Empty Names

Michael George Trevor Bench-Capon

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and the History of Science

September 2013

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

A lot of people have helped me, philosophically or otherwise, while I was writing and this. I would like to thank these people:

- The other philosophy postgraduates at Leeds from 2009 to 2013, especially Thomas Brouwer, Wouter Kalf, and Paul Ramshaw.
- The staff in the philosophy department at Leeds during that time, especially my supervisor, Andrew McGonigal.
- My examiners, Robbie Williams and Cian Dorr, for their input at the viva.
- My parents, my brothers, and my grandparents.
- My girlfriend, Aliya Vasylenko.

Abstract

Empty names are names which do not refer to anything. Apparently empty names are used in many different ways, and an analysis which looks good for one kind of use can look bad for another. I aim to get a wide enough angle on the issues that the solutions I propose won't run into that problem.

Chapter one is about names which are empty because they were introduced in the context of mistakes and lies. I see how we can assign truth values to utterances containing such names. I also look at how genuinely empty names could be meaningful at all, and examine how they could fit into a Davidsonian theory of meaning.

Chapter two is about mental states corresponding to the names dealt with in chapter one. I try to give an account of how beliefs could be subject to rational norms without appealing to their propositional contents. I do this by showing that puzzles about co-referring names can motivate such an account independently, and that the empty name beliefs can fit into this framework easily.

Chapter three is about attitude ascriptions and propositions. I consider different ways of responding to the problem of having propositions but no objects for the propositions to be about. I defend an account involving gappy Fregean propositions, and give a semantics for attitude ascriptions which incorporates them.

Chapter four is about the names that occur in fiction. I argue that we should take these to be polysemous between a use referring to an artistic creation and a use primarily suited to pretence. For the first use, I survey proposals for ontologies of fictional characters, and suggest one of my own. To make sense of the second use, I use a two-

dimensional semantics, which also helps with the problem of negative existentials.

Contents

Chapter Three: Contents without Constituents 103

Chapter Four: Fictional Names 147

Chapter 0 – Introduction

Once upon a time, at least according to philosophical folklore, most people were descriptivists about names. This means they thought that names, like ‘Aristotle’, had the same meanings as definite descriptions, like ‘the pupil of Plato and teacher of Alexander’. Bertrand Russell [1905] gave a famous analysis of definite descriptions which still makes sense even if nothing fits the description, and thus we could be forgiven for thinking that we had solved any problems associated with names which don’t refer to anything, or *empty* names as we'll call them. Apparently empty names are words like ‘Apollo’ (for a god some Greeks thought existed), ‘Vulcan’ (for a planet some astronomers thought existed), and ‘Santa’ (for a jolly fellow some children think exists).

The problem of empty names came back, so the story goes, when people like Ruth Barcan Marcus, Saul Kripke and David Kaplan started challenging the descriptivist consensus. They presented arguments that at least some of the time names behave very differently from definite descriptions. One view which gained some popularity was *Millianism*, named after John Stuart Mill, which says that the meaning of a name just is its referent. This immediately presents a problem for empty names, because it seems that a name without a referent won’t have a meaning. Even if we stop short of Millianism, we might still want to say that a name’s meaning is more intimately connected to its referent than a definite description’s meaning is to the thing (if any) fitting the description, and this makes empty names a problem.

One response is to fall back on descriptivism. I don’t want to do that, and will mostly only consider it as either a last resort or way of treating special cases. Another response is Meinongianism, which in its extreme form says there are no empty names, and all apparently empty names refer to non-existent objects. I don’t want to do that either, but I will leave the view on the table and give it a fair ride. The principal aim of this thesis is to see if we can make sense of the different ways we use apparently empty names without falling back on descriptivism. I will conclude that we can, a present my preferred way of doing so.

Chapter one looks at the consequences of there being meaningful names which are genuinely empty. It begins by looking at a proposal of David Braun's that utterances involving genuinely empty names express what he calls *gappy propositions*, which are entities structured like ordinary propositions but with gaps in one or more of the places where an ordinary proposition would have a constituent. Braun's theory has us evaluate the truth values of these utterances, and the corresponding beliefs, according to a two-valued negative free logic. I also look at some other principled ways of assigning truth values to utterances containing genuinely empty names, and defend the view that for at least some classes of apparently empty names, we can follow Braun in assigning the utterances what I call *pessimistic* truth values. The main defining feature of pessimistic evaluations is that atomic sentences in extensional contexts cannot be true if they contain any empty names. I examine an objection due to Anthony Everett that Braun's view does have enough to say about what 'Santa doesn't exist' has in common with 'Father Christmas does not exist', but not with 'Hamlet does not exist'. I respond that we should see the difference not in what the sentences represent, but in how they try to represent, and link this to work by David Chalmers, Gillian Russell and Sam Cumming on what can be called *metarepresentational* content. After looking at the consequences of meaningful names being apparently empty, chapter one finishes by looking at how we can make sense of empty names being meaningful within a Davidsonian theory of meaning. I argue that empty names as I have treated them present a distinctive puzzle for the Davidsonian project, but that this puzzle can be solved in more than one way, and that this gives us a concrete proposal for how to understand empty names as meaningful and learnable.

While chapter one is about language, chapter two is about thought. Communicative practices involving genuinely empty names will give rise to systems of beliefs which suffer from intentional failure in the way the names suffer from referential failure. We can carry the pessimistic truth assignments over from the utterances to the beliefs, but a new problem arises: how to make sense of the notions of consistency and valid deduction when the beliefs have either no contents or gappy contents. Especially if the beliefs have no propositional contents, it will be problematic to account for rational relations between the beliefs by invoking logical relations between the contents. Even if they have gappy contents, the fact that atomic gappy propositions cannot be true makes it hard for there to be non-trivial logical relations between them, which could used to explain rational norms governing the beliefs.

I argue that consideration of beliefs corresponding to co-referring names can motivate a framework which draws on work by Kit Fine to account for rational relations without reference to the beliefs' actual propositional contents. Instead, we cash them out in terms of the beliefs' potential contents or objects, as constrained by what Fine calls *co-ordination* relations between the beliefs. Beliefs suffering from intentional failure can fit into this framework without much trouble.

Chapter three is about attitude ascriptions involving apparently empty names, like 'Smith believes Vulcan is a planet'. These create a problem because ordinarily it is useful to take attitude ascriptions to involve reference to propositions, but where empty names are involved that will seem to mean referring to a proposition with some constituents missing. We can respond in any of three ways: reify the constituents, reject the propositions, or try to defend an ontology of propositions with constituents missing. David Braun's theory from chapter one is a version of the third option, but I argue that its Russellian gappy propositions are not fine grained enough, and that we would do best to look into the possibilities for an ontology of Fregean gappy propositions. I give a semantics for attitude ascriptions embodying this proposal, and show that it deals well with the problems we are trying to solve. This semantics is also presented more formally in an appendix.

The treatments of the first three chapters, with their pessimistic truth-value assignments in non-intentional contexts, are most plausible for names introduced in the context of a mistake or a lie. Sometimes, however, apparently names are introduced without anyone either being deceived or attempting to deceive, and the pessimistic truth value assignments seem less appropriate. Chapter three looks at fictional names, which include uses of names like ‘Sherlock Holmes’ and ‘Hercule Poirot’ as they appear within fiction, and also uses of the same names to talk about fiction as literary critics do. I argue that we should take the names as used to talk about fiction as referring to members of an ontology of fictional characters. I examine and evaluate some existing proposals for such an ontology, and sketch a proposal of my own which can solve the problems I identify. I argue for a separate analysis of uses of fictional names within fiction, and some derivative uses, which treats the fictional names as what I call *semi-rigid designators*. This analysis gives us a treatment of negative existential statements like ‘Holmes does not exist’ which takes them at face value and evaluates them as true. I also show how these treatments of fictional names fit in with my treatment of attitude ascriptions.

Chapter 1 – Names without Referents

1.0 Introduction

To begin with, we can set the problem up as a tension between these four theses:

1. When a name makes a semantic contribution to a sentence in which it occurs, that contribution includes a referent.
2. Some meaningful names do not have an existent referent.
3. Everything exists.
4. A meaningful name must make a semantic contribution to the sentences in which it occurs.

These four theses are more or less contradictory. If (2) and (3) are true, there must be meaningful names without referents. If (4) is true, then this means there are names which make semantic contributions to sentences containing them but do not have referents. This contradicts (1). Now, there are a lot of different kinds of uses of apparently empty names, and while all of these uses give rise to the tension between these four theses, you don't have to deny the same one each time.

(1) is denied by descriptivism about names, for example that defended by Russell [1905] and followed by Quine [1948: 5-9]. If names are all just disguised definite descriptions, they have the same semantic content whatever satisfies the description, and whether the description is satisfied uniquely or not. Empty names are as such not a big problem for descriptivism. I will not be discussing descriptivism much though[1],

[1] Exceptions will be §2.11, where descriptivism is used as an illustrative example, and §3.22, where descriptivism is considered (and rejected) as a way

because a large part of this project is to see if non-descriptivist theories can be made to work in the face of the problems of empty names. There are other ways of denying (1) besides descriptivism, and later on I will suggest that denying it is appropriate in at least some non-extensional contexts. (2) can be plausibly denied in some cases too, and in chapter four I will argue that fictional names as used in discourse about (rather than within) fiction can be treated that way. It is difficult to deny (2) across the board though, if for no other reason then because of the problem of negative existentials: ‘n does not exist’ would never be both meaningful and false. Denying (3) is Meinongianism: the view that there are non-existent objects, as defended notoriously by Meinong [1960] and more recently by Terence Parsons [1980] and Ed Zalta [1983]. I’m going to consider Meinongianism seriously in chapter three and particularly chapter four, but as it happens I will always end up supporting non-Meinongian solutions.

This chapter is mostly about what happens when we deny (4), and say that sometimes names can be meaningful without making any semantic contribution at all. It might seem hard to make sense of this: isn’t making a semantic contribution just what it is to be meaningful? Nonetheless, it seems that sometimes speakers could be in a situation where they blithely talk to each other, and appear to understand each other, even though the mechanisms for names making a semantic contribution, whatever they are, have broken down. Alternatively, maybe only the referential mechanisms have broken down: the names make a semantic contribution but no referential contribution, even though they are expressions for referring (rather than e.g. quantificational expressions). This can be seen as a kind of non-descriptivist denial of (1).

of analysing names as used in fiction, while keeping a non-descriptivist analysis of most names.

These kinds of case are the worst sort of empty name use, from the point of view of the speaker. Sometimes apparently empty names are introduced as part of fiction or make-believe, and while they may not have referents, nobody has to be making a mistake. Sometimes, on the other hand, names are introduced either because someone is making a mistake, or to induce a mistake in others. An example of the first kind is the name 'Vulcan', which (so the story goes) was introduced by Urbain Leverrier to name a planet he believed was closer to the sun the Mercury, and whose gravitational pull he thought caused some otherwise unexplained irregularities in Mercury's orbit. But there isn't a planet there for 'Vulcan' to refer to. An example of the second kind could be 'Santa', which adults use when telling tall tales to children around Christmastime. This time the introducers of the name aren't making a mistake, but the children who use it are as much in the dark as Leverrier. We can call names which really have no referents, instead of just appearing to have no referent, *genuinely empty names*, to contrast with *apparently empty names*, which would also include names we end up understanding as referring to a non-existent object, an abstract fictional character or something of that kind.

