Generalisation, substituting $\neg E!$ for A:

(57) $At \vdash \exists x Ax$

From (56) and (57) we get:

(58) $\exists x \neg E!x$

which contradicts the central tenet of (43) (and axiom of FL) $\forall x E!x$.

But given (55), EG should be weakened. (57) is thus revised so that:

(59) $At, E!t \vdash \exists x Ax$

It says that one can infer that there is an individual x such that Ax follows from At for t a singular term, only under the condition that t exists. Given this weakened version of existential generalisation, the problematic argument does not go through.

Hence, the weakening of existential generalisation (which is definitional of FL) is enough to be able to hold both NA and CQT. This puts FL in stark contrast with both the logical orthodoxy and heterodoxy. It is thus possible in FL to hold that everything exists ($\forall x E!x$) and that it is not true that Emma Woodhouse exists ($\neg E!(ew)$).

18.3.2 What is singular existence

The trick is thus done by introducing a logical predicate $E!$ obeying axioms (54) and (55). In particular, (54) says that the predicate of existence is true of everything. The now pressing philosophical question is: what does $E!$ mean?

There are two available interpretations of $E!$ in the literature. The first comes from Jaakko Hintikka and is by far the most popular. The second comes from Henry Leonard and is utterly unknown. I will argue against the first and then show how the second can be given an interesting intuitive interpretation.

The identity way

In [Hintikka 1959], it is shown how one can construct a predicate whose extension is trivially the whole domain of quantification using identity in the following manner:[31]

[31] Hintikka argues that his definition of singular existence is the formal counterpart to Quine's famous statement that "to be is to be the value of a variable", which he takes to be an interesting result.

(60) $E!t =_{def} \exists x(x = t)$

According to this definition, "everything exists" is a theorem of quantification theory with identity:

(61) $\forall y E!y =_{def} \forall x \exists y(x = y)$

The first critical remark about this definition of singular existence is that it makes negative existentials very difficult to understand. According to the definition, not to exist is defined as:

(62) $\neg E!t =_{def} \neg\exists x(x = t) \equiv \forall x(x \neq t)$

In particular, (20) means: "There is nothing which is identical to Emma Woodhouse". The least we can say is that this is a difficult sentence to parse. So it seems that a perfectly intelligible statement gets translated into a barely comprehensible one, using some technical jargon that one might want to avoid. This is a minor point, but, as we will see, Leonard's interpretation does not suffer this problem at all.

The second line of criticism concerns the suggested primacy of identity over existence. Indeed, it is not clear why identity (=) should be thought of as more primitive than existence ($E!$) so much that one can define the latter using the former as Hintikka suggests. There are no philosophical reason to think this way and there are even two reasons to resist such a reduction.

First, it is possible to construct a FL without identity. It has been done for the first time in [Lambert 1963].[32] [Bencivenga 1986] even made this point explicitly this in his introductory definition of free logics (my emphasis):

> A free logic is a formal system of quantification theory, *with or without identity*, which allows for some singular terms in some circumstances to be though of as denoting no existing object, and in which quantifiers are invariably thought of as having existential import.

Consequently, Hintikka's definition does not work for all languages. Interestingly, there is a stronger result shown in [Meyer, Bencivenga, and Lambert 1982] which says that $E!$ cannot be defined away in a quantification theory without identity. This suggests that $E!$ is indeed more primitive than = in FL.

Second, there are some conceptual reasons to believe that $E!$ should actually be thought of as more primitive than = even though it is technically derived. In [Scott 1979] for example, it is shown that the reflexivity axiom for identity requires

[32]For a standard presentation of FL without identity, see also [Priest 2008a] §13.2–5 (and then adding identity in the following paragraphs).

a predicate of existence. Indeed, it is somewhat controversial to think that every nonexistent object is identical to itself but everyone agrees that every existent object is identical to itself.[33] Thus, existence can be thought of as conceptually primitive and necessary for a good theory of identity, despite what Hintikka's definition suggests.

The modality way

The second interpretation comes from Leonard's meditations on Descartes's *cogito* argument in "The Logic of Existence":

> Now the tacit presupposition of the modern logic would be expressed by writing this formula as a postulate, true of any x, or by deducing it as a theorem:
>
> $$E!x =_{def} \exists\phi(\phi x)$$
>
> That is, to say that x exists is to say that x has at least one property.
>
> Shades of Descartes! "I think, therefore I am."
>
> [...]
>
> Notice that Descartes starts with "I think." There are other predicates than thinking with which he *might* have started. (Whether *he would* have regarded them as producing equally evident premises, I do not know or care. I should have been willing to grant them to him.) For example he might have said: "I am I." Or again said: "Either I think or I don't think." In other words, he might have taken *any* law of thought and affirmed it relative to himself, and I should have been willing to accept the affirmation as true.
>
> But I should *not* have granted that therefore Descartes existed. In other words, existence is not implied by necessary, or analytic, predicates. It is, rather, a consequence only of *contingent* truths. That I think is a contingent truth, rather than a necessary one, and I say this despite Descartes' *later* claim that thought is of my essence.[34]

Given standard notations of second-order, modal logic, we thus get Leonard's definition of singular existence:[35]

(63) $E!t =_{def} \exists\phi(\phi t \wedge \Diamond\neg\phi t)$

[33] See also [Lambert 1981] for a further discussion of the *substitutivity of the biconditional*, pp. 158-160.

[34] [Leonard 1956] pp. 56-8.

[35] One could be surprised by this sudden call for second-order logic here. But this is not that surprising. Indeed, consider the other standard logical predicate, namely identity (=). Identity is

We can thus define nonexistence in the following way:

(64) $\neg E!t =_{def} \neg\exists\phi(\phi t \wedge \Diamond\neg\phi t) \equiv \forall\phi(\neg\phi t \vee \neg\Diamond\neg\phi t) \equiv \forall\phi(\phi t \rightarrow \Box\phi t)$

Which says: t does not exist iff all the predicates which are true of t are necessarily true of t.

I think this definition of nonexistence is plainly intelligible and very intuitive. It seems that it captures what "being an artifice of a theory" or "being merely supposed". Indeed, one can stipulate or suppose that some individual exists. But the hypothetical individual *actually* exists only if the underlying theory or what is supposed is *actually* true. Otherwise, the hypothetical individual remains an artifice of the theory or a mere supposition. In such cases, the properties the individual has are all and only those properties speculated as well as the logical consequences of these speculations. The individual holds these properties by definition, hence necessarily.[36]

It might be objected that some intentional properties are contingently true of, say, Vulcan. For instance, it is contingently true that:

(65) Le Verrier thought the discovery of Vuclan would make him famous.

I think that these statements are not as bad for Leonard's definition as one might first think. Indeed, (65) implies that Le Verrier thinks that Vulcan exists. From Leonard's viewpoint, this is explained by the fact that he thinks Vulcan has a contingent property (that of making him famous in the near future), over and above the necessary properties predicted by the theory. So there are two distinct perspectives on Vulcan at play: that of Le Verrier believing that Newton's theory is right; and that of the enlightened astronomer knowing that Newton's theory is inaccurate. Within each perspective, Leonard's definition applies and makes good predictions. This suggests that the theorizing of intentional perspectives is compatible with Leonard's definition, despite *prima facie* counterexamples like (65).

---

usually taken to be defined by Leibniz's law which is the following second-order logic formula:

$$\forall x\forall y\forall\phi((\phi x \leftrightarrow \phi y) \leftrightarrow x = y)$$

Note that Leibniz's inspired definitions of identity can also be expressed in languages different than second-order logic. Martin-Löf's type theory is a case in point. See also the recent development of Homotopy type theory in which Voevodsky and others gave a precise definition of identity in [Univalent Foundations Program 2013]. I do not know if and how the same treatment could apply to Leonard's definition of singular existence, but it would be a very interesting result if it could.

[36] I suppose Williamson made the same observation independently in [Williamson 1990] while reflecting on the notion of "objecthood" as used in [Wittgenstein 1921]. He remarks that "objecthood is not obviously subject to contingency at all." He calls this the "principle that all objecthood is necessary objecthood".

In the remainder of this chapter, I will combine Leonard's definition of singular existence with pretence semantics to account for fictional names, building on this intuition. Before I can do this, I need to give a formal semantics to FL which will ne used to interpret external perspective statement. This is what I will do now.

18.4 Towards free pretence semantics

18.4.1 Choosing the right free semantics

If free logicians all agree at the proof-theoretical level in weakening EG and UI, there is no formal semantics which has been agreed upon. At first, intuitions were enough, "but, eventually, the task of defining a formal semantics could no longer be postponed; and here is where the conceptual issues started (however slowly) to come to the fore".[37]

The main problem is that the requirements of free logic seem to challenge extensional semantics. Indeed, in a classical setting, one uses an interpretation function I to define extensions for predicates and singular terms such that: for every predicate, $I(P^n) \subseteq D$; for every singular term t, $I(t) \in D$. Using this extensional interpretation of the language, one can then define truth-conditions for atomic formulae in the following manner:

$$v(P^n(t_1, ..., t_n)) = 1 \text{ iff } \langle I(t_1), ..., I(t_n)\rangle \in I(P^n)$$

But in FL, this is not possible, for it is not the case that *every* singular term t denotes a member of the domain of quantification. In other words, I is a partial function which is undefined for the empty terms of the language. So there is no straightforward way to give general truth-conditions for atomic formulae in FL. Indeed, when I is undefined, the classical truth-conditions cannot apply. The problematic cases are the simple external perspective statements like:

(17) Emma Woodhouse is a fictional character.

For (17) gets translated $F(ew)$. But ew, by definition, does not denote.

In order to interpret atomic formulae, some philosophical choices have to be made and several technical apparatuses have been designed to implement these choices. These discussions have given rise to different versions of free semantics.[38]

[37] [Bencivenga 2006], p. 297.
[38] See [Bencivenga 1986] for a seminal discussion of the different versions of free semantics.

One choice consists in saying that atomic formulae containing a non-denoting term are *always false*, for that reason it has been termed "negative" FL. The core intuition driving negative FL is to take at face value the biconditional of the classical truth-conditions. A formula is true whenever the extensional condition is satisfied (hence when I is defined), *and false otherwise*.[39] Though negative FL is a popular choice, it does not square well with our intuitions about external perspective statements involving a fictional name. Indeed, (17) appears to be *true*. But negative FL predicts it is false.[40] So we should look for another version of FL.

A second choice consists in saying that when I is undefined, there are no truth-conditions. In other words, atomic formulae containing non-denoting terms are truth-valueless. For this reason, such free semantics has been called "neutral" FL. The natural way of giving a semantics for neutral FL consists in assigning arbitrary classical truth-values to all atomic formulae containing a non-denoting term so as to get all the possible ways I could be completed. Once such completions are constructed, one can then define a super-valuation function over the set of completions.[41] Given that (17) is true and not truth-valueless, it follows neutral FL is not an option to model external perspective statements involving fictional terms.

18.4.2 Story Semantics

The last choice, that of "positive" FL, consists in providing truth-conditions when I is undefined in such a way that some atomic formulae containing a non-denoting term will be true and some other false. One version of positive free semantics is given in [Lambert and Fraassen 1972] under the name of "story semantics". It consists in defining a *story* S, which is the set of all the true atomic formulae containing some

[39]This idea goes back (at least) to the work of the medieval philosopher Buridan (see his *Sophismata* – §1.6.5). The modern motivation comes from the treatment of definite descriptions in [Whitehead and Russell 1912]. Negative FL was formalised for the first time in [Schock 1968]. Some other influential philosophical motivations are given in [Burge 1974]. For a discussion of the main philosophical objections to negative FL, see [Bencivenga 2006], pp. 297-8.

[40]Sainsbury argues for the adoption negative FL in [Sainsbury 2009], remarking that:

> [Negative free logic] leads to a very natural treatment of existential claims. "Pegasus exists" is false, just like "Pegasus flies", since the name is empty. So its negation is true. We typically express this as "Pegasus does not exist".

He does not comment upon other external perspective statements.

[41]This strategy was first given in [Van Fraassen 1966a] and [Van Fraassen 1966b]. See also [Bencivenga 1986] for a now standard version of the supervaluation free semantics which departs substantially from that of Van Fraassen, fixing problems about identity statements involving empty terms.

non-denoting terms. One can thus define free truth-conditions for atomic formulae in general:

(i) $v(P^n(t_1,...,t_n)) = 1 \begin{cases} \langle I(t_1),...,I(t_n)\rangle \in I(P^n), \text{ if } I(t_i) \text{ is defined (for } 1 \leq i \leq n) \\ P^n(t_1,...,t_n) \in S, \text{ otherwise.} \end{cases}$

As for the truth-conditions of other (complex) formulae, they are defined as usual:

(ii) $v(E!t) = 1$ iff $I(t)$ is defined.

(iii) $v(\neg A) = 1$ iff $v(A) \neq 1$

(iv) $v(A \wedge B) = 1$ iff $min(v(A), v(B)) = 1$.

(v) $v(A \vee B) = 1$ iff $max(v(A), v(B)) = 1$

(vi) $v(A \rightarrow B) = 1$ iff $v(A) \neq 1$ or $v(B) = 1$.

(vii) $v(\forall x Ax) = 1$ iff for every individual constant t, if $I(t)$ is defined, then $v(A[t/x]) = 1$.

The "story" of story semantics should not be thought of as the set of fictional propositions.[42] As such, "story" is a misnomer.

The story S rather contains all the atomic external perspective statements. In S, one should find "Emma Woodhouse is a fictional character" and the like. Once we have such an S, the external perspective statements will come out true.

So story semantics delivers the truth-conditions of external perspective statements in an extensional setting, but it does not really explain how we get them. Everything is hidden in the "story". The natural question is thus: how do we get to construct S in a systematic manner? So as to use fictional names in serious discourse afterward. Intuitively, what happens is that the name "Emma Woodhouse" is introduced into the language of FL (the language used for serious discourse) as an empty name *because* we have come to know about her in the fiction. So S should somehow be derived from the set of fictional propositions we have about Emma. This points towards a combination of pretence semantics and story semantics.

[42]Here, I depart substantially from [Lambert and Fraassen 1972]'s initial motivations (pp. 180-1) in which they want to model sentences like "Pegasus flies". I will not use story semantics to model fictional sentences because such sentence do not have truth conditions, but fictionality conditions.

18.4.3 Combining pretence semantics and free semantics

From existentially creative pretences to fictional names

Pretence semantics is designed to handle fictional statements, by providing fictionality-conditions. These fictionality-conditions are the result of principles of generation which transform real propositions about props into fictional propositions and generate the implicit fictional propositions from the explicit ones following general patterns of inference.

In this setting, we have seen that it is possible to import real names into the pretence by using the reality principle. This is how Tolstoi introduced the name "Napoléon Bonaparte" into his novel *War and Peace*. One can also introduce new names. This can be done in two ways: either the introduction is explicit by using a principle of type (T1) to this effect: "Let us say there is a witch called Wendy". Or it is done implicitly, the principle of appearances: this is the case of the practicing boxer. Jane Austin's *Emma* introduces the name implicitly, for she uses the name in the first sentence of her novel, presupposing that there is such a woman within the fiction. When one introduces a name within the pretence, either explicitly or implicitly, one defines what Evans called an "existentially creative pretence".[43]

I suggest that each time there is an existentially creative pretence, one stores the information about the names introduced and how they are introduced in a story S. Such information about the origin of empty names is what goes into the story semantics presented above. In other words, introducing an empty name into the language comes with some information about that name. This is built into the semantic apparatus so as to get the story semantics given above.

Proto-semantics and fictional terms

I think one can use some ideas from Antonelli's "proto-semantics" as presented in [Antonelli 2000] to explain how it can be made to work in detail. Antonelli's idea is to give a bivalent, extensional positive free semantics by introducing a linguistic parameter to relativise truth-conditions. The parameter is attached to empty terms, "so that one can speak, in analogy to modal logic, of *truth at* a term t".[44]

In order to formalise this idea, he introduces the what he calls a *proto-interpretation*. Here is an informal description of his strategy:

[43] It is not the only way of defining an existentially creative pretence, for one can use existentially quantified formulae and no name as discussed in subsection 6.2.2.

[44] [Antonelli 2000], p. 279.

> For example, consider a language containing an individual constant c and a one-place predicate symbol P. A proto-interpretation for such a language is a function π that (among other things) assigns a signed extension to P, where a "signed extension" is a subset S (of some given domain D of objects) associated with a special marker $+$ or $-$. [...]
>
> Once a class of signed extensions has been identified (by whatever device should turn out to be convenient), and thereby also a class of proto-interpretations, an interpretation for the language is defined as an assignment of a proto-interpretation to each term. Truth then works as follows: where π is the proto-interpretation assigned to c, and S the extension π assigns to P, the atomic sentence $P(c)$ is true on the given interpretation if and only if: either c denotes some object d and $d \in S$; or c does not denote, and S has a positive sign $+$. In the general case where predicates can have more than one argument, proto-interpretations are assigned not only to terms but to n-tuples of terms.[45]

The resulting formalisation is thus very similar to that of story semantics, but there is more structure. Indeed, instead of a set of sentences S, the atomic formulae containing empty terms are identified by signed extensions, that is at the sub-sentential level. Conversely, S can be seen as the resulting set of all the atomic formulae which we can get by associating the right non-denoting terms with the signed predicates.

I am not sure how this can apply across the board, for all predicates and terms of the language. But I think one can make explicit how the idea applies to important metafictional predicates used with fictional terms. Pretence semantics should be seen as a "device [that] turns out to be convenient" for identifying signed extensions for such predicates.

What it takes to "be fictional": Take the predicate "being fictional" F. The signed extension of F should be the empty set $\emptyset$ (given the intuitive meaning of F) and it is marked $+$. We thus have:

$$\pi(F) = (\emptyset, +)$$

This means that for no denoting term t is it true that Ft; but for some non-denoting term n, it is true that Fn. Now, we have to make sure that, among all the non-denoting terms, only the *fictional* ones get marked for F. Let us call the fictional terms n_f. It amounts to explaining how the following obtains:

[45] [Antonelli 2000], p. 279-80.

$$\forall n_f,\, Fn_f \in S$$

The answer is straightforward, given the discussion about existentially creative pretences given above. Indeed, each fictional term comes with a marker which says "origin: existentially creative pretence" (and probably more). For instance, "Emma Woodhouse" (ew) is thus marked. Postulating such a marker is quite plausible given the fact that if one does not know that "Emma Woodhouse" is a fictional name, one cannot be said to have properly *understood* the term. Hence, the putting Few into the general story S comes from the fact that F is marked for ew. This delivers the adequate truth-conditions for (17).

What it takes to "be nonexistent": A more interesting example is to consider the predicate "not existing" $\neg E!$. The signed extension of $\neg E!$ is, by definition, $\emptyset$ and it is marked $+$. We thus have:

$$\pi(\neg E!) = (\emptyset, +)$$

We now have to make sure that $E!$ is marked for every non-denoting term n. It amounts to explaining how the following obtains:

$$\forall n,\, \neg E!n \in S$$

Here, Leonard's definition of nonexistence comes out handy. Recall, by Leonard's standard, n does not exist iff all the predicates which are true of n are necessarily true of n. If we allow the plausible idea that markers are yield necessary or *a priori* truths, then it follows that for n not to exist is equivalent to "n being true of nothing else but what it is marked for".[46] For all fictional terms, by such standard, the corresponding negative existential is true. Indeed, the only true formulae involving a fictional term are external perspective formulae and these, by definition, are the ones which are marked by their origin.

This delivers the following truth-conditions for $\neg E!$:

> $v(\neg E!t) = 1$ iff each time ϕt is true, either it is a complex predicate, or it is an atomic formula and $\phi t \in S$ (i.e. ϕ is marked for t).

Interestingly, given the clause (ii) of story semantics, one can then use singular existence to predict the partiality of the interpretation function I. Indeed, the interpretation function will be undefined whenever the term does not exist. That is, whenever all the predicates which are true of it are those which are marked for it.

[46] This interestingly qualifies Leonard's definition in this setting. It suffices that all *atomic* formulae are necessarily true. Given that the truth of complex formulae is given recursively, I think it does fit Leonard's spirit, if not the letter.

This, I think, is really the essence of the meaning of $E!$ in FL. $E!$ is a "test" predicate which, when applied to a singular term gives 1 when the term has an extensional interpretation and 0 otherwise.

Naturally, one should add dynamics to the picture. Indeed, one can come to know that there are true atomic formulae about some term which brim over the predicates it marks (what is given by its origin). Discovering this is tantamount to discovering that the term denotes, i.e. that the individual at issue exists. I think this is a very good prediction of the theory.

18.5 Conclusions

In this chapter, I gave a way of combining pretence semantics and a version of free semantics, which is called "story semantics". This combination completes the defence of anti-realism, crucially using the tools of a functionalist theory of fictional "truth". I have thus shown how a general semantics for fictional discourse can be constructed in keeping with the pre-theoretical intuition that fictional characters do not exist and fictional events are not facts of the world.

First, I gave a motivation for introducing free logic which is independent of the analysis of fictional and metafictional sentences. It consisted in sending back to back the two main traditions which I called the logical orthodoxy and heterodoxy on the semantic analysis of negative existentials. Then I presented the core features shared by all free logicians from a proof-theoretical perspective. The crucial point consists in introducing a logical predicate for singular existence to weaken the rules of Universal Instantiation and Existential Generalisation. I then discussed two philosophical interpretations of this predicate, rejecting Hintikka's and adopting Leonard's. At this point, I had an axiomatic system and an intended interpretation of it: this is what I needed to choose between the many different free semantics that have been devised by free logicians.

Given the data I needed to model, only a version of positive free semantics could be useful. I introduced Lambert and Van Fraassen's "story semantics", a bivalent extensional semantics for positive free logic. I showed how it can easily be combined with pretence semantics to handle external perspective statements and give truth-values to them. The hinge is Evans's notion of existentially creative pretence. The intuitive idea is that when a fictional term is used in a metafictional context, it carries with itself some information about its origin. Postulating that such information is available for the semantics of atomic formulae is not unreasonable, given the fact that part of what *understanding* a fictional name is consists in knowing that it originates in an existentially creative pretence.

The formal apparatus of story semantics is somewhat rudimentary. One would like to know how the set of external perspective sentences becomes available. In order to answer this question and give some insights into the mechanism in the case of fictional names, I introduced the G. Aldo Antonelli's ideas which are formalised into his "proto-semantics". It is in essence the same as the story semantics except for the fact that it operates at a sub-sentential level. I showed how it can be used to give a precise analysis of what "being fictional" means as well as a way of interpreting "being nonexistent" in keeping with Leonard's definition.

This last development showed how interesting, from a philosophical viewpoint, a free pretence semantics can be.

As for further development, I think this framework should be extended so as to model two other pieces of data. One is the case of quantified negative existentials. Since the paradigmatic cases of true quantified negative existentials are derived from singular negative existentials involving fictional names, it seems that free pretence semantics should apply quite easily. However, the extension is not trivial for it requires to change the interpretation of quantifiers so that they can (momentarily at least) quantify over non-denoting terms.

The other is the application of the framework to ordinary empty terms. Everything that has been said here crucially involved the fact that fictional names are special kinds of empty terms, for they are *de jure* empty. Removing this premise might threaten several of the arguments given in this chapter. In principle, thus, many of what has been said about fictional names should apply *mutatis mutandis* to ordinary empty names, but I do not know to what extend some intermediary results are specific to fictional names. This going over the present defence of anti-realism with a fine-tooth comb with ordinary empty names in mind is an interesting, non-trivial task for the future.

Appendix B

Parafictional statements as metadata

In computer programming, there is an important practical distinction between the "data" and "metadata", i.e. information which is contained within a file and information about the file. At the level of the hardware, everything is a file, though. Metadata is a file, or part of a file. Only what contains the file is interesting because it tells something about another file, or part of a file. At the level of applications, i.e. the software level at which the human-machine interaction takes place, the distinction between data and meta-data is typically made visible. In other words, most of the time applications do not read and execute files containing data and metadata in the same manner.

Here is a first, very simple example. An email is a file which is designed to be sent *via* the internet protocol. The data of an email is its informational content, let us suppose it is a set of English sentences. The metadata of the email is a file which contains information about the email, typically answering to the following questions: "When was it sent?", "Who sent it?", "To whom?" It may also contain some more information about the machine on which it was written, the application used, the history of modification, etc. Usually, the email is a big file containing both data and metadata, travelling the internet.

At the level of the application, the difference between data and metadata is conspicuous. Indeed, what is of most interest for humans is the data. So the set of sentences is usually at the core of the layout of the email and the metadata is not straightforwardly visible. Of course, for the machine, the metadata is much more important. Your email application needs to "know" where the email comes from and its date so that it can be put in the right box in the layout and appear on your

screen where you expect to find it. Some metadata is useful for the user, such as the history of sendings and replies, the date, the participants in the discussion, etc. So it is generally easily accessible, although nicely disconnected from "the body" of the email.

Here is a little more sophisticated example. A digital photograph is a file which contains as data a long series of coding of how pixels on a screen should behave. It contains a lot of metadata so that the data can eventually become an image. One crucial metadata is the file extension (things like ".png") which is used by the machine to "read" the file using an appropriate application, i.e. to decode the large series of 0 and 1 so as to produce an image on the screen. If such metadata is missing, one would simply have to guess what application should be used, which can be, in practice, a very difficult guess. Much of the metadata comes from the camera which produced the photograph (if it originates in a camera). Such metadata typically include the signature of the camera, date and time, the aperture value, the exposure time, the focal length, etc. Such information might prove useful if one wants to work on the quality of the image. But in typical use, such information is not useful and this is why most applications will simply display the image and hide the metadata.

Interestingly, in this example, metadata is usually to be found in a text file. That is, its extension is crucially different from that of the image. Consequently different applications are needed to process the data and the metadata. If most image viewer include a text editor, the other way round is not usual. So we need the cooperation of different applications.

It is not so rare that pictures contain, in the metadata, a miniature reproduction of the photograph. The quality and framing of such an image is, for obvious reasons of space, typically very poor. But still, some pictorial information can be present in the metadata file. In order to see this image, one needs an image viewer.

The data and metadata are really independent files. It is possible to have data without metadata and it is possible to have metadata without data. By which I mean that it is possible to erase metadata without touching the data and the other way round. Of course, the two together are much more useful and workable than one alone. Indeed, metadata is crucial to the overall reliability of applications and operating systems. But in some cases, one is enough. For instance, in digital surveillance metadata of everything that is sent through the internet is systematically copied alone.

Now, I propose that we see parafictional statements as the metadata of a photograph which contains a miniature illustration of the photograph. It thus require something like a text editor and something like an image viewer to be read: the latter

is embedded in the former. The analogy with interpreting a parafictional statement is the following: one needs a module to interpret serious discourse and another module for interpreting pretence: the latter is embedded in the former. This is why parafictional statements have this unmistakable mixed perspective on the fictional events.

Appendix C

Pseudo-selves as virtual machines

A *virtual machine* is the emulation of a computer system at the functional level. It is a duplicate of a real computer machine within another computing architecture. Concretely, while running a physical computer, one starts a program which emulates an other operating system.