§1.1 approaches the problem of meaningful names without semantic contributions via a set of problems identified by David Braun [2005]. We can approach the problem via Braun because he proposes a solution which is fairly describable as taking empty names as being meaningful while making no semantic contribution. First we look at his problems, and then we look at his solutions, which involve a kind of propositional content he calls *gappy propositions*. Next we see if there are other ways of justifying assigning the same truth values to utterances containing genuinely empty names that Braun assigns them. If we can't, we look at what truth values we should assign them instead. I will end up arguing, with Braun, that we should assign truth values according to a two value negative free logic, at least in the non-intentional contexts under discussion. Intentional contexts are left for chapter three.

§1.2 looks at an objection to Braun's view due to Anthony Everett [2003]. This objection is that Braun can't explain what the contents of 'Santa doesn't exist' and 'Father Christmas doesn't exist' have in common with each other, but not with 'Hamlet doesn't exist'. This problem arises for Braun because he thinks they all express the same GP. It is also a threat to any proposal denying that genuinely empty names make a semantic contribution, because if two names don't make contributions at all, they don't make different contributions. We consider three possible solutions: a Millian solution similar to the one Braun adopts, a metalinguistic solution based on an idea from Keith Donnellan [1974], and a Fregean solution which only denies that genuinely empty names make a referential contribution. I also suggest how the Fregean and metalinguistic solutions can be combined, drawing on some work by Sam Cumming [2007, 2008], Gillian Russell [2008] and David Chalmers [2011, forthcoming]. This allows that empty names can have a kind of *metarepresentational* content. This is ultimately the view I prefer.

§1.3 looks a bit harder at the idea that genuinely empty names could be meaningful. It is one thing to say what truth values the sentences in question should have if there are meaningful genuinely empty names, and another to say how there could be such names. I approach the problem in the context of Donald Davidson's [1965, 1984] idea of using truth theories for languages as theories of meaning. This bears on the present project in two ways. First, it would be good if the present treatment of empty names was compatible with the Davidsonian project. Second, getting the two projects to fit together gives us a concrete way of understanding linguistic knowledge involving empty names. I show how genuinely empty names pose a distinctive problem for Davdidson's approach, and consider two solutions, one involving some semantic blindness, and another which modifies the shape of the truth theories in a way proposed by Mark Sainsbury [2005].

1.1 Meaningful names without semantic contributions

- ### 1.11 Braun's problems

Denying (4) means holding that a meaningful name need not make a semantic contribution to sentences in which it occurs. This uncomfortable-looking piece of logical space is defended by David Braun [1993, 2005]. He identifies five related problems that his proposal will need to address [2005: §1]:

- Meaningfulness for names: if a name's meaning is its referent, how can there be meaningful names without referents? This is the problem which arises most directly from denying (4).
- Meaningfulness for sentences: how can meaningful sentences have empty names in them, if the name leaves a gap in the semantics?
- Truth value: what are the truth values of (the contents of) these sentences? How can sentences without semantic contents have truth values, when the truth value of a sentence is meant to be the same as that of its content?
- Attitude ascriptions: if 'Vulcan is a planet' has no content, then 'that Vulcan is a planet' should fail to refer, so 'Leverrier believes that Vulcan is a planet' cannot be true.
- Belief and sincere assertive utterance: how can people sincerely assert 'Vulcan does not exist' if the assertion has no content for them to believe?

Braun puts forward his theory of gappy propositions (hereafter GPs) to solve all these problems. In this chapter I will only look at it as a potential solution to the first three. The problems of attitude ascriptions and belief and sincere assertive utterance will be dealt with in chapter three, with Braun's own solution examined in §3.42. Other issues about beliefs corresponding to genuinely empty names will be dealt with in chapter two.

- **1.12 Braun's Solutions**

Millians famously take the semantic value of a name to be just its referent, and this notoriously means co-referential names have the same semantic value. It also seems to mean that if a name has no referent, it has no semantic value, and so will leave a gap in the proposition the sentence expresses where a referring name would put a referent. Braun takes this in the most straightforward way possible, saying that the sentence will express a GP.

'Mars is a planet' and 'Vulcan is a planet' are both of the form [name]^[predicate], and (on Braun's Russellian view of propositions) this makes them express proposition with the structure represented by the ordered pair schema <object, property>. In 'Mars is a planet', the name refers to Mars, so the proposition is the one we can represent as <Mars, being a planet>. In 'Vulcan is a planet', the name does not refer, so the proposition is the GP we can represent as <___, being a planet>. Braun is clear that this is only meant as a way of representing the GPs: he is not identifying them with ordered pairs of blanks and properties, and has no declared interest in reducing structured propositions to ordered sets.[2] For Braun, GPs are propositional structures with gaps in,

[2] He emphasises [2005: n.6] that while ordered pairs can represent propositions, they need not be identical with them. Presumably some people will be less comfortable with an ontology of structured propositions which does not reduce in any natural way to one of sets. A fairly natural reduction can be effected if one is wanted, however. The proposition that *a* is *F* could be <{a}, {F}> and the corresponding GP could be <{}, {F}>. This is what Braun calls *Convention 2*. It could also be extended to allow the proposition that *a* and *b* were collectively *F* to be <{a,b}, {F}>, at least if we ignore propositions collectively ascribing a property to the members of a proper class.

and still count as propositions. If a gapless proposition is a propositional structure with all the gaps filled in, then a GP is a propositional structure without all the gaps filled in.

At this point we might want to ask for some clarification. What is a propositional structure? Is it the sort of thing that can have a gap in? It would be disingenuous to say that if you're committed to propositions then you're committed to propositional structures, and since you're committed to the structures you're committed to gappy instances of them. Some structures can have gappy instances and some can't. We know what it is for a car to have a wheel missing, but we don't know what it is for a set or n-tuple to have a member missing. Of course, Braun says that the propositions need not be identified with sets or n-tuples, so GPs don't commit us to gappy sets, but sets still provide a counterexample to the move from a commitment to structured entities to a commitment to gappy instances of their structure. Accepting GPs would thus be an additional theoretical commitment of some kind. Whether this addition is problematic will depend on the role we want propositions to have. If they are just for describing what utterances are about, GPs are probably alright. If propositions are supposed to play a robust explanatory role, then GPs may be more problematic. At this stage we can just note that they are an additional theoretical commitment, and they will have whatever properties they need to play the role Braun assigns to them.

Now we have some idea what Braun has in mind when he talks about GPs, we can see how they help with the problems set out in §1.11. First, consider the problems of meaningfulness for names and for sentences. Braun notes [2005: 600] that we do not judge empty names to be meaningless because we stand in significant cognitive relations to them which we do not stand in to nonsense strings like 'thoodrupqua'. We judge them meaningful whether they have a semantic content or not. This is true: we certainly think that empty names are meaningful, especially when we do not know they are empty, but also when we do. ‘Zeus’ doesn’t seem meaningless the way ‘thoodrupqua’ does.

Can we make anything of meaningfulness without semantic content? Let’s see how a view like that might go. We can start by considering the distinction between semantics and syntax. We can say that ‘Vulcan’ is syntactically a name, so it makes a syntactic contribution to the sentence by helping to determine its syntactic structure, which in turn determines the structure of the proposition it expresses. Combined with the view that ‘Vulcan’ does not refer, we can say that the name is syntactically meaningful but has no semantic value. The sentence is syntactically meaningful too, and if we have GPs then we can say it expresses a proposition, which could be enough to give it a semantic value when it appears embedded in an attitude context. What about when it appears by itself? Then its semantic value, if it has one, is its truth value. The truth value of a sentence expressing a proposition is generally taken to be the same as the value of the proposition. So to find the sentence’s truth value, if it has one, we need to find the truth value of the proposition it expresses, if it has one.

Braun says the GP expressed by ‘Vulcan is a planet’ is false, because it is atomic and has a gap where the object should be, and that makes it false. For Braun, all atomic GPs are false, including that represented by <__, existence>, and they compose truth-functionally just like gapless propositions. So the negation of an atomic GP is true, the disjunction of an atomic GP with another proposition is true iff the other is true (because the gappy disjunct is false), and so on. This gives us what is called a two-valued *negative free logic* for the language containing the empty names, with a standard syntax for predicate logic, and the semantics below:

- $V_A(x) = A(x)$, where A is an assignment of members of the domain to variables.

- $V_A(a) = V(a)$, where V maps some names to members of the domain and is undefined for the others.
- $V_A(F) = V(F)$, where V maps predicates onto sets of ordered n-tuples of members of the domain.
- $V_A(Ft_1…t_n) = T$ iff $V_A(t_1),…,V_A(t_n)$ are all defined, and $<V_A(t_1),…,V_A(t_n)> \in V_A(F)$; otherwise $V_A(Ft_1…t_n) = F$.
- $V_A(\neg\varphi) = T$ iff $V_A(\varphi) = F$; otherwise $V_A(\neg\varphi) = F$.
- $V_A(\varphi\&\psi) = T$ iff $V_A(\varphi) = V_A(\psi) = T$; otherwise $V_A(\varphi\&\psi) = F$
- $V_A(\varphi v\psi) = T$ iff $V_A(\varphi) = T$ or $V_A(\psi) = T$ or both; otherwise $V_A(\varphi\&\psi) = F$
- $V_A(\exists x\varphi) = T$ iff $V_B(\varphi) = T$ on some assignment B such that $B(y) = A(y)$ for all variables y: $y\neq x$.
- $V(\varphi) = T$ iff $V_A(\varphi) = T$ for all assignments A.

This gives us some reasonably intuitive results. ‘Vulcan is a planet’ will be false, ‘Vulcan is not a planet’ will be true, and ‘Vulcan is not a planet’ will not entail ‘something is not a planet’. Perhaps you don’t think the second of these is so intuitive, but Braun can mitigate this by identifying two readings of ‘Vulcan is a planet’, one true and one false. We can cash this out formally using lambda predicates, where $[\lambda x_1…x_n.\varphi]$ is true of some things when they satisfy a predicate determined by the open formula φ:

- $V_A([\lambda x_1…x_n.\varphi]t_{1,…}t_n) = T$ iff $A(t_i)$ is defined for all i: $1 \leq i \leq n$; and $V_B(\varphi) = T$, where $B(x_i) = A(t_i)$ for $1 \leq i \leq n$ and $B(x) = A(x)$ for other variables x. Otherwise $V_A([\lambda x_1…x_n.\varphi]t_{1,…}t_n) = F$.