The *host* is the term used to refer to the operating system in which the emulated program runs. If it is not a virtual machine, the host is typically a physical machine containing hardware, i.e. a processor and many different kinds of transistors. The *guest* is the term used to refer to the emulated operating system. The guest's architecture is not so connected to a hardware. As just hinted at, such constructions can in principle be embedded at will.

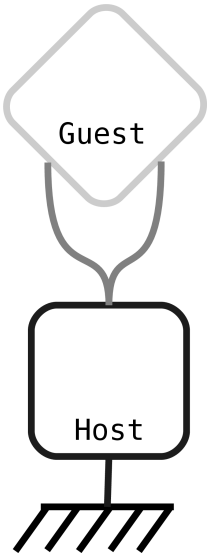


Figure C.1: Diagram of a virtual machine

If at the root of the guest's architecture there it no hardware, it does not mean that there is nothing. Indeed, all the computing that can be done within the guest's operating system is physically done on the same physical hardware as the one used by the host. So crucially, the guest is using the host's computational capacities. In order to do this, the host must allocate some of the hardware computations to the guest. Questions of priority become thus very important. Such questions raise very difficult theoretical problems when it comes to the underlying algorithms of virtual machine programs

whose aim is to allocate resources efficiently.

Both the host and the guest usually come with their own user's interface. When you are in the guest's user interface, the host is invisible, so to speak, i.e. you cannot change nor access the host's files through the guest's operating system. But it is there: it is its computing capacities which are exploited. Such presence is palpable when, for instance, the host is running a program which takes most of its CPU power: the guest will be significantly slowed down. Of course, one needs to define a way to switch back to the host. The simplest way is to shut down the virtual machine program.

Note that the operating system of the guest need not be the same as the host's. Indeed, virtual machines are especially useful when you want to run a program which is only compatible with a different operating system than the one running on your physical computer.

The root of a computer architecture can be seen as a self. The root of the host is an empirical self, in that it is directly connected to the hardware (as shown in Figure C.1). By contrast, the root of the guest is a pseudo-self, for the connection with the hardware is merely emulated. This makes a solid analogy between pseudo-selves and virtual machines.

Concluding remarks about anti-realism.
Or the story of how some (sane) intuitions met some (insane) philosophical arguments in the second half of the 20th century

There is no way I can conclude this dissertation by claiming that I have laid down the last word on any of the subjects I touched upon. The arguments which I have put forward are not knock-down in any sense. If some are conclusive, they will certainly be argued against in the near future. As for the proposals, they await being fleshed out by my future self or other philosophers of fiction. I will therefore not conclude about the content and will simply hope that my contributions make their way into the current debates.

That being said, I think it is fair to say that the running theme of this dissertation is a position one might call "anti-realism". I tried to take an anti-realist stance on fiction and go as far as I could. As concluding remarks, I would like to locate these kinds of intuitions in the philosophical landscape I have been working in, for I think the history of the second half of the 20th century is very confusing. This little clarification will also help understand who are the main philosophical characters in the philosophy of fiction, acting behind the scenes so to speak.

Definitions

Let us start with simple definitions. *Anti-realism* is the denial of *realism*.[1] "Realism" comes from the Latin word *res* which means "a thing". In general, *realism* is a claim which says that there is a thing behind a given discourse. One is thus realist relative to a kind of discourse. Consequently, there are as many "realisms" as there are kinds of discourse; and as many "anti-realisms" as there are "realisms". The point is nicely made in the opening paragraphs of [Dummett 1982]:

> It is clear that one can be a realist about one subject-matter, and not about another: though someone may have a general inclination towards realistic views, it is plain that there is no coherent philosophical position which consists in being a realist *tout court*. This may be expressed by saying that one may be realist about certain entities – mental states, possible worlds, mathematical objects – and not about others. But it seems preferable to say that realism is a view about a certain class of *statements* – for instance, statements in the future tense, or ethical statements - since certain kinds of realism, for instance realism about the future or about ethics, do not seem readily classifiable as doctrines about a realm of entities. So, in every case, we may regard a realistic view as consisting in a

[1] One could thus prefer to say "non-realism" or "irrealism". I take these terms to be synonymous for my purposes.

> certain interpretation of statements in some class, which I shall call "the given class".
>
> So construed, realism is a *semantic* thesis, a thesis about what, in general, renders a statement in the given class when it is true. The very minimum that realism can be held to involve is that statements in the given class relate to some reality independently of our knowledge of it.[2]

The kinds of discourse I have been interested about are the "fictional discourses", i.e. discourses associated with our fictional practices. More precisely, one can then define what realism and anti-realism about fictional discourse claim. Realists claim that fictional discourse is truth-conditional discourse, i.e. that fictional statements have a truth-value. In other words, they claim that fictional characters and events are *things*, although not *real* things. Anti-realist, by contrast, claim that fictional discourse is not truth-conditional, i.e. that fictional statements do not have a truth-value. In other words, they deny that fictional characters and events are *things*, and they rather say that they are, literally, nothing.

Eliciting intuitions

Though the above definitions are perfectly clear to a philosopher's mind, it is useful to flesh out the intuitions behind these views. In order to do this, I will start with a little anecdote.

A personal anecdote and a joke

We are in the Fall of 2016, at the Grand Palais. There is an exhibition on Hergé, the Belgian cartoonist best known for *The Adventures of Tintin*. There are many people. In one room, there are black and white draft plates (on the left) side by side with their corresponding original coloured plates (on the right): you can see the differences between Hergé's hand-drawing and what will be printed out.

There is a mother with her (8 year-old-ish) son between the plates and me. She tries to have her son interested, bends to him, points to the black and white plate on the left and says:

– Look there is Snowy! (it is Tintin's white Wire Fox Terrier)

Then she points to the right and says:

[2] [Dummett 1982], p. 55.

– And that's how it is in real life![3]

How shocking is this?

To ram the point, here is a joke one can find in the work of Lewis Carroll. It so happens that there is an English idiomatic expression for "realistic" fictional characters. One can say that they are "as large as life, and quite as natural". Lewis Carroll, in *Through the Looking-Glass*, turns the expression into: "as large as life, and twice as natural".

Don't say no, you smiled!

If you are shocked by the anecdote and you laugh at the joke, then you have anti-realist intuitions about the *life* of fictional characters. Indeed, the reason why the first anecdote is shocking and the joke is funny is because fictional characters (and events) are certainly life*like*, but this is an illusion. Nobody is alive. If you change "life" with *truth* and *reference*, you get the core anti-realist intuitions about fictional discourse which I am painfully trying to elicit.

The argument I just made is a "don't say no"-argument. This terminology comes from Jacques Prévert who wrote this line for the film *the Children of Paradise*, by Carné in 1945. Frédérick Lemaître, in trying to seduce a woman he just met in the busy streets of Paris, makes her laugh he then says: "don't say no, you smiled. Oh! It's wonderful: you smiled..." It is, I will show why later on, a very interesting form of argument when it comes to ontology.

Fictionalism is an Anti-realism

In many areas of philosophy, the term "fictionalism" is generally taken to be a synonym for "anti-realism". In fact, *fictionalism* is but one kind of anti-realism, albeit a popular one in many cases. It consists in denying realism about a kind of discourse by claiming that, in the relevant discourse, some fiction is at play. Here is the opening definition of "fictionalism" in general philosophy one can find in the *Stanford Encyclopedia of Philosophy*:

> Fictionalism about a region of discourse can provisionally be characterized as the view that claims made within that discourse are not

[3] In French: "Voilà comment il est dans la vraie vie".

best seen as aiming at literal truth but are better regarded as a sort of "fiction".[4]

Indispensability arguments

This preliminary definition is very interesting, for it emphasises the link between the views called "fictionalisms" and one very influential form of argument in favour of realism. The most wide-spread argument in favour of realism is of the following kind:

(i) Such discourse is true (or has a truth-value).

(ii) Given the semantic analysis of such discourse, if such discourse is true, then there must be Fs (a set of *things*).

(iii) Therefore, there must be Fs.

Arguments of this form are called "indispensability arguments" (IA). Though the reasoning behind an IA is well known for a long time in the philosophical tradition, the precise formulation of the argument schema originates in the work of Quine. He first formulated an IA leading to realism about mathematical discourse in [Quine 1948] and [Quine 1951].

One can see that IAs are valid arguments. As shown in [Yablo 2001], there are four possible responses to an IA. One can deny the truth of (i) or that of (ii); one can also stop uttering the problematic discourse. The fourth response consists in accepting the argument and thus embrace a kind of realism.

This shows that there are three distinct forms of anti-realism. Denying (i) is *fictionalism*; denying (ii) is a kind of anti-realism grounded on the so-called method of "paraphrasing away".[5] This kind of anti-realism is often called "nominalism" after Quine, though it is controversial to claim that the ancient and medieval nominalists are really "nominalists" in this modern sense. The third form of anti-realism can be called "quietism", though it has nothing to do with the eponymous view in Christian philosohy. [Yablo 2001] nicely argues that *fictionalism*, contrary to the other forms of anti-realism, is merely an *implicit* response to an IA in Quine's work.

[4]Eklund, Matti, "Fictionalism", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2017/entries/fictionalism/.

[5]This method was systematized by Russell's treatment of empty descriptions and later given a name by Quine.

A note on IAs

IAs have been tremendously successful in fastening the link between semantics and ontology in the second half of the 20th century analytical philosophy. They have subsequently become central in all kinds of ontological debates.

The philosophical interpretation of IAs has given rise to some debates in metaphysics, mainly revolving around the metaphysical status of scientific theories. Quine's forceful proposal to "naturalise" metaphysics on the ground that IA provides a steady bridge between epistemology and ontology has thus open a whole area of research. I think this focuses too narrowly on one kind of discourses, namely scientific theories, and the philosophical interpretation of IAs should aim for more generality.

Putnam, who has done much in this area to give its width to IAs in philosophy, gave in passing a very interesting interpretation of IAs. In [Putnam 1971], he claims that ground for IAs to work is a kind of intellectual honesty. In other words, someone who claims (or presupposes) that (i) and (ii) are true *and* proceeds in denying (iii) is "intellectually dishonest". I think this is very interesting, for it abstracts away from the kind of language and the intentions of the locutor.

Consequently, I think "don't say no"-arguments such as the one given above should be seen as dual in form to an IA. Honesty goes in both directions, so to speak. One should be honest when one is serious, i.e. one is committed to the existence of the Fs one seriously talks about, given a relevant IA. One should *also* be honest when one is non-serious, i.e. one is committed to the non-existence of the Fs one non-seriously talk about, given a relevant "don't say no"-argument. There is, *prima facie*, no reason to think the "serious language" is more fundamental to ontological debates.[6]

The Varieties of Fictionalism

Since fictionalism is one kind of anti-realism, it follows that there is no fictionalism *tout court*. Here is, indeed, a comprehensive list of recent fictionalist positions one can find in the *Stanford Encyclopedia of Philosophy*:

> More recently, Hartry Field (1980 and 1989), Joseph Melia (e.g. 2000), Mark Balaguer (e.g. 1998) and Stephen Yablo (see especially 2000, 2000a, 2001 and 2002) have defended fictionalism about mathematical discourse

[6] Of course, Quine gives an independent argument so as to promote classical logic as the "canonical language" when it comes to ontology. But such an argument can also be resisted on independent grounds, as free logicians have shown.

(see here the entry on fictionalism in the philosophy of mathematics); Bas van Fraassen (1980) has defended a version of fictionalism about scientific theories; Richard Joyce (2001, 2005), Mark Kalderon (2005) and Daniel Nolan, Greg Restall and Caroline West (2005) have defended moral fictionalism; **Kendall Walton (1985, 1990, 2000), Mark Crimmins (1998), Stuart Brock (2002), Mark Balaguer (1998a), Anthony Everett (2005), and Frederick Kroon (2000, 2004) fictionalism about things like negative existentials, identity statements, propositional attitude reports, and fictional characters**; Peter van Inwagen (1990) and Cian Dorr and Gideon Rosen (2002) fictionalism about ordinary object discourse; Bradley Armour-Garb and James Woodbridge (2015) and Alexis Burgess and John Burgess (2011) fictionalism about truth; and D.M. Armstrong (1989), Gideon Rosen (1990), John Nolt (1986), Seahwa Kim (2005) and John Divers (1999) modal fictionalism (see here the entry on modal fictionalism). Needless to say, this is a just a sample of recent defenses of fictionalist views. See the bibliography for some further references. For some recent overviews of fictionalism, see Caddick Bourne (2013), Divers and Liggins (2005), and Sainsbury (2010).[7]

I boldfaced the "fictionalisms about fictional discourse" on which I will now focus.

Fictionalism about fiction

Puzzling observations

There are IAs about fictional discourse.[8] Hence, there is a view called "fictionalism about fiction", which consists in denying the first premise of the relevant IAs. Now, here are two puzzling observations:

1. "Fictionalism about fiction" is not trivial.[9]

[7]Eklund, Matti, "Fictionalism", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2017/entries/fictionalism/.

[8]Among many other references, see [Thomasson 2003a] (p. 152) for an explicit mention of Quine's original argument and its relation to ontological debates about fictional characters.

[9]I know some philosophers who actually spend several years of their life trying to articulate the view...

2. "Fictionalism about fiction" is a minority view in the philosophical community.[10]

Note that for some obvious rhetorical reasons, nobody wants to call themselves "non-fictionalists about fiction".

Intuitions and arguments

At this point, one can see how philosophical debates substantially depart from intuitions. "Fictionalism about fiction" clearly sounds like an intuitive truism. But there are philosophical *arguments* leading to its denial. In philosophy, the rule is that one cannot fight arguments with intuitions, or jokes as I did above. One should produce counter-arguments.

This dissertation should be understood as an attempt to produce such counter-arguments. It is thus intended to be a defence of anti-realism. Those who want to restore intuitive claims in philosophy should like it.