This allows us to analyse ‘Vulcan is not a planet’ as ¬Planet(Vulcan), which is true, or as [λx.¬Planet(x)](Vulcan), which is false. This strategy of disambiguation using lambda expressions is always available, so if we think there is a false reading of ‘either Vulcan is a planet or grass is green’, we can take that as false too. Perhaps an unnatural-sounding natural-language paraphrase of this would be ‘Vulcan has the property of being such that either it is a planet or grass is green'. If you think this reading is never available, then don't analyse people's utterances of 'either Vulcan is a planet or grass is green' that way. Once we already have some GPs, however, it is not an obvious cost to admit such propositions and to be able to express them, even if in practice people usually don't.

The idea that sentences containing empty names could exhibit such scope ambiguities and that our analyses should accommodate them is not new. Russell [1905] held that names should be analysed as definite descriptions, and definite descriptions included an assertion that some unique thing satisfied the description. He also held that there is always a reading which gives this existential assertion wide scope. 'Either Vulcan is a planet or grass is green' can be understood as 'There is exactly one [insert description corresponding to 'Vulcan'] and either it is a planet or grass is green', which is false. Here is Russell talking about wide scope readings of the existential quantification in definite descriptions:

> ...when we say "George IV wished to know whether Scott was the author of *Waverley*," we normally mean "George IV wished to know whether one and only one man wrote *Waverley* and Scott was that man"; but we *may* also mean: "One and only one man wrote *Waverley*, and George IV wished to know whether Scott was that man". [Russell 1905: 489]

He goes on to say explicitly that sentences containing empty names like 'Apollo' will be false when the quantifier has wide scope, which he calls a primary occurrence of the name:

> A proposition about Apollo means what we get by substituting what the classical dictionary tells us is meant by Apollo, say "the sun-god". All propositions in which Apollo occurs are to be

> interpreted by the above rules for denoting phrases. If "Apollo" has a primary occurrence, the proposition containing the occurrence is false; if the occurrence is secondary, the proposition may be true. [Russell 1905: 491]

Likewise, Grice [1969: §3] held that names were subject to scope ambiguities in this way. He used a system of subscripts representing the orders in which the syntactic formation rules for sentences could be applied, and then used the subscripts in the semantics to get the different readings with their different truth values. The lambda predicates get equivalent results. Provided we allow a plenitude of GPs, including ones like <___, not being a planet> and <___, being such that either you're a planet or grass is green>, we can explain why utterances have their different readings and these readings have the truth values assigned by the negative free logical semantics given. At least, we can explain this if we can explain why atomic GPs are all false. Braun's argument [2005: §4] for this is the most tentative part of his proposal, and in §5 he argues that it would not matter much if atomic GPs lacked truth values, because we could still rationally believe or disbelieve them. However, let's consider his argument that they are false.

Braun's argument is essentially based on bivalence for propositions, combined with the claim that GPs are propositions. These GPs are not true, and since they are propositions they are therefore false, at least in some sense of 'false'[3]. He takes as his foil an argument he extracts from a footnote in Salmon [1998: n.54]. In fairness to Salmon, he was not

[3] It has been pointed out that bivalence does not entail that every proposition is either determinately true or determinately false (see e.g. Barnes and Williams [2011]). This distinction is worth mentioning because one might think that empty names were a candidate case for indeterminacy. However, Braun only needs bivalence, not determinate bivalence, because his argument is that atomic GPs are true or false, and not true, therefore false. Determinacy does not come into it.

arguing but explicitly assuming that GPs, which he calls *structurally challenged propositions*, are neither true nor false. Salmon says that the position is intuitive, and points out that if we enumerated the things that are bald and the things that at are not bald we would not find Nappy ('Nappy' is the empty name Salmon uses) in either list. We have already seen how we can use lambda predicates to give a false reading of 'Nappy is not bald', corresponding to the fact he is not among the non-bald. With that in mind, here is Braun:

> Salmon seems to assume that atomic gappy propositions are false only if all untrue things are false. But if all untrue things are false, then Piccadilly Circus and Russell's singleton set are false. The latter are not false, so atomic gappy propositions are not false. The weak link in this argument is the premise that atomic gappy propositions are false only if all untrue things are false. On the Gappy Proposition Theory, atomic gappy propositions are distinctive because they are objects of belief and assertion, and so are *propositions*. Only propositions, or items that express propositions, can bear truth values. Piccadilly Circus and Russell's singleton set are not propositions, and do not express propositions. So atomic gappy propositions are false, though Piccadilly Circus and Russell's singleton are not. [Braun 2005: §4.1; his emphasis]

Braun thinks that GPs must be propositions because they are the objects of belief and assertion. We can probably grant this, if we grant GPs at all. The whole point of introducing GPs was to have something to assign as the propositional contents of assertions involving empty names and the beliefs they express. On the other hand, admitting that GPs are propositions widens the category of propositions, which in turn makes unrestricted bivalence for propositions less appealing. We need an argument for the claim that all objects of belief and assertion are true or false, and that is a strong claim, particularly if we allow that assertions containing empty names have objects. The dialectic is moving in a fairly tight circle here. We'll try to break out of it §1.13. Braun's proposal has an internal coherence but it would be good if we could get similar results by an alternative and perhaps less controversial route.

First, however, it is worth looking at Braun's response to an objection to his truth values which he attributes to Fred Adams and Robert Stecker [1994]. If GPs are the semantic values of open formulas, and the semantic values of open formulas lack truth values, then GPs will too. Braun's reply is that GPs are simply not the semantic values of open formulas, and he points to some differences, in particular that the semantic values of open formulas vary with respect to assignments and that of 'Vulcan is a planet' does not. This response seems fair enough: nothing forces us to say that GPs are the semantic values of open formulas; indeed, there is no obvious reason to say that open formulas have contents except relative to assignments, and relative to assignments their contents will not have gaps corresponding to the variables. So I don't think this objection needs to worry Braun, or another GP theorist, but considering it may help clarify what GPs are supposed to be. The gaps aren't waiting to be filled, or sometimes filled by some things and sometimes by others. The gaps in GPs are as stable as the constituents of ordinary propositions.

• 1.13 Other routes to pessimistic truth values

The truth values assigned by the negative free logic are what we can call *pessimistic*. They are no kinder to people who say 'Vulcan is a planet' than to people who say 'Vulcan is an ostrich'. We could have what is called a *positive free logic*, where atomic sentences containing empty names can be true, and that could say that 'Vulcan is a planet' was true and 'Vulcan is an ostrich' was still false. Andrew Bacon [2013] defends a positive free logic for empty names, although he is more concerned to give optimistic truth values to sentences ascribing intentional or referential relations than to make ‘Vulcan is a planet’ come out true. A different positive free logic could be optimistic about ‘Vulcan is a planet’, though.

A two-valued negative free logic is not the only way of assigning pessimistic truth values, however. We could say that while atomic sentences containing empty names were neither true nor false, a disjunction of something neither true nor false with something true is itself true. This would correspond to the strong Kleene three valued logic K3. Alternatively, we could say that sentences containing genuinely empty names were never true or false, using the weak Kleene logic. Finally, we could say that they were always false, using the weak Kleene logic but interpreting both value 0 and ½ as kinds of falsity.[4] All of these assignments of truth values are distinguished by having the values insensitive to which empty name appears in a sentence: they can all be substituted for one another without changing the truth value.

Braun’s argument for GPs having the pessimistic truth values he takes them to have can be seen as metaphysical: he argues that GPs are propositions and then argues for their nature by analogy with other propositions. We can argue for pessimistic truth values, and ideally for the two-valued negative free logic, in a different way. We see what truth values we ought to assign the utterances and beliefs in question, and then argue that the GPs should be assigned these values because those are the values of the utterances and beliefs whose contents they are. This reasoning can either take the utterances and beliefs to be the primary truth-bearers and say the GPs have those values derivatively,

[4] K3 extends the classical truth tables by saying that the negation of a sentence taking value ½ has value ½, a disjunction has the maximum value of its disjuncts and a conjunction has the minimum value of its conjuncts. Weak Kleene extends the classical tables by saying that any compound including a sentence taking value ½ gets value ½. Both are from [Kleene 1952: §64]

or it can take the GPs as the primary truth bearers and assign them these truth values to explain why the utterances and beliefs have the truth values they do. How exactly that goes will depend on what you think propositions are for.

A big part of the defence of pessimism ultimately relies on the material in chapters three and four. These give alternative analyses of a lot of the uses of apparently empty names which we would be most tempted to assign optimistic truth values. If pessimism only applies to names introduced in the context of mistakes and lies, then it is more plausible. We will also have the option of giving optimistic truth values for sentences including even mistaken or mendacious names when they are used in attitude ascriptions. This strategy removes a lot of the data which the optimist might have used to support their case.

With the stakes thus lowered, we can look for alternative arguments to Braun's metaphysical argument. We could use a different metaphysical argument: sentences with empty names in them don't express propositions, so they don't have truth values, so the proper logic is weak Kleene, with ½ interpreted as 'neither true nor false'. If however we want to argue via the truth values of the utterances and beliefs, then one place to start is to think about what people are trying to do when they assert and believe.

Primarily, speakers and believers are trying to represent things accurately. The accuracy or inaccuracy of their representations is what truth-value assignments are supposed to track. Now, when someone says 'Santa is coming', they are representing just as inaccurately as someone who says 'the Taj Mahal is coming'. Furthermore, when someone says 'Santa isn't coming', knowing that 'Santa' doesn't refer, they are representing things correctly. They know how things are, and they use their utterance to spread their knowledge to others. Now, if a child says 'Santa isn't coming', maybe they aren't representing things correctly: they are trying to represent Santa staying away. This could motivate analysing it as [λx.¬Cx]s, rather than ¬[λx.Cx]s or ¬Cs. But in all these cases, the fact that the names don't refer does not put a middle ground between successful and unsuccessful representation. This can motivate the negative free logic, which in turn can motivate assigning GPs the truth values Braun assigns them, to either describe or explain the truth values of the utterances.

Is there a way to motivate the strong Kleene truth values? There may be, based partly on the nature of propositions and partly on the nature of utterances. The idea is that atomic utterances are meant to express atomic propositions, and get the atomic propositions' truth values, but compound utterances express something different about the atomic propositions of their atomic sentential constituents. So if I say 'either Vulcan is a planet or grass is green', that is true if one of the disjuncts expresses a true proposition, false if both express false propositions, and truth-valueless otherwise. If I say 'Vulcan is not a planet', taken as a negation, this is true if 'Vulcan is a planet' expresses a false proposition, false if it expresses a true proposition, and truth-valueless otherwise. This way we don't have to commit to GPs, but we can still have 'either Vulcan is a planet or grass is green' come out true.