As for the structure of this dissertation, it is naturally explained when one considers what is "fictional discourse". Fictional discourse comes in two varieties: the *fictional* and the *metafictional*. *Fictional discourse* (in the technical sense) corresponds to discourse *within* the fiction, whereas *metafictional discourse* corresponds to discourse *about* the fiction. For each kind of discourse, one can construct an IA leading to what can be called "fictional realism" and "metafictional realism". Both views are independent, though they share structural features. My defence of anti-realism consists in providing counter-arguemtns on both grounds: in the first part I argued against fictional realism; in the third part I argued against metafictional realism.

Main philosophical characters in this story

There are thus specific philosophers behind the scenes of this dissertation: the ones who formulated the relevant IAs about fiction for the first time.

First, Willard von Orman Quine, as already mentioned above, is a central figure for he is the father of IAs, the origin of all modern realists. However, Quine did not theorise about fiction. His influence in the philosophy of fiction, as I have tried to show in this conclusion, is both huge and indirect. Interestingly, Quine's idiosyncratic position has also been hugely influential. As it happens, he started his career as

[10] I do not have a quantitative study on which I can ground this claim. It is a personal feeling, given what I know of the community.

a champion of nominalism about mathematics (among other theoretical entities), which is, as we have seen above, a form of anti-realism. But he then became a "reluctant realist" (this is Quine's phrase) after being convinced by his own IA about mathematical discourse.

Second, David Lewis who should be thought of as the arch-realist about *fictional discourse* (in the technical sense). Lewis defended fictional realism in two independent steps. First, he defended modal realism, i.e. realism about modal discourse, in [Lewis 1973] and [Lewis 1986]. Then, he reduced fictional discourse to modal discourse in [Lewis 1978]. Part one of this dissertation mainly consists in denying the second step of Lewis's argument.

Third, Saul Kripke who should be thought of as the arch-realist about metafictional discourse. Kripke argued for metafictional realism in two independent steps. First, he gave an argument for direct reference when it comes to names, thus attacking descritivism which was the majority view at the time of [Kripke 1972]. Then, he showed that direct reference about names entails realism about metafictional discourse in [Kripke 1973/2013]. Part three of this dissertation mainly consists in trying to resist the second step of Kripke's argument.

It is difficult to describe precisely what is common to the three philosophers when it comes to ontology. Lewis was a student of Quine's and explicitly claims that his defence of modal realism is Quine's ideas taken to an end.[11] As for Kripke, it is not clear that he was influenced by Quine directly, although his way of defending metafictional realism closely resembles Quine's "reluctant" posture. I think a plausible answer should take into account the fact that both Lewis and Kripke contributed to setting up possible world semantics as an extension of classical logic. Quine's realism heavily depends on his taking classical logic as the benchmark for ontological debates. It might be that classical logic and realism have intimate connections which were subsequently extended with the advance of modal logic. If this is true, then anti-realist's inclinations should force one to adopt non-classical views on logic.[12] This claim, though difficult to back up in detail and quite speculative, is one of the central claim of [Dummett 1982].

[11] See especially [Williamson 2014] for a detailed explanation and a textual proof of this claim.

[12] This much should explain why the constructive parts of this dissertation contain non-classical formalisms: pretence semantics and free logic.

Résumé substantiel en français

Grandes lignes du projet de recherche

Cette thèse de doctorat en philosophie est une défense d'une forme d'anti-réalisme à propos du discours fictionnel. Autrement dit, l'ensemble de ce travail a pour but de développer une analyse sémantique des discours fictionnels qui soit en accord avec les intuitions anti-réalistes, selon lesquelles les discours fictionnels n'ont *pas de condition de vérité* et les événements et personnages de fiction *n'existent pas*.

Au cours de cette élaboration, dans la seconde partie, je m'éloigne des débats proprements sémantico-ontologiques pour proposer une contribution en philosophie de la litérature. Il semble que les critiques litéraires soient assez régulièrement en désaccord sur ce qui est le cas dans une fiction. Or, être en désaccord sur quelque chose suppose qu'il y a quelque chose. Ainsi, la reconnaissance de cette critique batailleuse semble entrer en conflit avec l'anti-réalisme à propos des discours fictionnels. Cette seconde partie consiste donc à montrer comment on peut se quereller sur l'interprétation d'un texte fictionnel sans pour autant sombrer dans le réalisme.

Le débat entre réalisme et anti-réalisme à propos des discours fictionnels est ouvert et donne lieu à des oppositions très franches dans le champ de la philosophie analytique du langage. C'est ainsi le désaccord, la controverse, la querelle qui est la notion qui traverse mes contributions tant en philosophie du langage qu'en philosophie de la litérature. Pour cette raison, j'ai choisi un titre volontairement incisif qui se prête assez facilement au registre familier. Ainsi, la traduction "être en désaccord sur la fiction" ou "se disputer sur la fiction" rate complètement cet effet. Il faudrait préférer "se mettre sur la gueule à propos de la fiction" qui décrit précisément le sous-entendu familier du titre anglais. Cependant la connotation grossière de cette expression me semble trahir la subtilité de l'original. Finalement, s'il me fallait décider d'une traduction malgré tout, je dirais: "Pas d'accord sur la fiction".

Cette thèse comporte une introduction générale et trois grandes parties d'une centaine de pages chacune, découpées en chapitres et une conclusion générale. Je vais en faire le résumé dans l'ordre ci-dessous.

Introduction: Est-ce bien un problème philosophique?

Dans l'introduction, je montre en quoi le discours fictionnel ne pose des problèmes philosophiques qu'étant donné l'adhésion à des présupposés d'ordre réalistes en

philosophie du langage. Pour ce faire, je m'appuie sur une proposition que l'on trouve dnas [Rorty 1981] et qui consiste à imputer les présupposés réalistes en question à Parménide. Après avoir reconstitué l'argumentaire de Rorty (et souligné l'ironie mordante de l'auteur), je montre les limites de cette proposition. Je propose de repenser la proposition de Rorty dans un cadre plus récent, hérité du programme de recherche initié par Brentano sur la naturalisation de l'intentionalité. Je montre dans ce contexte que les présupposés en question dérivent d'une conception dite "combinatoire" des relations en général. L'interprétation combinatoire des relations, qui est au fondement conceptuel et historique de la philosophie analytique, débouche sur un problème lié à l'interprétation sémantique du discours fictionnel. Si l'on suit l'autre voie, l'interprétation dite "fondationelle" des relations, le discours fictionnel ne pose pas de problème sémantique spécifique. En revanche, l'interprétation fondationelle, qui est au fondement conceptuel et historique de la tradition phénoménologique, débouche sur un problème lié à l'interprétation sémantique du discours factuel.

Les différents problèmes et débats liés à l'analyse sémantique du discours fictionnel discutés dans cette thèse sont donc des problèmes locaux, en ce sens qu'ils n'ont de portée que dans le cadre de la philosophie analytique. Dans la mesure où changer de cadre ne reviendrait pas à résoudre les problèmes liés au discours fictionnel mais à les dissoudre, il serait fâcheux de critiquer le cadre analytique au cours du travail. C'est la raison pour laquelle il ne sera plus mention du cadre théorique après cette introduction. Cette thèse a donc pour modeste objectif de bricoler une solution nouvelle, résolument anti-réaliste, aux problèmes sémantiques formulés à l'intérieur du cadre analytique. Naturellement, si le cadre de la philosophie du langage analytique lui-même était faux ou incohérent, alors mes contributions ne seraient que la pointe avancée de l'incohérence dans le domaine du discours fictionnel. Sous cette hypothèse qu'il faut bien envisager, on pourra encore apprécier l'intégrité de ma position, toute risible soit-elle.

Première partie: vers la sémantique de la feinte

Dans la première partie, je m'intéresse au "problème de la vérité dans la fiction". Ce problème consiste à expliquer le contraste entre les phrases suivantes:

(1) Hamlet est un humain.

(2) Hamlet est un crocodile.

Il y a un contraste évident entre ces deux phrases. À première vue, une manière évidente de décrire ce contraste consiste à dire que (1) est *vraie* dans la pièce de

Shakespeare, tandis que (2) est *faux* dans la même pièce. Cependant, l'idée que le contraste ci-dessus doit être expliqué en terme de valeur de vérité est *très* controversé. Dans le CHAPITRE 1, j'introduis les deux théories en conflit sur ce point. La première, rendue crédible depuis l'article fondateur [Lewis 1978], s'appelle la théorie modale de vérité dans la fiction. La seconde, associée au travail de Kendall Walton, s'appelle la théorie fonctionnelle de la vérité dans la fiction. Dans la suite de cette partie, je commence par donner différents arguments contre la théorie modale puis je contribue à l'avancement de la théorie fonctionnelle.

Avant de me lancer dans la *pars destruens*, je caractérise précisément la théorie modale en analysant la réduction opérée par Lewis du fictionnel au modal. Ce faisant, on pourra m'objecter que je simplifie la pensée de Lewis, ce qui est juste. Cette simplification, cependant, me permet d'embrasser toute une famille de théories modales qui s'est développée à partir des intuitions fondamentales de Lewis et c'est contre ces différentes générations de théoriciens modaux que portent les contre-arguments des chapitres suivants.

Contre la théorie modale

Arguments indirects

Dans le CHAPITRE 3, je développe les arguments dits "indirects" contre la théorie modale. Ils sont indirects car ils ont la forme générale d'un *reductio ad absurdum*: il s'agit de poser l'hypothèse que la théorie modale est correcte puis argumenter en disant que la théorie modale créé en fait plus de problèmes qu'elle n'en résoud. Les deux problèmes qui apparaissent lorsque l'on réduit le fictionnel au possible viennent du fait que les "mondes" fictionnels sont souvent incomplets et parfois contradictoires. Or, si les "mondes" fictionnels se réduisent au mondes possibles (de la sémantique des mondes possibles), alors ils doivent êtres complets et consistants. Cet argument indirect contre la théorie modale consiste donc à dire que la sémantique des mondes possibles, i.e. l'appareil formel créé pour rendre compte de la sémantique des énoncés modaux, est *trop forte*, ou trop contraignante pour l'analyse sémantique des énoncés fictionnels.

Il existe deux manières de répondre à cet argument, qui sont très souvent entremêlées dans la litérature spécialisée mais qu'il est bon de distinguer, selon moi. J'appelle "stratégie de la tête brûlée" la réponse qui consiste à remettre en cause la sémantique des mondes possibles afin d'intégrer des mondes dits "impossibles" ou "non-normaux" (je préfère cette seconde terminologie qui vient de [Berto 2017]). La stratégie de la tête brûlée admet que l'ensemble des mondes fictionnels n'est pas un sous-ensemble des mondes possibles; cependant, elle soutient que l'ensemble des

mondes fictionnels est inclu dans l'union des mondes normaux et non-normaux. Il suffit donc d'étendre la sémantique des mondes possibles de façon à obtenir une sémantique des mondes non-normaux. Ce travail a été fait par plusieurs philosophes dans les dernières années. Je me réfère, pour les besoins de mon exposition, à la sémantique non-normale que l'on trouver dans [Berto and Jago 2019] car elle me semble la plus aboutie et aussi parce que Berto discute expressément le discours fictionnel dans d'autres ouvrages. On a ainsi les moyens, dans les mondes non-normaux, de définir la *vérité* et la *fausseté*. On explique finalement le contraste entre (1) et (2) en terme de vérité dans le monde de *Hamlet* et en général la "vérité" dans la fiction en terme de vérité dans un monde (non-)normal.

Par opposition, j'appelle "stratégie de la tête froide" la réponse aui consiste à garder la sémantique des mondes possibles intacte mais à considérer que les mondes fictionnels doivent être identifiés à des ensembles de mondes possibles, plutôt qu'à des mondes possibles singuliers. La stratégie de la tête froide admet que l'ensemble des mondes fictionnels n'est pas un sous-ensemble des mondes possibles; cependant, elle soutient que l'ensemble des mondes fictionnels est inclu dans l'ensemble des parties de l'ensemble des mondes possibles. Sur un ensemble de mondes possibles, on peut définir la *vérité* et la *fausseté* en utilisant les méthodes bien connues de la super-valuation et de la sub-valuation. La stratégie de la tête froide consiste à dire que la "vérité dans la fiction" peut être réduite à un mélange subtile de super et de sub-valuation sur un ensemble de mondes possibles. Je décris l'algorithme permettant d'obtenir la "vérité" dans la fiction en général.

Ma contribution principale consiste à dire que la stratégie de la tête froide est la plus intéressante, contrairement à ce qui est la plupart du temps affirmé dans la littérature. D'abord, on peut modéliser bien plus que ce que l'on pourrait penser à première vue. Je montre ainsi qu'en suivant la stratégie de la tête froide on peut rendre compte de la "vérité" dans la nouvelle de Priest "Sylvan's box", bien qu'il affirme le contraire dans [Priest 1997]. Ensuite, je pense que cette seconde stratégie ne renonce pas à *expliquer* ce qu'il se passe dans une fiction incomplète ou contradictoire. Il me semble que la technique de la sub-valuation correspond assez précisément à l'idée qu'une fiction contradictoire est une instance d'un phénomène plus général d'injonction contradictoire.

Je situe enfin le débat entre les deux stratégies au niveau de l'interprétation philosophique de ce qu'est une contradiction. La tête brûlée accepte l'idée qu'il puisse y avoir des contradictions qui soient simples logiquement, tandis que la tête froide considère que toutes les contradictions sont complexes logiquement, i.e. articulant deux constituants qui se contredisent. Quoi qu'il en soit, je pense que ce débat est relativement orthogonal au problème de la "vérité" dans la fiction, montrant donc

que la théorie modale est intrinsèquement problématique: elle créé effectivement plus de problèmes qu'elle n'en résoud ici. Je propose finalement de revitaliser la théorie modale par une interprétation fonctionaliste de la sémantique des mondes possibles, en remettant à plus tard cette discussion qui suppose le développement à venir sur la théorie fonctionaliste de la "vérité" dans la fiction.