Another way getting truth values for the utterances without committing to GPs is tentatively suggested by Kripke [2013: 156-60]. The idea is to take some sentences, like 'Vulcan is red' or 'There are no bandersnatches in the Arctic' as saying that there is no true proposition that Vulcan exists, or no true proposition that there are bandersnatches in the Arctic. He doesn't want to view this analysis as metalinguistic, saying that this reading would be 'subject to the same kind of difficulties as the metalinguistic analysis is elsewhere' [2013: 157]. If we do take it as metalinguistic – saying that 'Vulcan is red' does not express a true proposition – then it will indeed be subject to problems like Alonzo Church's [1950] translation test. (The analysis makes the statement about an English sentence, and so becomes implausible when translated into another language, since the e.g. German for "'Vulcan is red'" is "'Vulcan is red'".) If we do not understand it that way, however, it seems that the new analysis won't express a proposition either. If there is no proposition that bandernatches exist because there is no such kind of thing as a bandersnatch, then it seems there will be no proposition that true propositions *that bandersnatches exist* exist, because there is no such kind of thing as a true proposition that bandersnatches exist. Basically Kripke's proposal assimilates all the problematic statements to existential statements, but the new existential statements are themselves just as problematic. As such, I don't think Kripke's proposal works as it stands, and we would be better off with one of the others.

We still have a range of options, some of which involve commitment to GPs, and some of which don't. Ultimately I will argue in §4.42 that a Fregean ontology of GPs more fine-grained than Braun's can be well motivated, and this fits well with a two-valued negative free logic. However, we don't have to rely on that to get pessimistic truth values for sentences containing genuinely empty names. Which option we go for will depend on how willing we are to commit to an ontology of GPs, and what exactly we think the relationship is between a sentence's truth value and the proposition if any that it expresses. None of the options undermines the possibility that empty names and the sentences containing them can be syntactically meaningful while being semantically, or at least referentially, defective. We will look more at what the meaningfulness could amount to in §1.3.

1.2 Everett's objection

Anthony Everett [2003] makes the point that these three are not sufficiently differentiated by Braun's theory:

1. Santa Claus does not exist.
2. Father Christmas does not exist.
3. Hamlet does not exist.

Everett says that (1) and (2) say the same thing and (3) says something else, or at least if this is wrong then the intuition needs explanation. This objection is reasonable, but it can be met, and in more than one way. Examining the different ways of meeting the objection will help us understand what is going on when people use genuinely empty names.

- **1.21: A Millian solution**

Forgetting about empty names for a moment, consider these three sentences:

A. New York does not exist.
B. The Big Apple does not exist.
C. Chicago does not exist.

Here we have an analogous situation with referring names instead of empty ones. There is an intuition that (A) and (B) must say different things from each other because one could rationally believe one and not the other, if they thought that 'the Big Apple' was the English translation of 'El Dorado', for example. (C) must say something else, because it could be true (or truth-valueless) while the others were false. The intuition that (A) and (B) say different things is related to Frege's [1952] argument that 'Phosphorus is Hesperus' and 'Phosphorus is

Phosphorus' must have different contents because one is informative and the other not, and one could believe one and not the other.

Millians, especially Salmon [1986: ch. 8], tend to argue that (A) and (B) actually say the same thing, but (usually) suggest different modes of presentation when embedded in attitude contexts. The intuition Frege was playing off might be unsatisfied by this position, but if so then the Fregean should not say that (1) and (2) say the same thing either. If they concede that they do, then given the foregoing, all the same considerations might apply equally to (1), (2) and (3) saying the same thing, if anything. The idea is the same: that of grasping the same content under different modes of presentation and so believing it one way but not the other. Everett's objection plays two conflicting intuitions off against each other: the one which individuates content by mode of presentation and the one which individuates contents by truth conditions.

That was pretty quick, but it should make us think a bit harder about the objection and the intuitions behind it. Saying that (1) and (2) have the same content (or both have no content) allows for a certain amount of content misrecognition, and lumping (3) in with them as well could be more of the same. The distinctions in play here may not be distinctions of content. In fact, I will argue in §4.41 that the things said in defence of Millians here, by for example Jennifer Saul [1998], cannot all be applied equally to Braun's GPs, at least when we are dealing with attitude ascriptions. Nonetheless, if we have accepted the idea of content misrecognition in one place, as the Millian must, it gives us an initial response to Everett. It would however be nice if the Millian could say something principled about what (1) and (2) have in common with each other but not with (3). In the next section I will give them something to say.

1.22: A metalinguistic solution

The explanation we will give of what (1) and (2) have in common with each other but not with (3) need not be confined to cases involving empty names. The same phenomenon probably arises in cases such as Kripke's [1979] 'Pierre' and 'Paderewski' puzzles about belief, and Salmon's [1986: §7.2] puzzle about Elmer and Bugsy Wabbit[5]. It will however plausibly apply to any puzzle involving empty names similar to Everett's. The explanation uses Donnellan's [1974] metalinguistic notion of a *block*, which he introduces to give truth conditions for singular existentials. 'Block' is not rigorously defined, but is supposed to capture the idea of a name being introduced without successfully attaching to a referent. He puts forward this rule:

> (R) If *N* is a proper name that has been used in predicative statements with the intention to refer to some individual, then ⌜*N* does not exist⌝ is true if and only if the history of those uses ends in a block. [Donnellan 1974: 25][6]

(R) is not meant to give the meaning of '*N* does not exist', but it is supposed to be true. 'Zeus exists' is not synonymous with 'The history of "Zeus" does not end in a block', because the former sentence is not about 'Zeus' (or blocks) and the latter is. Donnellan is aware that he has not given an analysis, but (assuming everything exists and that negative existentials containing empty names are true) he is right about the rule,

[5] The Pierre puzzle is more discussed in the literature, but the Paderewski puzzle is in some ways harder. The extra complication is that 'Paderewski', if it is equivocal at all, is only so in the believer's idiolect. The same issue arises in Salmon's puzzle. There is more discussion of these puzzles in the next chapter.

[6] Donnellan's formulation 'proper name that has been used in predicative statements with the intention to refer to some individual' may incidentally help us to get a handle on what it would take for something to be syntactically a name but lack a semantic value.

since the histories of all and only empty names will end in blocks. Fred Adams and Robert Stecker [1994] say that on Donnellan’s view, if a negative existential statement expresses a proposition, it must be the metalinguistic proposition that the name does not refer, or that its history leads to a block.

> Although Donnellan gives us the ‘truth’ rule (R), he never tells us exactly which proposition 'Vulcan does not exist' expresses. One possible proposition expressed is ‘“Vulcan” does not refer’. Indeed, other than our view that no proposition is expressed, this is the only possibility we can think of for a proposition expressed (dismissing description theories, as Donnellan does). [Adams and Stecker 1994: 395]

As I read Donnellan, he definitely does not think that the metalinguistic proposition is what is expressed. He discusses the issue in §7 of his paper, and seems fairly clear that the metalinguistically expressed truth-conditions of an utterance come apart from the proposition expressed by that utterance, even in non-empty cases. ‘Cicero is wise’ is true iff the referent of ‘Cicero’ is wise, while ‘Tully is wise’ is true iff the referent of ‘Tully’ is wise, even though they express the same proposition. He does not settle on an answer for the proposition expressed by a sentence like ‘Vulcan does not exist’. Braun thinks it expresses a GP, and it is obvious that if there are GPs then they are candidate contents for ‘Vulcan does not exist’, in competition with the metalinguistic content which Donnellan rejects.

Donnellan worries that ‘Santa Claus does not exist’ and the French ‘*Père Noël n’existe pas*’ will not turn out to be translations of one another, because they involve different names. However, different names’ histories do not always lead back to different blocks. The referent of ‘London’, if any, must be the same as that of ‘*Londres*’ and *‘Londinium’*, because the histories of those three names all (let’s assume) converge, dividing either by corruption or translation, but leading back to the use of a single name[7]. Likewise with 'Father Christmas', 'Santa Claus' and '*Père Noël'*. Donnellan recognizes this:

> ...in the example before us, and others one can think of, our inclination to say that people using different empty names express the same negative existence proposition seems to be a matter of historical connection between the blocks involved. In our example, it seems to me that the reason we think both children express the same proposition is that the story of Santa Claus and the story of Père Noël, the stories passed on to the two children as if they were actual, have a common root. And if there were not this common history, I think we should rather hold that the two children believed similar, perhaps, but not identical falsehoods... [Donnellan 1974: 30]

Now we have something to say to Everett. Since the histories of 'Santa Claus' and 'Father Christmas' converge as they are traced back, it will have to be the same block for each name. The two uses of 'Bugsy Wabbit' in Elmer's idiolect in Salmon's puzzle are similar. There are facts about history from which it follows (given some facts about how reference is determined and transmitted) that (1) is true iff (2) is true. It also follows that if 'Santa Claus' and 'Father Christmas' refer, then

[7] The idea of the history ending (starting?) in a block should be taken with a pinch of salt. Gareth Evans [1973; 1982: ch.11] points out that the connection between a name and its referent is more complicated than a simple baptism followed by transmission of reference from speaker to speaker. Names can probably become empty, cease to be empty, and change referent. However, while this makes the situation more complicated, it should not change the consequences for the present argument. The required convergence is that the histories of use converge more recently than any changes in the referents of the names, so they have the referent now that the name had then, if any, and otherwise no referent.

‘Santa Claus is Father Christmas’ will be true. This does not mean ‘Santa Claus is Father Christmas’ is actually true[8]. (Depending on our other commitments, we could have an intensional predicate tracking the metalinguistic phenomenon, such that ‘s.c. ≈ f.c.’ was true. I include this option in the semantics given in the appendix to chapter three.) Donnellan wants to say that there are particular truths and falsehoods being believed here, and that (1) and (2) have the same contents and (3) has a third. We need not commit ourselves to anything that strong at this point, since all we want to do is point to something that (1) and (2) share with each other and not with (3), to explain the undeniable feeling that they do. However, perhaps we do also want to commit ourselves to some kind of sameness of content.

1.23: A Fregean solution

Sam Cumming [2007, 2008, forthcoming] identifies a level of content he calls *discourse content*[9]. This is a kind of Fregean content constituted by discourse entities, which are socially constructed abstract objects which form the scoreboards in language games, in the terminology of Lewis [1979]. I will go into Cumming’s view in more detail in §2.12, but for

[8] It is tempting here to use the connection between existence and identity to argue the point, saying that existence is being something, so identity statements involving empty names must be false. It’s a defective argument though, because if Santa is said to be identical to Santa, and Santa is not something, then the identity statement doesn’t entail the existential one. We need to argue on independent grounds that the identity statement is false, and give an alternative account of the connection between the two names which denies that they are coreferential, or that the identity statement is true. That is what I am doing.

[9] It may understate Cumming’s ambitions to call it just a level of content. I will not get into whether it is the primary level here though, the important thing being that Cumming offers us a way to say what (1) and (2) have in common if we are willing to reify a certain level of content.

now the important thing is the idea of having a level of content which isn't individuated by what is represented, but by what is being done to try to represent the world. Even if these attempts are unsuccessful, we can talk about the attempts. If they are attempts which more than one person can participate in, for example by using the same word in deference to the same people, then we can get a level of content capturing interpersonal generalizations off the ground. This is what Cumming does. We could think of it as a kind of metalinguistic content, or since not all attempts to represent the world are linguistic, *metarepresentational* content. David Chalmers [2011] does something similar when he constructs something like Fregean senses out of A-intensions, which are functions from epistemically possible scenarios to semantic values.