Arguments directs

Dans le CHAPITRE 4, je développe les arguments dits "directs" contre la théorie modale. Ils consistent à attaquer les présupposés fondamentaux de la théorie modale.

Le premier argument consiste en un retournement de la perspective modale. Il s'agit de nier qu'il faille réduire le fictionnel au modal et au contraire de défendre qu'il faut réduire le modal au fictionnel. Cette approche fictionalisante du modal requiert une définition de ce que peut être un *monde* fictionnel qui de présuppose pas la notion de monde possible. Je propose plusieurs manières pour penser cette notion de monde en général qui se spécialise en mondes possibles et fictionnels.

Le second argument est celui de Evans contre Lewis, énoncé très brièvement dans [Evans 1982] et jamais développé par la suite. Il s'agit de montrer que l'analyse modale de Lewis est incompatible avec une théorie causale du contenu sémantique. En d'autres termes, si l'on regarde bien la paraphrase contrefactuelle du discours fictionnel proposée par Lewis, on doit conclure que le contrefactuel ne sélectionne *aucun* monde possible. L'argument présuppose une théorie causale de la référence et se généralise si l'on adopte une théorie causale de l'information sémantique. En conséquence, il est clair que la théorie causale de la référence et l'anti-réalisme à propos de la "vérité" fictionnelle (ou le fonctionalisme) vont de paire, formant même un cercle vertueux.

Il faut souligner que ce résultat est un résultat d'incompatibilité seulement. Lewis, en refusant la théorie causale de la référence dans plusieurs de ses travaux, se met hors de toute critique sur ce point. En revanche, dans [Lewis 1978], il a des propos qui peuvent prêter à confusion sur ce point. Cet argument d'Evans sera donc retenu contre la théorie modale à la mesure du crédit qu'on accorde par ailleurs à la théorie causale de la référence, qui est un sujet orthogonal.

Pour le fonctionalisme

Je me tourne ensuite vers le fonctionalisme, qui est un anti-réalisme à propos du discours fictionnel et consiste à nier que le contraste entre (1) et (2) soit analysé comme une différence en terme de valeur de vérité. Dans le CHAPITRE 5, je présente les idées fondamentales de [Walton 1990]. Dans le CHAPITRE 6, je discute la formalisation

de ces idées proposées dans [Everett 2013]. C'est l'introduction de la "sémantique de la feinte" (*pretence semantics*). La sémantique de la feinte est un appareil formel qui permet de donner des conditions de fictionalité de manière compositionelle en combinant deux types d'information sémantique. D'un côté, il y a les propriétés réelles des accessoires (*props*) pertinents utilisés dans les jeux de faire-semblant. De l'autre, il y a les principes de génération (*principles of generation*) qui permettent d'importer des informations et de procéder à des inférences. La sémantique de la feinte permet ainsi de donner des conditions de fictionalité à des énoncés comme (1) et (2). Ces conditions de fictionalités miment, simulent des conditions de vérité. Il est important de noter que la sémantique de la feinte est une extension de la sémantique véri-conditionnelle et elle est donc conceptuellement dépendante d'une analyse véri-conditionnelle du snes linguistique.

Dans le CHAPITRE 7, je me propose d'appliquer le formalisme d'Everett au discours fictionnel en allant dans les détails. Cela consiste à chercher le ou les principes de génération qui sont essentiellement liés au discours fictionnel. C'est le principe dit du "reportage" ou du témoignage qui est le plus plausible et proposé par Everett. Cependant, ce principe se heurte à des contre-exemples: des fictions dans lesquelles il est impossible que les événements soient rapportés sous la modalité du reportage ou du témoignage. C'est les problème du "reportage de l'irreportable" (*reporting-the-unreported*). Je résouds ce problème en mettant au jour une structure double de feinte qui me semble caractéristique du discours fictionnel. Je montre ensuite les potentielles connexions que cette structure double entretient avec des problèmes classiques en narratologie. Enfin, je montre que le problème du reportage de l'irreportable risque de se reproduire au second étage de la structure double. Afin d'éviter la contamination, je propose qu'au second étage de la structure, le principe de génération fondamental soit celui qu'Evans nomme le "principe d'apparence", selon lequel les choses *sont* ce qu'elles *apparaissent*. On bloque ainsi les potentiels contre-exemples qui nous pousseraient à une régression à l'infini. La structure double proposée est donc plus complexe qu'un simple enchâssement de feintes, c'est bien une structure où deux jeux de faire-semblant *essentiellement différents* s'imbriquent.

Contenu additionnel

Dans le dernier CHAPITRE 8, je prolonge la réflexion en m'interrogeant sur le statut de l'auteur. Je montre que le fonctionalisme que je défends ne présuppose nullement d'auteur réel mais uniquement un lecteur réel. Il me semble que c'est très utile dans la mesure où l'on peut ainsi se débarrasser de la prérogative de l'auteur sur la "vérité" dans la fiction. Il me semble que des pratiques comme celle des fan-fictions,

en particulier la méthode de la "continuité rétroactive" (*retcon*), nous poussent vers une "démocratie" des lecteurs où l'on peut négocier (parfois à la majorité) la "vérité" dans la fiction. Cela est aussi en lien avec des cas où c'est la justice qui tranche ce qui est "vrai" dans la fiction. Je trouve un cas de différend juridique au États-Unis qui dépend de la doctrine juridique du *fair use*, incluse dans la loi sur le *copyright*. Il semble donc que, parfois, l'auteur n'a pas son mot à dire. Et pourtant le cadre fonctionnaliste fonctionne très bien.

Dans l'APPENDICE A, je fais un pas de côté et montre comment on peut considérer que le problème de la "vérité" dans la fiction (tel que défini dans le CHAPITRE 1) est un problème d'apprentissage statistique non-supervisé.

Le grand débat du scarabée

Dans cette seconde partie, je m'intéresse à un cas d'étude en ce qui concerne la "vérité" dans la fiction: c'est le "grand débat du scarabée" (GDS). Cette controverse littéraire sur l'interprétation de *La Métamorphose* de Kafka vient d'un cours que Nabokov donna dans les années 1950, lorsqu'il était professeur de littérature étrangère à l'Université de Cornell. Ces cours ont été retranscrits dans [Nabokov 1980]. Cette controverse a été retrouvée et nommée par Stacie Friend qui l'utilise à des fins d'illustration dans [Friend 2011].

Dans le CHAPITRE 9, je commence par construire un espace logique pour penser les désaccords en général. Je propose de considérer une première distinction hérité de [Kölbel 2004] entre les désaccords *fautifs* et les désaccords *sans faute*. Typiquement, les désaccords sans fautes sont basés sur des divergences de goût, subjectives. Une seconde distinction, inspirée par [Friend 2011], sépare les désaccords *réels* et les désaccords *fictionnels*, selon que c'est la vérité ou la fictionnalité d'une proposition qui est sujet à controverse. Dans la mesure où le GDS est incontestablement un désaccord fictionnel, il s'agit de se demander si c'est un désaccord sans faute ou fautif. Ce sera la question centrale de cette partie.

Dans le CHAPITRE 10, je réponds à une objection qui consiste à dire que le GDS n'est pas un désaccord, car les deux parties imaginent simplement à leur façon sans confronter leurs lectures. C'est l'analyse *descriptive* du GDS. Je montre comment cette analyse descriptive peut avoir des connexions intéressantes avec le phénomène de la résistance imaginative. Cependant, cette analyse du débat ne peut pas épuiser les interprétations du GDS: il y a un élément irréductiblement *normatif* qui justifie que l'on parle de désaccord.

Le GDS comme un désaccord sans faute

Dans lle CHAPITRE 11 je montre comment on peut construire une notion de *faute* qui soit suffisante pour parler de désaccord sans faute dans le cas de *La Métamorphose* de Kafka. Je montre ensuite que l'on peut soutenir plusieurs lectures non fautives de *La Métamorphose* de Kafka en produisant des contre-arguments à la lecture Nabokov. Pour Nabokov, imaginer que Gregor Samsa s'est transformé en cafard est une faute. En effet, cela va contre le principe de réalité à l'œuvre dans *La Métamorphose*: les cafards ne restent pas coincés sur le dos dans la réalité, or Gregor reste coincé sur le dos dans la fiction. Cela supporte l'idée que Gregor est, fictionnellement, un scarabée. J'explore les différentes manières de mettre en défaut cet argument, proposant ainsi des lectures non fautives de *La Métamorphose* dans lesquelles Gregor est bel et bien un cafard.

La conséquence de cette discussion est une forme de pluralisme interprétatif. Afin de caractériser précisément ce pluralisme, j'introduis la terminologie de [Friend 2017a] et défend le pluralisme de l'élucidation.

Dans le CHAPITRE 12, je me demande d'où vient ce pluralisme, ce qui le rend possible. Je défends qu'il provient de la liberté d'imaginer du lecteur, liberté d'imaginer les détails. Cependant, la liberté du lecteur est contrainte par les engagements antérieurs qu'il a pris ainsi que d'autres engagements plus généraux. En conséquence, on peut décrire la situation en utilisant le notion d'équilibre réflexif, due à Nelson Goodman dans [Goodman 1955]. Je rapproche ainsi l'indétermination de Gregor à la notion d'indétermination pragmatique de Quine ([Quine 1960]), qui est inspirée des travaux de Goodman.

Le GDS comme un désaccord fautif

Dans le CHAPITRE 13, j'en viens finalement à me poser la question des conditions de possibilité d'un débat comme le GDS. Pour ce faire, je me propose de construire une paire minimale, inspiré par la méthodologie de la linguistique formelle. Je construis cette paire minimale en mettant en regard *La Métamorphose* de Kafka et *Le Nez* de Gogol. Il n'y a pas de "grand débat du Nez" possible. Les conditions de possibilité d'un grand débat sont donc les suivantes: il faut que soit disponible soit une science naturelle soit une convention acceptée culturellement.

En conséquence, j'arrive à l'idée que le GDS est un épiphénomène d'un désaccord *réel* à propos de la disponibilité d'une science naturelle ou d'une convention. En fait, le GDS peut se réduire à la question de savoir si on peut appliquer l'entomologie *telle quelle* au monde de Gregor Samsa. Cette question est en réalité une question sur *le genre* de *La Métamorphose*, i.e. de la nouvelle bien réelle écrite par Kafka.

La distinction entre désaccord réel et fictionel, arrivé à ce degré d'analyse, semble devenir inopérante. On peut voir le GDS comme un désaccord réel et l'espace logique ouvert au début de cette partie se referme.

Un problème avec les noms fictionnels

Dans le CHAPITRE 14, je commence par définir le problème. Il s'agit de donner une théorie de la référence fictionnelle qui rende compte à la fois des usages dits "fictionnels" et des usages dits "métafictionnels". Plus précisément, il faut expliquer comme un nom fictionnel pourquoi et comment "Emma Woodhouse" semble ne pas faire référence à une personne réelle dans une phrase comme:

(16) Emma Woodhouse, belle, intelligente et riche, avec un foyer confortable et un tempérament heureux, semblait unir certaines des meilleures grâces de l'existence.

tandis que *le même nom* semble faire référence à ce que l'on peut appeler un "individu de papier" dans:

(17) Emma Woodhouse est un personnage de fiction et c'est le personnage principal du roman de mœurs de Jane Austin intitulé *Emma*.

La raison pour laquelle il est permis de penser que "Emma Woodhouse" fait bien référence dans (17) est que la phrase semble indubitablement avoir une valeur de vérité et donc, d'après un principe de compositionalité raisonnable, il faut bien que le nom réfère. La possibilité de produire des énoncés métafictionnels est clairement une caractéristique essentielle des noms fictionnels, peut-être même de tous les noms vides.

Contre le réalisme à propos du discours métafictionnel

Dans le CHAPITRE 15 je montre comment ces données linguistiques sont difficiles à théoriser à la fois pour le réaliste et l'anti-réaliste, i.e. la thèse qui dit que "Emma Woodhouse" est un nom réel et celle qui dit que "Emma Woodhouse" est un nom vide. Il semble que l'on soit forcé à être antiréaliste à propos du discours fictionnel et réaliste à propos du discours métafictionnel.

Dans le CHAPITRE 16, j'explore les contre-arguments au réaslisme. Je distingue deux types de réalisme: l'un qui est une théorie de l'ambiguité et l'autre une théorie de la polysémie. Le second réalisme est le plus convaincant, il prend son origine dans

[Kripke 1973/2013]. Je discute une version plus récente présentée notamment dans [Récanati 2018] et [Récanati forthcoming].

Contre la théorie de la polysémie des noms fictionnels, j'affirme que le type de polysémie associé à un nom fictionnel ne peut pas être de nature cognitive, car c'est un phénomène purement grammatical qui est en jeu. Or les noms fictionnels, *qua* entités grammaticales, sont soit vides, soit réels. Le passage d'une catégorie grammatical à l'autre, tel que Kripke le stipule et Récanati le construit, me paraît contradictoire. J'offre une contre-proposition pour décrire le phénomène sous-jacent. Je pense que ce qui se passe entre les énoncés fictionnels et métafictionnels est que l'on introduit un nouveau terme vide dans le langage.