Gillian Russell [2008] offers another proposal which would allow us to cash out Donnellan's suggestion, this time without invoking an extra level of content. She proposes to rehabilitate the analytic/synthetic distinction by thinking of analyticity as truth in virtue of *reference determiner*. Some things are done to make words have the referents they have, if any, such as pointing, baptizing or using indexicals. Sometimes these things are enough to ensure that utterances of those sentences are true. For Russell, these are the analytic sentences. 'I am here now' is one of her examples, following Kaplan [1989: 508-9]. Perhaps a more common situation is when a sentence will be true in virtue of reference determiner unless there is reference failure, in which case they might not be true. She calls these *pseudo-analytic* [2008: 100-5]. 'Hesperus is Hesperus' is a straightforward example, and she offers 'Mohammed Ali is Cassius Clay' as a less trivial example[10]. If

[10] The reason this works is that (let's suppose) one name was used in fixing the reference of the other: "The referent of *Mohammed Ali* was introduced in a slightly different way, when Elijah Muhammad, the leader of the Nation of Islam, said *Let's use 'Mohammed Ali' to name Cassius Clay*. *Mohammed Ali* thus refers to whatever object, if any, *Cassius Clay* refers to." [Russell 2008: 58-9]

we buy into this, we can say that two sentences are *pseudo-analytically equivalent* iff their reference determiners ensure that they have the same truth values if any. Now we can say that (1) and (2) are pseudo-analytically equivalent, as are any sentences with 'Santa Claus' substituted for 'Father Christmas' or vice versa. We can also say that 'Father Christmas' and '*Père Noël*' are pseudo-analytic translations of one another, in that they would be suitable translations for producing pseudo-analytically equivalent French and English sentences. Obviously the other words would have to be pseudo-analytically translated too, which might be difficult in many cases, but it should at least allay Donnellan's worry that 'Santa Claus does not exist' and '*Père Noel n'existe pas*' will not turn out to be translations of each other. Even if we ultimately decide they have no content worthy of the name, we can still say that they are pseudo-analytic translations in this sense.

Russell's proposal does not demand that sentences containing empty names have contents, because their reference determiners may fail to give them one. If two sentences are pseudo-analytically equivalent, the reference determiners of both will succeed or fail together, and the truth values if any will always be the same. A difference between empty and non-empty names is that the latter can co-refer even if their reference determiners do not demand it. The histories of 'Phosphorus' and 'Hesperus', for example, do not converge until you get to Venus itself and the initial baptisms[11]. This special kind of convergence is unavailable for empty names: two empty names cannot have been successfully used to baptize the same object because they were not successfully used to baptize any object. We could however have two

[11] One might think they have converged now because everyone either knows that the morning star is the evening star or defers to someone who does. This kind of thing means that any uncontroversial example of non-convergence would have been a controversial example of co-reference. We can however say uncontroversially that 'Phosphorus' and 'Hesperus' used to be an example of co-reference without convergence.

empty names which were analytic translations of each other without their histories converging, if their references were fixed by the same descriptions. Thus if two isolated communities both introduced a name for the thing than which none greater could be conceived, or the set of all non-self-members, they could be said to be having pseudo-analytically equivalent thoughts and making pseudo-analytically equivalent utterances even if the names were empty. I don't see this feature as a cost.

1.3 Empty names and truth theories

- **1.31: Davidsonian theories of meaning**

The rest of this chapter is about Donald Davidson's idea of using truth theories for languages as theories of meaning for those languages[12]. This relates to the present project in two ways. First, we have argued for assigning pessimistic truth values to sentences containing empty names. If the truth theory for the language is meant to serve as its theory of meaning, then we need to make sure the truth values we assign do not generate implausible consequences for the theory of meaning. Also, if we can fit genuinely empty names into a Davidsonian theory of meaning, this gives at least one framework for understanding how names and sentences can be meaningful (and learnable) while being semantically defective in the way we have suggested. This applies to David Braun's proposal on which sentences involving empty names express Russellian GPs, but similar issues will arise for any theory evaluating sentences according to a negative free logic, including the system of Fregean gappy propositions I will cautiously endorse in §3.42. The discussion in §1.33 especially will tie in with that, introducing Fregean elements into Davidson's proposal. In short, we are interested in Davidson's framework for two reasons: we want to show that negative free-logical treatment of empty names does not clash with it, and it might provide one way of understanding how there could be a theory of meaning for a language containing genuinely empty names.

[12] Davidson [1965] introduces the idea as a constraint on interpreting a language, on the grounds that a language which couldn't be given an axiomatic truth theory would be unlearnable. He develops the idea in several other papers, most of which are in Davidson [1984]. Notable contributions to the programme by others are Davies [1981], Evans [1981] and Larson and Segal [1995]. It is possible that Davidson himself viewed the role of a theory of meaning slightly differently from his successors; I am thinking of it in terms of the propositional component of speakers' knowledge of their language.

So, what is a Davidsonian theory of meaning? One way to view a theory of meaning is as the set of propositions you need to know to be able to understand a language. Maybe you need some non-propositional knowledge too, or some skill which isn't knowledge at all, but it is reasonable to say you need at least some propositional knowledge. This applies to speakers, and it also applies to people who study the speakers and say what they mean. Davidson's idea is that the theory of truth for the language can serve as a theory of meaning, or be systematically transformed into one. Now we need to know what a theory of truth is.

A truth theory for a language is one where for each sentence of that language there is a T-sentence in the metalanguage, of this form:

> 'S' is true iff _____

These T-sentences would be true whenever the blank was filled by a sentence with the same truth value as 'S'. To serve as a theory of meaning, however, the truth theory must be *interpretive*. This means that the blank must be filled by a sentence of the metalanguage synonymous with 'S', and then the T-sentence can give the meaning of 'S'. For example:

> 'La neige est blanche' is true iff snow is white.

This is the T-sentence for 'La neige est blanche', where French is the object language and English is the metalanguage. The hope is that somebody could know a language by internalizing the contents of T-sentences for indefinitely many sentences of that language, by deriving them from finitely many axioms giving the semantic values of the words in the language and the ways the values of complex expressions depend on the values of their simpler parts. If someone knew the content of an interpretive truth theory for a language, that would be all the information they would need to understand the language. They would also probably need some practice at using it, but no new information.

There is an objection to (this retelling of) Davidson which is worth addressing here because it illuminates how this sort of thing is meant to work. The concern is that interpretive truth theories cannot be informative because all the T-sentences might be like the following uninformative-looking sentence:

T ‘Snow is white’ is true iff snow is white.

This would happen (at least for the context-independent fragment of the language) if the metalanguage contained the object language. Hilary Putnam objects along these lines [1975: 258-62]. As Putnam tells it, Davidson sounds to have been a little cowed by this objection in conversation, but I don’t think there is reason to be cowed by it now. There are two issues here. First, it is useless for you to give me a theory of meaning for (without loss of generality) English if you do it in English. This is to be expected though: either I will not understand the theory (because I don’t speak English) or I will already have the information (because I do). Second, it seems that sentences like T can’t contain any information because they are trivial instances of the disquotation schema. This isn’t right either though, since a theory of meaning is a set of propositions, not a set of sentences. It happens that we have a convention according to which propositions like that expressed by T can be expressed by sentences which, given how the meanings of their words are determined, could not be false. (This makes the sentences analytic, or at least pseudo-analytic, in Gillian Russell’s sense.) T is such a sentence. The speakers will not be able to internalize the proposition (in the first instance) under this mode of presentation though, because they don’t understand the language. They will have to learn it under a different mode of presentation. The proposition expressed by T is a non-trivial, contingent proposition about a sentence, snow and whiteness, and it is the proposition you need to know to decide whether to assent to ‘snow is white’ or not.

One might press the objection that someone could, on encountering an unfamiliar sentence of a familiar language, apply the disquotation schema to produce an interpretive T-sentence without thereby coming to know what the unfamiliar sentence meant. The correct response is that while they would know that the T-sentence they produced was true, they would not form a belief with the T-sentence’s content because they would not understand the sentence. You might as well have given them a T-sentence in an unfamiliar language and told them it was true.

We should bear in mind then that a T-theory, insofar as it is internalized by competent speakers, is embodied by a system of beliefs, not a system of natural language sentences. Each competent speaker, insofar as they are competent, will have a system of beliefs which have the same contents (if any) as the sentences of an interpretive truth theory for the language, and which would be correctly verbally expressed by giving such a theory. We can call the beliefs corresponding to the T-sentences *T-beliefs*. Call the beliefs corresponding to axioms of the T-theories *A-beliefs*. Since these beliefs may correspond to empty name sentences, they may lack contents or at least have gappy contents. The next chapter will include an explanation of how the speakers’ deductions can be valid even if the beliefs suffer from referential or intentional failure, which helps make sense of A-beliefs with gappy contents or no contents featuring as axioms of the speakers’ internalized T-theories. Now we have a sense of how Davidson’s idea works, we can see how empty names fit into it.

1.32 The Davidsonian problem of empty names

The principal issue with empty names arises because the meanings of the words in a language are supposed to be fixed by the truth theory for that language internalized by its competent speakers. That is how we distinguish the true interpretive T-theories from the true uninterpretive ones. The worry is that the only things fixing the meaning of an empty name are the speakers' A-beliefs saying what it refers to, and these beliefs are not true. As such, it seems like there is nothing left to give meaning to an empty name at all, and yet the speakers seem to understand each other. What do they know that non-speakers of the language don't?

The axiom for 'Vulcan', for example, is this:

(V) 'Vulcan' refers to Vulcan.

It must be so, because the speakers, not knowing that 'Vulcan' is an empty name, internalize an axiom much like the one for a non-empty name, e.g.:

(P) 'Pluto' refers to Pluto.

However, while (P) is true, (V) is on my account not true. Its truth condition is that the value of '"Vulcan"' refers to the value of 'Vulcan'. There is no value of 'Vulcan', so (V) is not true. This is fine in principle, because (V) simply captures the belief of the speakers that 'Vulcan' is non-empty, and their corresponding belief that Vulcan exists. They are indeed wrong about this, so it is understandable that the corresponding A-beliefs would not be true. This captures what they are getting wrong, but we still need to capture what they are getting right.

- ### 1.33 Semantic competence and semantic ignorance

We can start by pointing out that (V) is not really the only thing competent speakers internalize. They also know that 'Vulcan' is a name. This way they know how to get truth conditions for sentences in which it occurs. This may not sound like much, but we will see that it is all we need if the only meaning facts about 'Vulcan' are that it is a name and that it does not refer. The speakers are right that it is a name and wrong to think it refers to anything. That is the right result: they have the name's syntax right and its semantics wrong.

The reason this may be no bar to their understanding the language is that it may not give them any false or uninterpretive T-beliefs. Since what you come across and come out with when using a language are sentences, if you understand the meanings of those then you can understand the language well enough to use it. If someone asks you about the semantics of the language – for example if they ask you what 'Vulcan' refers to – you will give the wrong answer. This is not a problem of language mastery though, or at least not the aspect of language mastery which the Davidsonian approach is trying to account for, i.e. the ability to use a language as opposed to the ability to talk about it.[13]

[13] The view that the T-sentences are the only part of the T-theory constitutive of language mastery can be separated from the view, also associated with Davidson, that dispositions to respond to sentences are the only evidence admissible to the linguist in setting up a theory of meaning. Putnam [1975: 258-262] is understandably critical of this, saying that in practice and in theory the linguist can learn what words mean by asking speakers directly. Even if we hold that (beliefs in) T-sentences are constitutive of language-mastery, we can still agree with Putnam that information about semantic axioms is admissible evidence, because the T-sentences are derived from the axioms and so information about axioms is evidence for what the T-sentences are. However, admissible evidence need not be infallible, and it is possible that

We have the makings of a Davidsonian explanation for the speakers' incorrect semantic axioms not interfering with their language mastery. What about our language mastery? We don't believe the false axioms, but it isn't very plausible that to master a language you need to join its speakers in their ignorance – semantic ignorance – of which names are empty. A truth-theory with true axioms as well as true T-sentences, perhaps as established by the field linguist but in any case known by anyone who knows all the semantic facts about the language (rather than just enough to speak it), will presumably contain this sentence instead of (V):

(V*) 'Vulcan' has no referent.

The truth theory internalized by the speakers of a language does fix the syntactic categories of the expressions, presumably, but it cannot guarantee that the names have referents. This reflects that the speakers' implicit beliefs about their language's syntax are not fallible in the way that their implicit beliefs about semantics are. To be in a syntactic category a word just has to be used as such, whereas to have a referent its history needs not to end in a block.

Replacing (V) with (V*) in the field linguist's theory of meaning is not a complete solution to the problem, however, because the speakers will also have T-beliefs they would express like this:

(VP) 'Vulcan is a planet' is true iff Vulcan is a planet.

This sentence is (at least according to the negative free logic argued for in §1.1) true, since it is a biconditional both sides of which are false. It is

when we ask the speakers what a name means they will still answer falsely, e.g. by saying '"Vulcan" refers to Vulcan'.

also interpretive, since if ‘Vulcan is a planet’ is meaningful at all then the right-hand side of the biconditional is a fine translation. For the speakers to derive (VP) is easy, because they have internalized (V), and they can derive it the same way they derive (PP) using (P):

(PP) ‘Pluto is a planet’ is true iff Pluto is a planet.

The conscientious field linguist has more trouble, because she wants to put a synonym of ‘Vulcan is a planet’ on the right hand side, but has no word ‘Vulcan’ in her language. Why should she? Empty names are introduced by accident. Unless we have a very good reason for thinking otherwise, it makes sense to hold that everything true should be statable in a language that contains no empty names. Cian Dorr says something similar about translating ‘phlogiston’ into a language used by people who never came up with a phlogiston theory:

> It would be absurd for the Tritonians [who never came up with the phlogiston theory] to advocate linguistic reform on the grounds that without a new word, they will be unable to express the fact about chemistry expressed in English by the sentence ‘there is no phlogiston’. And their lack of any word equivalent to ‘phlogiston’ need not, intuitively, prevent them from stating a perfectly excellent semantic theory for English. [Dorr 2005: §16]

How easy it is to meet this desideratum will depend on how much there is to the meaning of a genuinely empty name. In this section we see what we can do if the only facts about a genuinely empty name’s meaning are that it is empty and that it is a name. If there is more to the meaning of a genuinely empty name than this, the account in §1.34 may be more appropriate, even though it may not be able to fully eliminate empty names from our metalanguage.

Since (VP) contains ‘Vulcan’, the linguist needs another interpretive T-sentence which only includes words in her own language or respectable additional technical terms. She knows (V*), that ‘Vulcan’ is a name, that the value of ‘is a planet’ is {x: x is a planet}, that ‘is a planet’ is a predicate, and that a sentence of the form [name]^[predicate] is true iff the value of the name is a member of the value of the predicate. I can offer two possible solutions. They complicate the canonical derivations of T-sentences but not in any damaging way, since the truth values of the object language sentences and the T-sentences for straightforward cases are left the same as before.

The simplest solution is this: whenever she comes across an atomic sentence which must be false because it contains an empty name she replaces the sentence with a sentential constant ‘⊥’, which is defined as being always false. So instead of (VP), she has this:

> (VP*) ‘Vulcan is a planet’ is true iff ⊥.

This captures only that ‘Vulcan is a planet’ is semantically defective with the effect of always being false. If this is all there is to say about the meaning of ‘Vulcan is a planet’, then (VP*) is interpretive. Braun thinks there is more to its meaning: it expresses the GP represented by <__, planethood>. In that case, what we need is a term in the language to indicate the semantic defectiveness of ‘Vulcan’ without entailing the complete defectiveness of ‘Vulcan is a planet’ and the like. We can use ‘__’. It is arguably not an empty name because is not a name at all; it is just a term whose semantic rule is that where it appears in an atomic sentence that sentence is false. If this is enough to make it a name, perhaps we can console ourselves with the fact there is only one of it. Where our conscientious linguist comes across an empty name *N*, she derives T-sentences by applying the canonical rules to the axiom ‘*N* refers to __’. There is nothing to stop her doing this, and the fact that ‘*N* refers to __’ is false does not matter, since she does not believe that axiom. If we choose this option the T-sentence will be (VP**):

> (VP**) ‘Vulcan is a planet’ is true iff __ is a planet.

(VP**) may be an improvement even if we don’t want GPs. Even if ‘Vulcan is a planet’ expresses no proposition gappy or otherwise, it is still plausibly about planethood, which (VP*) does not capture, and so (VP*) is arguably not interpretive. In any case, this solution is at least safer than that using ‘⊥’. If Braun is right and the contents of atomic sentences with empty names are GPs, ‘⊥’ is not fine-grained enough to accommodate the rich variety of contents. If he is wrong and they all mean the same thing – nothing more than is needed for them to be false – then sentences like ‘__ is a planet’ will mean that too. So we may as well go for the latter option. ‘__ is a planet’ (and ‘⊥’ if you prefer that option) are both well-formed sentences and enter into compounds like any others, so we get the correct result that the interpretive truth theory contains T-sentences like this:

> ‘Vulcan is a planet or grass is green’ is true iff __ is a planet or grass is green.

Two more examples:

> ‘Father Christmas is Santa Claus’ is true iff __ is __.

> ‘“Vulcan” refers to Vulcan’ is true iff ‘Vulcan’ refers to __.

This gets us what was wanted. In particular, it captures the speakers’ mistake in thinking that a name is not empty, while explaining their ability to understand their language because their T-beliefs are all true and interpretive. However, someone learning the language but knowing that the names are empty is also able to learn an interpretive T-theory, but derived from a set of axioms modified to say correctly that the empty names have no referents, and expressed in a metalanguage which contains no empty names itself. This is what Dorr and I wanted.

- **1.34 More tentative semantic axioms**

If the worst thing about the proposal we have just seen was that it attributed false A-beliefs to the speakers, this would probably not be a problem. The speakers do think 'Vulcan' refers to Vulcan, and that's false. If we give them credit for thinking 'Mars' refers to Mars, why not criticize them for thinking 'Vulcan' refers to Vulcan? This is not the only odd thing about the proposal though. The odd thing is that it doesn't distinguish between the meanings of any empty names, and so 'Vulcan is a planet' comes out synonymous with 'Santa is a planet'. Now, Braun thinks that they do indeed have the same content, and in §1.21 and §1.22 we saw how you might defend this. But maybe we still want to say they have different contents, and so "'Vulcan is a planet' is true iff __ is a planet" will not count as an interpretive T-sentence. For that, we need another proposal.

Mark Sainsbury [2005] provides one. His idea is to replace the semantic axioms for names with universally quantified ones with this form:

$$\forall x[\text{'n' refers to } x \leftrightarrow x = n]$$

According to the two-valued negative free logic Sainsbury works with, these will be true when 'n' refers, but they will also be true when 'n' is empty. The T-sentences will follow from these more tentative semantic axioms just as they do in the original proposal. This takes speakers as having true A-beliefs, even when they think the name is non-empty, because this existential commitment is not a commitment about meaning. As I said, not much turns on the truth of the speakers' A-beliefs: they are wrong about something and it does not affect their language mastery, and whether we call their ignorance strictly semantic is not a particularly substantive issue. There are two more important features of Sainsbury's proposal, one welcome and one potentially unwelcome.

The welcome feature is that the field linguist can believe the universally quantified semantic axioms, whatever they think about whether the name refers. From this they can derive T-sentences like "'Vulcan is a planet' is true iff Vulcan is a planet". We could even say "'Vulcan is a planet' means that Vulcan is a planet", where 'means' is taken to be a relation between an expression and the proposition referred to by the following 'that' clause.[14] This lets us say that T-theories have to put the right empty names in the right places to count as interpretive. This is how things should be if we agree, with the proponents of metarepresentational content but perhaps against Braun, that empty names are not all synonymous. This set-up also allows us to say that co-referring names need not be synonymous.

The potentially unwelcome feature of this account is that we lose Dorr's desideratum that it should be possible to give a theory of meaning for a language containing empty names in a language which doesn't contain correspondingly empty names. I agreed that this desideratum had some intuitive appeal. Is there anything we can say to feel better about not meeting it? Perhaps it would help to recall the newly tentative semantic axioms, and how they don't commit us to the names not being empty. Using a name and thinking it non-empty are just two different things.

[14] For a concrete proposal of a formal language admitting that kind of sentence, we could fit it into the language given in the appendix to chapter three. It would be formalized as *MEANS('Pv', THAT(Pv))*, where *'Pv'* refers to an expression and *THAT(Pv)* refers to a proposition. *MEANS* would work much like *BEL*, taking a term in the first position and a proterm (propositional term) in the second.

But it still seems weird that we need empty names to express all the facts.

Let’s step back and think about what the information is that we’re using the empty name to capture. It is the sense of the name. Saying “the sense of ‘Santa’” won’t do, if for no other reason then because we are trying to get acquainted with the sense of ‘Santa’ under a mode of presentation other than using the description ‘the sense of “Santa”’[15]. In principle, we could however presumably find out everything there was to know about the sense of the name, and present this information in such a way that we could have an operator O which combined with an appropriate expression E of this information to produce a term with the sense of the name we started with. Then we could have semantic axioms which looked like this:

$$\forall x[\text{‘n’ refers to } x \leftrightarrow x = O(E)]$$

These don’t contain any empty names as primitive expressions, and since ‘$O(E)$’ is stipulated to have the same sense as ‘n’, the resultant T-theory will presumably be interpretive. It won’t be practical to use this method, but that is to be expected: it really isn’t practical to capture the sense of the word ‘Santa’ without using the word or a synonym. But in principle, we probably could, and it is only our limitations as

[15] Kripke [2011d: 343-4] makes the similar point that the motivation of computability theory speaks in favour of what Quine [1961: 330] disparagingly called a ‘frankly inequalitarian attitude towards the various ways of specifying [a] number’. Just as ‘f(x)’ is not an informative answer to ‘what is f(x)?’, ‘the sense of “Santa”’ is not an informative answer to ‘what is the sense of “Santa”?’. This objection is different from Putnam’s objection discussed in §1.31, as can be seen if we translate the parts of the T-theories into another language. ‘Santa’ will still appear on both sides even when the metalanguage is not homophonic, whereas in the kind of case Putnam is talking about it wouldn’t.

investigators and speakers that force us to use the shortcut. Perhaps that is close enough to Dorr’s desideratum to keep its supporters happy.