L'anti-réalisme au travail

Afin de mettre en œuvre cette contre-proposition anti-réaliste, je dois reprendre les données linguistiques dans cette optique. Je commence donc au CHAPITRE 17 à classer systématiquement les énoncés métafictionnels. Il faut distinguer les énoncés métafictionnels purs des énoncés qui comportent une perspective mixe sur la fiction. Je le fais en séparant l'acte de référence de l'acte d'assertion au niveau des actes de langage, chacun ayant une version simulée, ou "pseudo". J'en arrive à une tripartition assez classique dans la litérature spécialisée: le discours interne (ou fictionnel) lorsqu'on a une pseudo-assertion contenant une pseudo-référence; le discours externe (ou métafictionnel pur) lorsqu'on a une assertion contenant de la référence; et le discours à la perspective mixe lorsqu'une assertion contient de la pseudo-référence.

Je montre qu'il y a une continuité dans les discours mixes, ce que j'explique par l'introduction de contraintes sur un pseudo-self. Puis, je pontre que si l'on peut simuler des conditions de vérité dans une perspective anti-réaliste, en jouant sur lesdites contraintes, il est impossible d'obtenir des conditions de vérité authentiques. Ainsi, le discours métafictionnel pur résiste définitivement aux techniques habituelles utilisées pas les anti-réalistes.

Parmi les énoncés métafictionnels purs, on trouve notamment les énoncés existentiels négatifs. Ces énoncés sont particulièrement importants et gênants pour l'anti-réaliste. En effet, par définition, un anti-réaliste est quelqu'un qui pense que les personnages de fiction n'existent pas. Et pourtant, il n'arrive pas à rendre les énoncés existentiels négatifs correspondant vrais! C'est la raison pour laquelle il me semble que, parmi tous les énoncés métafictionnels purs, ce sont les existentiels négatifs qui mettent le plus nettement en tension la position anti-réaliste.

Dans le CHAPITRE 19 je me propose de donner une théorie des existentiels négatif qui soit en accord avec les principes de l'antiréalisme. Pour ce faire, après avoir

présenté le problème en toute généralité, j'introduis la logique libre qui est le formalisme le plus adéquat pour donner une analyse sémantique des existentiels négatifs. Je choisis une logique libre positive, très inspirée de [Antonelli 2000]. Je montre comment cette logique libre positive peut se combiner naturellement avec la sémantique de la feinte développée en première partie. Cela permet de rendre cohérente la position anti-réaliste, et donc potentiellement attractive.

Remarques conclusives

En guise de conclusion, je reviens sur l'anti-réalisme en général. Je montre le mouvement historico-conceptuel qui a conduit de nombreux philosophes à adopter une position anti-réaliste à propos des discours fictionnels dans la seconde moitié du vingtième siècle. Cela permet à la fois de comprendre la structuration profonde de ma thèse qui est un dyptique anti-réaliste ainsi que de faire clairement apparaître le lien entre mon introduction métaphilosophique et le reste de la thèse.

Bibliography

Altshuler, Daniel and Emar Maier (2018). "Death on the Freeway: Imaginative resistance as narrator accommodation". In: *Philpapers online*.

Antonelli, G Aldo (2000). "Proto-semantics for positive free logic". In: *Journal of philosophical logic* 29.3, pp. 277–294.

Armstrong, David (1973). *Belief, Truth and Knowledge*. Cambridge University Press.

Arnauld, Antoine and Pierre Nicole (1996). *Logic or the Art of Thinking*. Cambridge University Press.

Asher, Nicholas. (2011). *Lexical Meaning in Context: A Web of Words*. Cambridge University Press.

Austen, Jane (1815). *Emma: A Novel*. Collection of British and American authors. Bernhard Tauchnitz (1877).

Bach, Kent (2008). "Applying pragmatics to epistemology". In: *Philosophical issues* 18.1, pp. 68–88.

Badura, Christopher and Francesco Berto (2019). "Truth in Fiction, Impossible Worlds, and Belief Revision". In: *Australasian Journal of Philosophy* 97.1, pp. 178–193.

Barthes, Roland (1968). "The reality effect". In: *The Novel: An Anthology of Criticism and Theory, 1900-2000*, pp. 229–34.

Barwise, Jon and John Perry (1983). *Situations and Attitudes*. MIT Press.

Beardsley, Monroe C. (1958). *Aesthetics: Problems in the Philosophy of Criticism*. 1st ed. Harcourt, Brace and World, Inc.

Bencivenga, Ermanno (1986). "Free logics". In: *Handbook of philosophical logic*. Springer, pp. 373–426.

— (2002). "Free logics". In: *Handbook of philosophical logic*. Springer, pp. 147–196.

— (2006). "Putting Language First: The 'Liberation' of Logic from Ontology". In: *A Companion to Philosophical Logic*. Blackwell Publishing, pp. 293–304.

Benoist, Jocelyn (2002). *Représentations sans objet aux origines. Aux origines de la phénoménologie et de la philosophie analytique*. Presses Universitaires de France. DOI: 10.3917/puf.benoj.2001.02.

Berto, Francesco (2017). "Impossible worlds and the logic of imagination". In: *Erkenntnis* 82.6, pp. 1277–1297.

Berto, Francesco and Mark Jago (2019). *Impossible worlds*. Oxford University Press.

Booth, Wayne (1983). *The Rhetoric of Fiction*. 2nd. University of Chicago Press.

Borges, Jorge Luis (1964). *Labyrinths, Selected Stories & Other Writings*. New Directions.

— (1967). *The Book of Imaginary Beings*. 2nd. 1974. Penguin (transl. Norman Thomas di Giovanni.

Braun, David (1993). "Empty names". In: *Noûs* 27.4, pp. 449–469.

Brentano, Franz (1874). *Psychology from an Empirical Standpoint*. Translated by Antos C.Rancurello, D.B.Terrell and Linda L.McAlister, 2nd edition 1995. Routledge.

Brooke-Rose, C. (1983). *A Rhetoric of the Unreal: Studies in Narrative and Structure, Especially of the Fantastic*. Cambridge paperback library. Cambridge University Press.

Burge, Tyler (1974). "Truth and singular terms". In: *Noûs*, pp. 309–325.

Byrne, Alex (1993). "Truth in fiction: the story continued". In: *Australasian Journal of Philosophy* 71.1, pp. 24–35.

Caillois, R. (1965). *Au coeur du fantastique*. Gallimard.

Calvino, Italo (1980). *The Literature machine*. English transl. Patrick Creagh 1986. Secker and Warburg.

— (1988). *Six Memos for the Next Millennium*. transl. Creagh, P. Charles Eliot Norton lectures. Harvard University Press.

Capinera, J.L. (2008). *Encyclopedia of Entomology*. Encyclopedia of Entomology. Springer Netherlands.

Castañeda, Hector-Neri (1975). "Individuation and non-identity: a new look". In: *American Philosophical Quarterly* 12.2, pp. 131–140.

Chomsky, N. et al. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Church, Alonzo (1965). "Review of Lambert [1963]". In: *Journal of Symbolic Logic* 30, pp. 103–104.

Cobreros, Pablo et al. (2012). "Tolerance and Mixed Consequence in the S'valuationist Setting". In: *Studia logica* 100.4, pp. 855–877.

Cocchiarella, Nino (1966). "A logic of possible and actual objects". In: *Journal of Symbolic Logic* 31, p. 688.

Coleridge, Samuel Taylor (1817). *Biographia Literaria; or, Biographical sketches of my literary life and opinions*. London : Rest Fenner.

Cooper, Robin (2007). "Copredication, dynamic generalized quantification and lexical innovation by coercion". In: *Fourth International Workshop on Generative Approaches to the Lexicon*. Paris, France, pp. 269–288.

Crimmins, Mark (1998). "Hesperus and phosphorus: Sense, pretense, and reference". In: *The Philosophical Revieal* 107.1, pp. 1–47.

Currie, Gregory (1990). *The nature of fiction*. Cambridge University Press.

Donnellan, Keith S (1975). "Speaking of nothing". In: *Contemporary Research in Philosophical Logic and Linguistic Semantics*. Springer, pp. 93–118.

Dummett, Michael (1982). "Realism". In: *Synthese* 52.1, pp. 55–112.

D'Alessandro, William (2016). "Explicitism about truth in fiction". In: *British Journal of Aesthetics* 56.1, pp. 53–65.

Eco, Umberto (1979). *Lector in fábula: le role du lecteur ou la cooperation interpretative dans les textes narratifs*. Grasset.

— (1994). *Six walks in the fictional woods*. Harvard University Press.

Erlich, Victor (1956). "Gogol and Kafka: a Note on "Realism" and "Surrealism"". In: *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*. Mouton and Co, The Hague, pp. 100–109.

Evans, Gareth (1982). *The Varieties of Reference*. ed. J. McDowell. Clarendon Paperbacks.

Everett, Anthony (2003). "Empty names and gappy propositions". In: *Philosophical Studies* 116.1, pp. 1–36.

— (2005). "Against fictional realism". In: *The Journal of Philosophy* 102.12, pp. 624–649.

— (2013). *The nonexistent*. Oxford University Press.

Fauconnier, Gilles (1994). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge University Press.

Fine, Kit (1975). "Vagueness, truth and logic". In: *Synthese* 30.3, pp. 265–300.

— (1982). "The Problem of Non-Existents. I. Internalism". In: *Topoi* 1, pp. 97–140.

— (2012). "Counterfactuals without possible worlds". In: *The Journal of Philosophy* 109.3, pp. 221–246.

Frege, Gottlob (1892). "Sense and reference". In: *The philosophical review* 57.3, pp. 209–230.

— (1980). *Philosophical and Mathematical Correspondence*. ed. Gabriel, G. and McGuinness, B. and Kaal, H. University of Chicago Press.

Friend, Stacie (2008). "Imagining fact and fiction". In: *New waves in aesthetics*. Springer, pp. 150–169.

— (2011). "The great beetle debate: a study in imagining with names". In: *Philosophical Studies* 153.2, pp. 183–211.

Friend, Stacie (2017a). "Elucidating the truth in criticism". In: *The Journal of Aesthetics and Art Criticism* 75.4, pp. 387–399.
— (2017b). "The real foundation of fictional worlds". In: *Australasian Journal of Philosophy* 95.1, pp. 29–42.
Frye, Northrop (1957). *Anatomy of criticism*. Princeton University Press (2000).
García-Carpintero, Manuel (2004). "Assertion and the semantics of force-markers". In:
Garcia Marquez, Gabriel (1967). *One Hundred Years of Solitude*. transl. Gregory Rabassa (1960). Avon Books.
García-Salgado, Tomás (2005). "The Perspective of Leonardo's Last Supper". In: 3. URL: https://perspectivegeometry.com/sitebuildercontent/sitebuilderfiles/lastsupperpg.pdf.
Gendler, Tamar Szabó (2000). "The puzzle of imaginative resistance". In: *The Journal of Philosophy* 97.2, pp. 55–81.
Gendler, Tamar Szabó and Shen-yi Liao (2016). "The problem of imaginative resistance". In: *phipapers.org*.
Genette, Gérard (1991). *Fiction et diction*. Seuil.
Gilbert, Sandra and Susan Gubar (2000). *The madwoman in the attic : the woman writer and the nineteenth-century literary imagination*. Yale University Press.
Ginzburg, Carlo (1989). "Traces. Racines d'un paradigme indiciaire". In: *Mythes, emblèmes, traces. Morphologie et histoire*, pp. 139–180.
Goodman, Nelson (1955). *Fact, Fiction, and Forecast*. 4th ed. 1983. Harvard University Press.
— (1976). *Languages of Art*. 2nd ed. Hackett Pub Co Inc.
— (1978). *Ways of Worldmaking*.
Gotham, Matthew (2016). "Composing criteria of individuation in copredication". In: *Journal of Semantics* 34.2, pp. 333–371.
Grice, H Paul (1975). "Logic and conversation". In: *1975*, pp. 41–58.
Haldane, John BS (1926). "On being the right size". In: *Harper's Magazine* 152, pp. 424–427.
Hanley, Richard (2004). "As good as it gets: Lewis on truth in fiction". In: *Australasian Journal of Philosophy* 82.1, pp. 112–128.
Hardy Jr, James D and Leonard Stanton (2002). "Magical Realism in the Tales of Nikolai Gogol". In: *Janus Head: Journal of Interdisciplinary Studies in Literature, Continental Philosophy, Phenomenological Psychology* 5, pp. 126–39.
Harrison, J.F, A. Kaiser, and J.M. VandenBrooks (2010). "Atmospheric oxygen level and the evolution of insect body size". In: *Proceedings of the Royal Society B: Bi-*

ological Sciences, 277(1690). Royal Society for the Biological Sciences, pp. 1937–1946.

Hintikka, Jaakko (1959). "Towards a theory of definite descriptions". In: *Analysis* 19.4, pp. 79–85.

— (1962a). "Cogito, ergo sum: Inference or Performance?" In: *The philosophical review* 71.1, pp. 3–32.

— (1962b). *Knowledge and belief: An introduction to the logic of the two notions*. Contemporary philosophy. Cornell Univ.

Hume, David (1757). *Of the Standard of Taste*.

Husserl, Edmund (1913). *Ideen zu einer reinen phänomenologie und phänomenologischen philosophie: buch. Allgemeine einführung in die reine phänomenologie*. Ideen zu einer reinen phänomenologie und phänomenologischen philosophie. M. Niemeyer.

Hyde, Dominic (1997). "From heaps and gaps to heaps of gluts". In: *Mind* 106.424, pp. 641–660.

Iatridou, Sabine (2000). "The grammatical ingredients of counterfactuality". In: *Linguistic inquiry* 31.2, pp. 231–270.

Jacquette, Dale (1991). *Meinongian logic: The semantics of existence and nonexistence*. Perspektiven der analytischen Philosophie. De Gruyter.

Jakobson, Roman (1929). *Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves*. Pražskỳ Linguistickỳ Kroužek: Travaux du Cercle Linguistique de Prague. Jednota československỳch matematikú a fysikú.