1.4 Conclusion

There are lots of ways people use apparently empty names. This chapter has been about the worst kinds of case: names which are introduced in the contexts of mistakes and lies. People use them to try to represent the world, and in a sense they seem to be failing. On the other hand, the names seem to be meaningful in some sense, and there seems to be a difference between people who understand the names and people who don't. We have been trying to make sense of this.

§1.1 was mostly about assigning truth values to utterances containing genuinely empty names. David Braun has one proposal, involving gappy propositions (GPs). Utterances containing genuinely empty names express GPs, and GPs have truth values assigned in accordance with a two-valued negative free logic. The commitment to GPs is neither innocuous nor outrageous, and we looked at some other ways of assigning truth values to the utterances in question. The main claim was that we can assign pessimistic truth values to the utterances, at least in extensional contexts, and there is a reasonable case for using the two-valued negative free logic Braun uses.

A consequence of the pessimistic truth values is that you will be able to freely substitute one genuinely empty name for another (at least in extensional contexts) without affecting the truth value. If Braun is right, then it won't affect the sentence's content either. §1.2 thus considered Anothony Everett's objection that Braun's position does not have the resources to explain what 'Santa Claus' and 'Father Christmas' have in common with each other that they don't have in common with 'Hamlet'. We saw that it may be open to Braun to use some of the resources Millians use to challenge some other intuitions about co-referring names. Even if this does not ultimately work, we can still explain what needs explaining in metalinguistic or metarepresentational terms, following Donnellan. The metarepresentational explanation offers some resources for classifying utterances by a level of metarepresentational content, following Cumming and Chalmers. Alternatively we can regiment the phenomena using Gillian Russell's notion of pseudo-analyticity, without committing to a level of content.

§1.3 was about Davidsonian theories of meaning. While we seemed to be able to make sense of what would follow from names being meaningful but genuinely empty, situating them within a Davidsonian theory gives a specific proposal for explaining how such names are possible. It also removes the potential objection that genuinely empty names cannot be meaningful because the Davidsonian theory is correct and they can't fit into it. The Davidsonian problem of empty names is that the competent speakers' semantic axioms fixing the meanings of empty names will all be false, leaving nothing in virtue of which for the names to be meaningful. We saw two possible responses to this. We can say that the speakers' semantic axioms are false but their T-beliefs are nonetheless true and interpretive, explaining their semantic competence. This solution was most plausible if the only facts about empty names' meanings are that they are names and they are empty. If there are other meaning facts to capture, we can get an interpretive T-theory, for the speakers and for the linguist who knows the names are empty, by adopting Sainsbury's proposal for non-existentially committing semantic axioms. While this seems to mean that not all facts are statable in a language containing no empty names, we saw that this limitation may be practical rather than theoretical, which might be less of a problem.

The account in this chapter is only supposed to apply to the worst cases of apparently empty names. Other uses will not be treated so pessimistically, as we will see in chapter three when we look at attitude ascriptions, and in chapter four when we look at fictional names. First, however, we need to look at the beliefs people have corresponding to the kind of usage we have dealt with in this chapter, and the deductions people make involving those beliefs. That is the topic of chapter two.

Chapter 2 – Beliefs without Objects

2.0 Introduction

The last chapter was mostly about language. This chapter is about thought. It is reasonable to suspect that if there are linguistic practices involving names without referents then they will give rise to beliefs, or at least mental states like beliefs, that are not about things in the world in the way ordinary beliefs are. Exposure to communicative practices involving the referring name 'Winston Churchill' and the empty name 'Santa' induce psychological states in a child which are intrinsically the same kinds of thing, and which are commonly described as beliefs about Churchill and Santa. The latter are instances of the kind of thing which this chapter is about. We can call such psychological states 'gappy beliefs', while remaining neutral on whether gappy beliefs are really beliefs. Perhaps to count as a belief a state has to have a propositional content, and I won't return to the issue of gappy beliefs' contents until chapter three. Instead of talking in terms of contents, this chapter will talk about beliefs (and gappy beliefs) and the things and properties they are about. We will also postpone discussion of attitude ascriptions to chapter three: now we are concerned with the attitudes themselves (a cognitive phenomenon), rather than our ascriptions of them (a linguistic phenomenon).

A lot of this chapter will be about Frege's puzzle about co-referring names and the corresponding beliefs. This is because the puzzles both arise from a mismatch between beliefs and the things they are beliefs about. Frege's puzzles arise when we have co-reference, and empty name puzzles arise when we have no reference. To explain how gappy beliefs fit into an account of the relationship between beliefs and the things they are about, we first need an account to fit them into. Such an account needs to be able to deal with the various forms of Frege's puzzle.

This chapter has two parts. §2.1 motivates a solution to Frege's puzzle at the level of individual psychology, rather than at the level of communication or linguistic meaning. I discuss communication-based accounts with specific reference to Sam Cumming's [2007, 2008, forthcoming] account in terms of discourse content, but I argue that consideration of Saul Kripke's [1979] puzzle about belief leads to a dilemma which no communication-based account can solve.

§2.2 presents my preferred psychological solution. I consider views which take belief tokens to have their truth-conditions explained by the language of thought hypothesis defended by Jerry Fodor [1975, 2008]. While Fodor's solution would work, the hypothesis is controversial and substantive and I will try to get by without it. Drawing in particular on work by Kit Fine [2003, 2009], I examine some less controversial folk-psychological assumptions which would be explained by the language of thought hypothesis but do not presuppose it. I consider some objections to Fine's key concept of *co-ordination* based on work by Timothy Williamson and Laura Schroeter, and defend the concept against them. With the concept found to be in good standing, I argue These assumptions are all we need to explain the data thrown up by Frege's and Kripke's puzzles. I give a fairly formal framework embodying these assumptions, which lets us say when someone's beliefs are inconsistent or when their deductions are valid, without invoking the contents of their attitudes. Instead we talk about the potential contents or objects of their attitudes, as constrained by the co-ordination relations between them. Once we have done that, I demonstrate that it is simple to accommodate gappy beliefs, and to allow deductions involving them to be governed by rational norms in the way that deductions involving ordinary beliefs are.

2.1 Frege's puzzle

Frege's puzzle is a family of problems which arise when speakers are unaware that two names, or two classes of occurrences of the same name, refer to the same thing[16]. For example (ignoring the fact that they are fictional), Lois Lane doesn't know that 'Superman' and 'Clark Kent' co-refer. She has conflicting beliefs about him, for example she has a belief that he can fly which she expresses by saying 'Superman can fly', and a belief that he cannot fly which she expresses by saying 'Clark Kent can't fly'. This is true even though she understands both names. The puzzle arises because she seems to have beliefs whose truth conditions logically could not jointly be met, but without being guilty of the kind of irrationality normally associated with (logically) inconsistent beliefs. If she assertively said 'Superman can and cannot fly' she would have made a logical mistake, but since she is acquainted with him in two ways she can have (in a sense) inconsistent beliefs about him without being guilty of this kind of irrationality.

There are two other kinds of puzzle relating to co-referring names, which this chapter will not directly discuss. The first is a puzzle about identity statements, analyticity and informativeness. It seems that 'Superman is Superman' is analytically true and uninformative, while 'Clark Kent is Superman' is synthetic and informative. This is sometimes taken to be an argument against names' meanings being exhausted by their referents. Here is Quine:

> Frege's example of 'Evening Star' and 'Morning Star' and Russell's of 'Scott' and 'the author of *Waverley*', illustrate that

[16] Similar problems can probably arise in the absence of any co-referential linguistic expressions too, assuming that there can be non-linguistic modes of presentation, as there presumably can. I will focus on cases involving language though, both for ease of exposition and to keep close to the existing literature.

> terms can name the same thing but differ in meaning. [Quine 1951/1953: 21]

I won't discuss whether this puzzle about identity statements (and some other statements[17]) is adequate motivation for rejecting Millianism (the view that a name's meaning is just its referent), although for the record my own view is that it is not. You could however try to use the same machinery to explain both the difference in meaning and Lois's rationality, and that would count as a communication-based account of the kind this chapter argues against.

The other category of Frege puzzle concerns attitude ascriptions and Leibniz's law. It is fairly natural, at least in some contexts, to say that Lois believes that Superman can fly but not that Clark Kent can. Since 'Superman' and 'Clark Kent' ordinarily co-refer, this appears to be a failure of Leibniz's law, in that co-referring names in attitude contexts seem not to be substitutable for one another *salva veritate*. This issue can also be used as the basis for an argument against Millianism. I am more sympathetic to this than the argument from informativeness, but this chapter is about attitudes rather than attitude ascriptions, and this issue about ascriptions will be deferred until chapter three.

The parallels between Frege's puzzle and the puzzles of empty names are starkest when we look at the issues about beliefs and rationality. This should become apparent when we consider the problem in some more depth. Lois is not only rational to hold her inconsistent beliefs, but she can also make justified deductions whose rationality cannot be explained by the things the beliefs are about. For example, if she believes that (to speak naturally but perhaps loosely, depending on your view of attitude ascriptions) Superman can fly and Superman

[17] For example 'If Superman flies then Superman flies' seems uninformative, while 'If Superman flies then Clark Kent flies' seems informative.

wears a cape, she can infer that some cape-wearer flies. She could not infer this from beliefs that Superman flies and Clark wears a cape. This means that the rationality or otherwise of her inferences cannot be explained just by the logical relations between the propositional contents of her beliefs, if these contents are individuated in so-called Russellian fashion by the things that the beliefs are about. Both are about Superman. That there is an issue about deduction as well as one about consistency should be no surprise, because (classically) valid arguments are those whose premises are inconsistent with the negations of their conclusions. A solution to Frege's puzzle needs to be able to explain why Lois can make the deductions she can make, and can't make the deductions she can't make.

Gappy beliefs can also be the premises and conclusions of real-life chains of reasoning, and this reasoning is subject to the same rational constraints. If a child believes that Santa is jolly and Santa is fat, she can infer that someone fat is jolly. She cannot make the same inference from the belief that Santa is jolly and the Tooth Fairy is fat. Assuming there is no Tooth Fairy or Santa, these cannot be explained by the things the child's beliefs are about, because there are no people for them to be about. According to David Braun's account of GPs discussed in chapter one, these rational constraints cannot be fully explained by the beliefs' contents either, because he takes the GPs that Santa is fat and that the tooth fairy is fat to be identical. This does not doom his view immediately, because as we saw with Lois and Superman, Frege's puzzle already creates problems for an account which explains the constraints purely in terms of the beliefs' contents, if those contents are individuated by the objects the beliefs are about. One thing to bear in mind is that even if GPs are individuated more finely, however, the logical relations between them might be unsuitable, on the grounds that atomic GPs cannot be true.