Kafka, Franz (1915). *The Metamorphosis*. transl. Johachim Neugroschel. Scribner – 2010 *The Metamorphosis, in The Penal Colony,and Other Stories*.

Kamp, Hans (1981). "The paradox of the heap". In: *Aspects of Philosophical Logic*. Uwe Münnich (ed.), 1996. Dordrecht: D. Reidel, pp. 225–277.

Kamp, Hans and Uwe Reyle (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy. Springer Netherlands.

Kanger, Stig (1957). *Provability in Logic*. Stockholm, Almqvist & Wiksell.

Kaplan, David et al. (1973). "Bob and carol and ted and alice". In: *Approaches to natural language* 1, pp. 490–518.

Karttunen, Lauri (1974). "Presupposition and linguistic context". In: *Theoretical linguistics* 1.1-3, pp. 181–194.

Kazin, A. (1941). *The Inmost Leaf: A Selection of Essays*. Noonday paperbacks. Noonday Press (1959).

Kim, Hanna, Markus Kneer, and Michael T Stuart (2018). "The Content-Dependence of Imaginative Resistance". In: *Advances in Experimental Philosophy of Aesthetics*, p. 143.

Kölbel, Max (2004). "III—Faultless disagreement". In: *Proceedings of the Aristotelian Society (Hardback)*. Vol. 104. Wiley Online Library, pp. 53–73.

Kripke, Saul (1959). "A completeness theorem in modal logic". In: *The journal of symbolic logic* 24.1, pp. 1–14.

— (1963). "Semantical analysis of modal logic i normal modal propositional calculi". In: *Mathematical Logic Quarterly* 9.5-6, pp. 67–96.

— (1972). *Naming and Necessity*. 1st ed. Harvard University Press.

— (1973/2013). *Reference and Existence: The John Locke Lectures*. Oxford University Press, USA.

Kroon, Frederick (2003). "Quantified negative existentials". In: *Dialectica* 57.2, pp. 149–164.

Lamarque, Peter (2008). *The Philosophy of Literature*. Foundations of the Philosophy of the Arts. Wiley.

Lamarque, Peter and S.H. Olsen (1994). *Truth, Fiction, and Literature: A Philosophical Perspective*. Clarendon library of logic and philosophy. Clarendon Press.

Lambert, Karel (1963). "Existential import revisited". In: *Notre Dame journal of formal logic* 4.4, pp. 288–292. URL: https://projecteuclid.org/download/pdf_1/euclid.ndjfl/1093957655.

— (1981). "On the philosophical foundations of free logic". In: *Inquiry* 24.2, pp. 147–203.

— (2003). "The Philosophical Foundations of Free Logic". In: *Free Logic. Selected Essays*. Cambridge University Press, pp. 122–175.

Lambert, Karel and Bas van Fraassen (1972). *Derivation and counterexample; an introduction to philosophical logic*. Encino, Calif, Dickenson Pub. Co.

Lavocat, Françoise (2010a). *La Théorie littéraire des mondes possibles*. Ouvrage collectif sous la direction de Françoise Lavocat. Editions du C.N.R.S.

— (2010b). "Les genres de la fiction. État des lieux et propositions". In: *La théorie littéraires des mondes possibles*. Éditions du C.N.R.S., pp. 15–51.

— (2016). *Fait et fiction: Pour une frontière*. Poétique. Le Seuil.

Leblanc, Hugues and Richmond Thomason (1968). "Completeness theorems for some presupposition-free logics". In: *Fundamenta Mathematicae* 62, pp. 125–164.

Leonard, Henry (1956). "The logic of existence". In: *Philosophical studies* 7.4, pp. 49–64.

Lévi-Strauss, Claude (1949). *Les Structure Elementaries de la Parenté*. Beacon paperback no. 340 (English ed. 1969). Eyre & Spottiswoode.

Lewis, David (1973). *Counterfactuals*. 2nd. John Wiley & Sons (2013).
— (1978). "Truth in fiction". In: *American philosophical quarterly* 15.1, pp. 37–46.
— (1979a). "Counterfactual dependence and time's arrow". In: *Noûs*, pp. 455–476.
— (1979b). "Scorekeeping in a language game". In: *Semantics from different points of view*. Springer, pp. 172–187.
— (1983). "Truth in fiction with postscripts". In: *Philosophical papers, vol. 1*. New York: Oxford University Press, pp. 261–280.
— (1984). "Putnam's paradox". In: *Australasian journal of philosophy* 62.3, pp. 221–236.
— (1986). *On the plurality of worlds*. Vol. 322. Oxford Blackwell.
— (1990). "Noneism or allism?" In: *Mind* 99.393, pp. 23–31.
— (1999). "Individuation by acquaintance and by stipulation". In: *Papers in Metaphysics and Epistemology*. Cambridge University Press, pp. 373–402.
— (2015). "Reply to Dana Scott,"Is There Life on Possible Worlds?"" In: *A Companion to David Lewis*, pp. 18–21.
Lorand, Ruth (2001). "Telling a story or telling a world?" In: *The British Journal of Aesthetics* 41.4, pp. 425–443.
Luo, Zhaohui (2012). "Formal semantics in modern type theories with coercive subtyping". In: *Linguistics and Philosophy* 35.6, pp. 491–513.
MacColl, Hugh (1905). "Symbolic Reasoning (VI.)" In: *Mind* 14.53, pp. 74–81.
Maier, Emar (2018). "Lying and Fiction". In: *The Oxford Handbook of Lying*. Ed. by Jörg Meibauer. Oxford University Press, Part IV, §23.
Mally, Ernst (1914). "Gegenstandstheoretische Grundlagen der Logik und Logistik". In:
Malrieu, J. and B. Vercier (1992). *Le fantastique*. Hachette Éducation (réédition numérique FeniXX).
Matravers, Derek (2014). *Fiction and Narrative*. 1st ed. Oxford University Press.
McDowell, John (1977). "On the sense and reference of a proper name". In: *Mind* 86.342, pp. 159–185.
Meinong, Alexius (1904). "Uber Gegenstandtheorie". In: *Untersuchungen zur Gegenstandtheorie und Psychologie (edit. A. Meinong), Barth, Leipzig*.
Mellier, D. (2000). *La littérature fantastique*. Mémo (Editions du Seuil) Lettres. Seuil.
Merleau-Ponty, Maurice (1945). *Phénoménologie de la perception*. Bibliothèque des idées. Gallimard.
Meyer, Robert K, Ermanno Bencivenga, and Karel Lambert (1982). "The ineliminability of E! in free quantification theory without identity". In: *Journal of Philosophical Logic* 11.2, pp. 229–231.

Millikan, Ruth Garrett (1998). "A common structure for concepts of individuals, stuffs, and real kinds: More Mama, more milk, and more mouse". In: *Behavioral and Brain Sciences* 21.1, pp. 55–65.

Montague, Richard (1970). "English as a Formal Language". In: *Linguaggi nella societa e nella tecnica*. Ed. by Bruno Visentini. Edizioni di Communita, pp. 188–221.

Moran, Richard (1994). "The expression of feeling in imagination". In: *The Philosophical Review* 103.1, pp. 75–106.

Nabokov, Vladimir (1955). *Lolita*.

— (1959). *Nikolai Gogol*. New Directions.

— (1980). *Lectures on Literature*. 1st ed. by Fredson Bowers and John Updike. Mariner Books.

Nietzsche, Friedrich (1908). *Ecce homo: Wie man wird, was man ist*.

Nolan, Daniel (1997). "Impossible worlds: A modest approach". In: *Notre Dame Journal of Formal Logic* 38.4, pp. 535–572.

— (2013). "Impossible worlds". In: *Philosophy Compass* 8.4, pp. 360–372.

Nolt, John (2007). "Free logics". In: *Philosophy of logic*. Elsevier, pp. 1023–1060.

Parsons, Terence (1980). *Nonexistent objects*. New Haven: Yale University Press.

Paśniczek, Jacek (2001). "Can Meinongian logic be free?" In: *New Essays in Free Logic*. Springer, pp. 227–236.

Pavel, Thomas (1975). ""Possible Worlds" in Literary Semantics". In: *The Journal of Aesthetics and Art Criticism* 34.2, pp. 165–176. ISSN: 00218529, 15406245.

— (1986). *Fictional Worlds*. Harvard University Press.

Peacocke, Christopher (1978). "Necessity and truth theories". In: *Journal of Philosophical Logic* 7.1, pp. 473–500.

Phillips, John (1999). "Truth and inference in fiction". In: *Philosophical Studies* 94.3, pp. 273–293.

Plato. *Theaetetus and Sophist*. Edition and translation by Christopher Rowe (2015). Cambridge Texts in the History of Philosophy. Cambridge University Press.

Priest, Graham (1997). "Sylvan's Box: A Short Story and Ten Morals". In: *Notre Dame Journal of Formal Logic* 38.4, pp. 573–582.

— (2008a). *An introduction to non-classical logic: From if to is*. Cambridge University Press.

— (2008b). "The closing of the mind: How the particular quantifier became existentially loaded behind our backs". In: *The Review of Symbolic Logic* 1.1, pp. 42–55.

Proudfoot, Diane (2006). "Possible worlds semantics and fiction". In: *Journal of Philosophical Logic* 35.1, pp. 9–40.

Proust, Marcel (1906). “Sur la lecture”. In: *Sésame et les lys*. Préface à la traduction du livre de John Ruskin. Société du Meercure de France.

Pustejovsky, James (1998). *The Generative Lexicon*. Revised edition. Bradford Books. MIT Press.

Putnam, Hilary (1971). *Philosophy of Logic*. Harper essays in philosophy. Harper & Row.

Quine, Willard Von Orman (1939). “Designation and existence”. In: *The Journal of Philosophy* 36.26, pp. 701–709.

— (1948). “On what there is”. In: *The Review of Metaphysics* 2.1, pp. 21–38.

— (1951). “Main trends in recent philosophy: Two dogmas of empiricism”. In: *The philosophical review*, pp. 20–43.

— (1960). *Word and Object (Studies in Communication)*. 2nd. The MIT Press (1964).

— (1969). *Ontological relativity: and other essays*. First Edition. Columbia University Press.

— (1986). *Philosophy of logic*. 2nd ed. Harvard University Press.

Rapaport, William J (1978). “Meinongian theories and a Russellian paradox”. In: *Noûs*, pp. 153–180.

Reboul, Anne (2011). “La résistance imaginative: émotions, valeurs et fiction”. In: *Les ombres de l’âme: penser les émotions négatives*. Ed. by Tappolet et al. Markus Halter, pp 131–143. URL: https://hal.archives-ouvertes.fr/hal-00925833.

Récanati, François (2012). *Mental Files*. Oxford University Press.

— (2018). “II—Fictional, Metafictional, Parafictional”. In: *Proceedings of the Aristotelian Society*. Vol. 118. 1. Oxford University Press, pp. 25–54.

— (forthcoming). “Fictional reference as simulation”. In: *The Language of Fiction*. Ed. by Emar Maier & Andreas Stokke. Oxford University Press.

Renauld, Marion (2013). “Vous avez dit ’fiction’? Analyse contextualiste du jugement de fictionalité et approche pragmatiste des œuvres romanesques”. PhD thesis. Université de Lorraine.

— (2014). *Philosophie de la fiction. Une approche pragmatiste du roman*. Presses Universitaires de Rennes.

Rescher, Nicholas (1960). “Many-sorted quantification”. In: *Atti del XII Congresso Internazionale di Filosofia*. Vol. 5, pp. 447–453.

Richard, Mark (2000). “Semantic pretence”. In: *Empty names, Fiction, and the Puzzles of Non-Existence*. Ed. by Anthony Everett and Thomas Hofweber. CSLI Press, Stanford, pp. 205–232.

Ripley, David (2013). “Sorting out the sorites”. In: *Paraconsistency: Logic and applications*. Springer, pp. 329–348.

Rivenc, François (1995). “Husserl avec et contre Frege”. In: *Les Études philosophiques* 1, p. 13.
Rorty, Richard (1981). “Is there a problem about fictional discourse?” In: *Funktionen des Fictiven: Poetik und Hermeneutik,* 10.
— (1982). *Consequences of Pragmatism: Essays, 1972-1980*. Univ of Minnesota Pr.
Rosch, Eleanor and Carolyn B Mervis (1975). “Family resemblances: Studies in the internal structure of categories”. In: *Cognitive psychology* 7.4, pp. 573–605.
Rosenblatt, Louise (1938). *Literature as exploration*. New York: Appleton-Century.
Rouillé, Louis (2015). *Derrière les noms dans la fiction*. CogMaster thesis, dir. François Récanati.
Routley, Richard (1979). “The semantical structure of fictional discourse”. In: *Poetics* 8.1-2, pp. 3–30.
— (1980). *Exploring Meinong’s Jungle and Beyond*. Departmental monograph Philosophy Dept., Research School of Social Sciences. Australian National University.
Russell, Bertrand (1905). “On denoting”. In: *Mind* 14.56, pp. 479–493.
— (1915). “The ultimate constituents of matter”. In: *The Monist* 25.3, pp. 399–417.
— (1918). *The Philosophy of Logical Atomism*. Routledge (1972).
— (1919). *Introduction to Mathematical Philosophy*. London: Allen & Unwin.
— (1923). “Vagueness”. In: *The Australasian Journal of Psychology and Philosophy* 1.2, pp. 84–92.
Ryan, Marie-Laure (1991). *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Indiana University Press.
Ryle, Gilbert (1949). *The Concept of Mind*. 60 Anv (2009). University of Chicago Press. URL: ./BIBLI/Ryle_1949-The_Concept_of_Mind.pdf.
Sainsbury, Mark (2009). *Fiction and Fictionalism*. 1st ed. New Problems of Philosophy.
— (2018). *Thinking about Things*. Oxford University Press.
Salmon, Nathan (1987). “Existence”. In: *Philosophical perspectives* 1, pp. 49–108.
— (1998). “Nonexistence”. In: *Noûs* 32.3, pp. 277–319.
Sartre, J.P. (1943). *L’être et le néant: essai d’ontologie phénoménologique*. Bibliothèque des idées. Gallimard.
Schaeffer, Jean-Marie (1999). *Pourquoi la fiction?* Poétique. Seuil.
Schiffer, Stephen (2003). *The Things We Mean*. Oxford University Press.
Schock, Rolf (1968). *Logics without existence assumptions*. Almqvist & Wiksell Stockholm.
Schuppert, Guillaume (2019). “Le "faire-semblant" en procès. Examen et défense de la philosophie de Kendall L. Walton”. PhD thesis. Université de Lorraine.

Schwartz, Lillian F (1988). "The staging of Leonardo's Last Supper: a computer-based exploration of its perspective". In: *Leonardo* 21.5, pp. 89–96.
Scott, Dana (1979). "Identity and existence in intuitionistic logic". In: *Applications of sheaves*. Springer, pp. 660–696.
Searle, John (1969). *Speech Acts: An Essay in the Philosophy of Language*.
— (1975). "The logical status of fictional discourse". In: *New literary history* 6.2, pp. 319–332.
Shukman, Ann (1989). "Gogol's The Nose or the Devil in the Works". In: *Nikolay Gogol*. Springer, pp. 64–82.
Sicher, Efraim (1990). "Dialogization and Laughter in the Dark, or How Gogol's Nose Was Made: Parody and Literary Evolution in Bachtin's Theory of the Novel". In: *Russian Literature* 28.2, pp. 211–233.
Sorensen, Roy (2007). "Bald-faced lies! Lying without the intent to deceive". In: *Pacific Philosophical Quarterly* 88.2, pp. 251–264.
Soria Ruiz, Andrés (2019). "The Place of Value in Natural Language. A non-factualist approach to evaluative terms and judgment." PhD thesis. IJN, ENS.
Stalnaker, Robert (1968). "A theory of conditionals". In: *Ifs*. Springer, pp. 41–55.
— (1977). "Pragmatic presuppositions". In: *Proceedings of the Texas conference on performatives, presuppositions, and implicatures. Arlington, VA: Center for Applied Linguistics*. ERIC, pp. 135–148.
— (1980). "A defense of conditional excluded middle". In: *Ifs*, pp. 87–104.
Stock, Kathleen (2017). *Only Imagine: Fiction, Interpretation and Imagination*. Oxford University Press.
Stojanovic, Isidora (2017). "Context and Disagreement". In: *Cadernos de Estudos Linguísticos*. URL: https://hal.archives-ouvertes.fr/hal-01625608.
Strawson, Peter (1950). "On referring". In: *Mind* 59.235, pp. 320–344.
— (1967). "Is existence never a predicate?" In: *Crítica: Revista Hispanoamericana de Filosofía*, pp. 5–19.
Suits, Bernard (1978). *The Grasshopper: Games, Life and Utopia*. University of Toronto Press.
Terrone, Enrico (2017). "On Fictional Characters as Types". In: *British Journal of Aesthetics* 57.2, pp. 161–176.
Thomasson, Amie (1999). *Fiction and Metaphysics*. Cambridge Studies in Philosophy. Cambridge University Press.
— (2003a). "Fictional characters and literary practices". In: *The British Journal of Aesthetics* 43.2, pp. 138–157.
— (2003b). "Speaking of fictional characters". In: *Dialectica* 57.2, pp. 205–223.
Todorov, Tzvetan (1970). *Introduction à la littérature fantastique*. Éditions du Seuil.

Twardowski, Kazimierz (1912). "Actions and Products. Comments on the Border Area of Psychology, Grammar, and Logic". In: *Semiotics in Poland 1984–1969*. Springer (1979), pp. 13–27.

Uckelman, Sara L (2019). "Fanfiction, Canon, and Possible Worlds". In:

Unger, Peter (1980). "The problem of the many". In: *Midwest studies in philosophy* 5.1, pp. 411–468.

Univalent Foundations Program, The (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study: https://homotopytypetheory.org/book.

Van Fraassen, Bas (1966a). "Singular terms, truth-value gaps, and free logic". In: *The journal of Philosophy* 63.17, pp. 481–495.

— (1966b). "The completeness of free logic". In: *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 12, pp. 219–234.

Van Inwagen, Peter (1977). "Creatures of fiction". In: *American philosophical quarterly* 14.4, pp. 299–308.

Varzi, Achille (2001). "Vagueness in geography". In: *Philosophy & Geography* 4.1, pp. 49–65.

Vuillaume, Marcel (1990). *Grammaire temporelle des récits*. Propositions. Les éditions de minuit.

Walton, Kendall (1978). "How remote are fictional worlds from the real world?" In: *The Journal of Aesthetics and Art Criticism* 37.1, pp. 11–23.

— (1990). *Mimesis as Make-believe: On the Foundations of the Representational Arts*. Harvard University Press.

— (1993). "Metaphor and Prop Oriented Make-Believe". In: *European Journal of Philosophy* 1.1, pp. 39–57.

— (1994). "Morals in fiction and fictional morality". In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 68, pp. 27–66.

— (2006). "On the (so-called) puzzle of imaginative resistance". In: *The architecture of the imagination: New essays on pretence, possibility, and fiction*. Clarendon Press Oxford, pp. 137–148.

— (2013). "Fictionality and imagination reconsidered". In: *From Fictionalism to Realism*. edited by C. Barbero, M. Ferraris, and A. Voltolini. Cambridge Scholars Publishing.

Wang, Hao (1952). "Logic of many-sorted theories". In: *The Journal of Symbolic Logic* 17.2, pp. 105–116.

Weatherson, Brian (2004). "Morality, fiction, and possibility". In: *Philosophers' Imprint* 4.3, pp. 1–27.

Whitehead, Alfred North and Bertrand Russell (1912). *Principia mathematica*. Vol. 2. University Press.

Whitehead, Claire (2017). *The Fantastic in France and Russia in the 19th Century: In Pursuit of Hesitation*. Routledge.

Williamson, Timothy (1990). "Necessary identity and necessary existence". In: *Wittgenstein—Eine Neubewertung/Wittgenstein—Towards a Re-Evaluation*. Springer, pp. 168–175.

— (2002). *Knowledge and Its Limits*. Oxford University Press, USA.

— (2013). *Modal Logic as Metaphysics*. Oxford University Press.

— (2014). "How did we get here from there? The transformation of analytic philosophy". In: *Belgrade Philosophical Annual* 27, pp. 7–37.

Wittgenstein, Ludwig (1921). *Tractatus Logico-Philosophicus*. 2nd ed. Routledge Classics. Routledge (2001).

— (1953). *Philosophical Investigations*. 3rd - translation Elizabeth Anscombe. Blackwell Publishers (1986).

Woods, John (1974). *The Logic of Fiction: A Philosophical Sounding of Deviant Logic*. 1st. The Hague, Paris: Mouton, and Co.

— (2018). *Truth in Fiction: Rethinking its Logic*. 1st ed. Synthese Library 391. Springer International Publishing.

Woodward, Richard (2011). "Truth in fiction". In: *Philosophy Compass* 6.3, pp. 158–167.

Yablo, Stephen (2001). "Go figure: A path through fictionalism". In: *Midwest studies in philosophy* 25.1, pp. 72–102.

— (2002). "Coulda, woulda, shoulda". In: *Conceivability and possibility*, pp. 441–492.

Zalta, Edward (1983). *Abstract Objects: An Introduction to Axiomatic Metaphysics*. 1st ed. Synthese Library 160. Springer Netherlands.

Zwicky, Arnold M and Jerrold M Sadock (1973). *Ambiguity tests and how to fail them*. Ohio State University. Department of Linguistics.

RÉSUMÉ

Dans cette thèse, je contribue aux débats contemporains en philosophie analytique sur la vérité, l'interprétation et la référence dans la fiction. Je défends une version du "fonctionnalisme" (issu de l'œuvre de Kendall Walton) selon lequel le concept clé pour analyser les fictions est la feinte ou le faire-semblant. Dans la première partie, je m'oppose à la théorie modale de la vérité dans la fiction et j'introduis ensuite une "sémantique de la feinte". La théorie modale dit que les énoncés fictionnels sont comme les énoncés contrefactuels, qui peuvent se voir attribuer des conditions de vérité dans la sémantique des mondes possibles. Mais le fictif déborde le possible, et même l'impossible (de l'hypothétique sémantique des mondes impossibles). De plus, la théorie modale est incompatible avec une théorie causale de l'information sémantique, qui a des bons arguments pour elle. Quant à la sémantique de la feinte, il s'agit d'un appareil formel donnant des conditions de fictionnalité (au lieu de conditions de vérité) provenant des caractéristiques réelles des accessoires utilisés dans les jeux de faire-semblant et des "principes de génération". Dans la deuxième partie, j'affine la sémantique de la feinte à partir d'une étude de cas. C'est un débat littéraire à propos ce qui est vrai dans la nouvelle de Kafka, "la Métamorphose". Nabokov a un jour donné un argument contre les critiques qui affirment que Gregor Samsa s'est transformé en un monstrueux cafard. Les cafards ne restent pas coincés sur le dos ; or, Gregor est coincé sur le dos dans la première scène. Nabokov affirme donc qu'il ne peut pas être un cafard ; ça doit être un gros scarabée. Pour analyser ce "grand débat du scarabée", j'utilise la notion de "désaccord sans faute" qui vient de l'épistémologie. J'étudie finalement l'indétermination des événements fictionnels qui, à mon sens, n'est ni linguistique ni ontologique, mais pragmatique. Dans la troisième partie, je défends l'antiréalisme des noms fictifs selon lequel les personnages de fiction n'existent pas et les noms fictifs ne font pas référence. L'antiréalisme est le pendant du fonctionnalisme sur la question de la référence. Bien qu'intuitive, cette théorie doit répondre à un contre-argument puissant basé sur des "emplois métafictionnels" des noms. Par exemple, on peut dire : "Emma Woodhouse est un personnage de fiction". Étant donné un principe de compositionalité, il s'ensuit que le nom "Emma Woodhouse" fait référence dans de tels contextes. Cet argument conduit à une forme de réalisme : les noms fictifs font référence à une sorte "d'artefact abstrait". Je montre que les meilleures théories réalistes sont inadéquates. Je fournis ensuite une analyse des données linguistiques en introduisant une notion de perspective. Cela me permet de circonscrire les énoncés métafictionnels problématiques pour l'antiréaliste. Ironie de l'histoire, les anti-réalistes buttent principalement sur des existentiels négatifs bien que ceux-ci affirment précisément l'idée centrale de l'antiréalisme, à savoir que les personnages de fiction n'existent pas. La sémantique de la feinte (qui donne des conditions de fictionnalité) est impuissante à leur donner des conditions de vérité. Pour les expliquer, j'utilise une logique libre positive qui se combine avec la sémantique de la feinte. L'antiréalisme est donc à la fois intuitif et tenable.

MOTS CLÉS

Fiction, vérité dans la fiction, noms fictifs, sémantique de la feinte, philosophie de la littérature, désaccord sans faute, emplois métafictionnel des noms, existentiels négatifs, logique libre

ABSTRACT

In this dissertation, I contribute to contemporary debates in analytic philosophy about truth, interpretation and reference in fiction. I defend a version of "functionalism" (originating in Kendall Walton's work) which says that the key concept for analyzing fictions is pretence or make-believe. In the first part, I argue against the modal account of truth in fiction and then introduce "pretence semantics". The modal account says that fictional statements are similar to counterfactual statements, which can be given truth-conditions using possible-world semantics. But the fictional well exceeds the possible, and also the impossible (of hypothetical impossible-world semantics). Moreover, the modal account is incompatible with a causal theories of semantic information which can be argued for independently. As for pretence semantics: it is a formal apparatus delivering fictionality-conditions (instead of truth-conditions) which derive from real features of the props used in games of make-believe and some "principles of generation". In the second part, I fine-tune pretence semantics on a case study. It is a literary debate about what is true in Kafka's story "the Metamorphosis". Nabokov once argued against critics who say that Gregor Samsa has turned into a monstrous cockroach. Cockroaches do not get stuck on their backs; Gregor is stuck on his back in the opening scene of the story. So, Nabokov argues, he cannot be a cockroach; he must be a big beetle. In order to analyze this "great beetle debate", I use the notion of "faultless disagreement" which comes from epistemology. I thus investigate the indeterminacy of fictional events which, I argue, is neither linguistic nor ontological but pragmatic in nature. In the third part, I defend anti-realism about fictional names which says that fictional characters do not exist and that fictional names do not refer. Anti-realism is the same doctrine as functionalism applied to reference. Though intuitive, the view has to meet a powerful counterargument based on "metafictional uses" of names. For instance, one can say truly: "Emma Woodhouse is a fictional character". Given compositionality, it follows that the name "Emma Woodhouse" refers in such contexts. This argument leads to a form of realism: fictional names refer to some kind of "abstract artefact". I show that the best realist theories are inadequate. Then I provide an analysis of the linguistic data introducing a notion of perspective. It enables me to circumscribe the problematic metafictional statements for the anti-realist. Ironically, anti-realists mainly struggle with negative existentials, although these put into words the central tenet of anti-realism, namely that fictional characters do not exist. Pretence semantics (which yields fictionality-conditions) is helpless for giving truth-conditions to them. To account for them, I use a version of positive free logic which combines with pretence semantics. Anti-realism is thus both intuitive and tenable.

KEYWORDS

Fiction, truth in fiction, fictional names, pretence semantics, philosophy of literature, faultless disagreement, metafictional uses of names, negative existentials, free logic