The co-reference and empty name cases give rise to similar problems, and it would be good to give a unified solution to them if we can. That is what I will try to do. Another way of putting it is that everyone needs to solve puzzles of co-reference, whatever they think about empty names and gappy beliefs. This means we can use Frege's puzzle to motivate some machinery while remaining neutral on empty names and gappy beliefs. Once we have the machinery, we can show that gappy beliefs fit into it easily, without causing any new problems. The point is that you don't need new machinery like non-existent objects when you have independently motivated machinery that can already solve the problem.

2.11 Descriptivism

To get a feel for how communication-based solutions might work, let's look briefly at descriptivism. Russell [1905] is the *locus classicus* for the view that the meaning of a name is the same as the meaning of some definite description. On that view, co-referring names can then be synonymous with different descriptions which are satisfied by the same object, and different non-referring names can be synonymous with descriptions which are not satisfied by anything. This means that the propositions expressed by Lois's utterances 'Superman can fly' and 'Clark Kent cannot fly' will be different propositions, equivalent to the propositions expressed by sentences involving different descriptions. This way the rational permissibility of her beliefs can be explained by the consistency of their contents. The descriptions are in fact satisfied by the same person, but there is no incoherence in their being satisfied by different people, only one of whom can fly. The extension of this idea to rational constraints on deduction should be simple.

The same idea can used in the empty name case: 'Santa' and 'the Tooth Fairy' are associated with different unsatisfied definite descriptions. This gives 'Santa is fat' and 'the Tooth Fairy is not fat' consistent contents, and the content of ‘someone jolly is fat’ follows from that of ‘Santa is jolly and Santa is fat’ but not that of ‘Santa is jolly and the Tooth fairy is fat’.

Descriptivism has lost a lot of popularity since Russell’s day, in large part due to Kripke [1980]. It is rare to hold the same version of it that Russell held, but the general strategy for solving Frege’s puzzle is simple and attractive. Instead of explaining the constraints on rationality by assigning beliefs contents individuated by the things the beliefs are about (if any), we individuate their contents in some other way. The logical relations of consistency and consequence between these contents can then explain the rational constraints on belief sets and deductions.

It is possible that the view I will end up endorsing could be recast as one in which rational relations between belief tokens could track logical relations between their contents. I don’t think that is the most natural way of putting it and won’t put it that way, but perhaps it could be done. Whether or not it can be done is not the point though, because there is still a substantial difference between communication-based accounts and the one I am putting forward. One way of thinking about communication is that I believe something, express my belief with an assertion, and then you end up believing what I believe. This kind of communication aims to get the hearer to resemble the speaker in some way, and this resemblance can be described as us having beliefs with the same contents. We can also say that assertions have the same contents as the beliefs they are used to transmit. I will argue that no level of content which can play this role in understanding communication can also explain the rational relations between belief tokens. With Russell’s theory no longer in vogue, however, in what follows I will use Sam Cumming’s [2007, 2008, forthcoming] view as a contemporary representative of the communication-based strategy.

2.12 Anaphora and drefs

Cumming [2007: ii] says he is conducting an experiment in 'what happens if you treat names as anaphoric expressions on a par with pronouns'. When people use words like 'she' and 'it' they have to keep track of who and what are being talked about, and what they keep track of will be different in different conversations. The way Cumming thinks about this is to take reference to be mediated in these cases by entities which attach to the objects in question, and which are denoted by the words. These entities are called *discourse referents* (hereafter drefs)[18]. Cumming's idea is to take names as referring via drefs as well, although while the drefs denoted by pronouns will usually (though not always) be confined to one discourse, drefs denoted by names will generally span many discourses. These discourses may be far apart in time, and the participants in one may not even know about the participants in another. Since it would be hard to keep track of a dref across such distances using expressions used to denote as many different drefs as 'it' or 'she', we introduce a name. While indistinguishable names sometimes attach to different drefs, using names seems to combine with context to narrow the options enough that in practice we can keep track in a way that we couldn't if we always used pronouns. In the rest of this section I will give some more detail on how this machinery works, and in the next section I will show how Cumming applies it to Kripke's puzzle, and argue that it is unsatisfactory.

In a textbook case of anaphora, an indefinite expression is used, introducing an object, and then a definite expression is used to refer back to that object:

[18] 'Denote' is Cumming's word for what a word does to a dref, so I will follow him in this. To keep the terminology uniform, I will say a words and speakers denote drefs, drefs attach to objects, and words and speakers refer to objects.

> ‘Wash a bunch of fresh spinach well and then shred it finely. Sauté it in a little butter until it is wilted, drain __, then put a little into each ramekin.’ [Huddleston and Pullum 2002: 1457; their emphasis]

Note that some pronouns in the second sentence have their antecedent in the first. Cumming [2007: §1.2] shows how something similar can happen with names: you start off with an expression like (to use his example) ‘Jessica Rett, a prominent fashion designer’ and subsequently you just say ‘Rett’ or something like that. You introduce her into the discourse, indicating which name you’ll be referring back to her with, and then use the name anaphorically. In subsequent discourses you can still refer back to the original referent, because anaphora can cross discourses. Allowing this is not just an *ad hoc* measure to solve problems about names, because it is useful for ordinary anaphora to cross discourses too:

> We can even imagine an individual being introduced in the first discourse (e.g. a new love-interest) and becoming salient enough to be retrieved by a pronoun at the beginning of the second discourse (`Did he call?'). Now, a natural way of describing this would be to say that the pronoun at the start of the second discourse is anaphoric to some indefinite expression embedded in the first. However, this explanation is impossible if discourse boundaries are impervious to anaphora. [Cumming 2007: 17]

One might be resistant to the idea that this is anaphora at all: if the man is salient enough then maybe you can refer to him anyway, just by his salience. This is may not be a question best settled from the armchair: for example, there might be empirical data showing that people’s responses to names are more like people’s responses to deictic pronouns than to anaphoric pronouns, or the other way around. It does not matter for present purposes whether Cumming’s treatment of names as anaphora is ultimately the best way of carving things up, however, since I am only using it as a worked example to show that communication-based solutions to Frege's puzzle do not work. If treating names as anaphoric is empirically implausible then that is another reason to reject Cumming's specific communication-based solution, but I am interested in showing what is wrong with those types of solution in general. For present purposes we can largely ignore criticisms of Cumming's position which are unrelated to Frege's puzzle.[19]

The application of Cumming's proposal to simple variants of Frege's puzzle is quite straightforward. Words refer to objects by denoting drefs which attach to those objects. When I say 'Phosphorus is big but Hesperus is not', the two names denote distinct drefs which both attach to Venus. We can imagine a conversational scorecard, following David Lewis [1979], showing what the common assumptions of the participants in a conversation are. Drefs will correspond to columns on the scorecard which help us keep track of objects. The conversational

[19] There is at least one advantage of treating the 'did he call?' case and classic anaphora in the same way which can be seen from the armchair. It allows you to treat anaphoric reference back to earlier sentences the same way you treat anaphoric reference back to earlier in the same sentence (as in the spinach example), without having an arbitrary cut-off between different-sentence anaphora and Cumming's case. If it is a stretch to call it anaphora in the latter case, you can preserve what is important in Cumming's proposal by having classic anaphora, names and non-demonstrative discourse-initial personal pronouns all refer via drefs, and allowing drefs to survive across discourses. This should still exclude cases where a personal pronoun is accompanied by a demonstration, as when you point at someone and say 'he's tall'. One way of drawing the distinction is between referents which are eligible because they have already been talked about, so there is a dref to reuse, and referents which are eligible for some other reason (e.g. because you're pointing at them or they just walked in carrying a gun), where there is no dref to reuse and so the pronoun cannot be anaphoric.

scorecard is updated in order to keep track of the conversational common ground; for example, if I say 'Cicero was bald' and everyone accepts this then 'bald' (or baldness) is added to a column of the scorecard corresponding to the dref denoted by 'Cicero'.

There are various ways of thinking about the status of the scorecard, but for present purposes it is easiest to think of there being one objective scorecard for the conversation, and if we have our own personal ones they are just there for keeping track of the communal one. It gives a more realistic picture of communication if the columns correspond to drefs rather than directly to objects, because that allows that not all true identities and their consequences will be trivially part of the common ground: sometimes asserting a true identity can entail a non-trivial update to the scorecard.[20]

This will mean that when I say 'Phosphorus is big but Hesperus is not', the change in the score mandated by the first conjunct need not conflict with that mandated by the second conjunct, as it would be if I said 'Phosphorus is big but Phosphorus is not'. This is because the first conjunct updates the column for the dref denoted by 'Phosphorus' and the second updates the column for the dref denoted by 'Hesperus'. Since the score does not incorporate the fact that 'Phosphorus' and 'Hesperus' denote drefs attaching to the same object, the worldly inconsistency which my utterance introduces to the scorecard does not create any

[20] I have not wanted to get into the metaphysics of drefs because the problems with Cumming's account are supposed to generalize to other communication-based accounts, so changing the metaphysics of drefs will not help. To help get a handle on what they are, one can view them as the same sorts of things as conversational scorecards: presumably abstract and possibly dependent on social practices in the sense defended in Thomasson [1999], although somewhere could probably be found for them in a systematic ontology of abstracta, such as an ontology of impure sets or the ontology of property-encoding objects due to Zalta [1983].

difficulties in updating it. The contradiction in the information represented is not intrinsic to the scorecard, because the contradictory properties are kept in separate columns.

2.13 Paderewski

While Cumming's account works quite straightforwardly for simple cases like that of 'Phosphorus' and 'Hesperus', it will run into trouble if puzzle cases still arise where there is not only sameness of referent but also sameness of dref. The natural candidates for such cases are in Kripke's puzzles about Pierre and London and Peter and Paderewski[21]. I will focus on the latter, since the bilingual issue in the former is an unnecessary complication for present purposes.

In the Paderewski puzzle [Kripke 1979: 449], a man called Peter is familiar with Paderewski as a politician and as a pianist, both under the name 'Paderewski'. He thinks they are two different people, and assents to 'Paderewski has musical talent' when talking about him as a pianist,

[21] The puzzles are laid out and discussed in Kripke [1979], who notes [pp. 448-9] that some similar puzzles appear in Putnam [1975]. Here is one:

> "[S]uppose Oscar is a German-English bilingual. In our view, in his total collection of dialects, the words 'beech' and *Buche* are *exact synonyms*. The normal form descriptions of their meanings would be identical. But he might very well not know that they are synonyms! A speaker can have two synonyms in his vocabulary and not know that they are synonyms!
>
> … Oscar may well believe that *this* is a 'beech' (it has a sign on it that says 'beech'), but not believe or disbelieve that this is a '*Buche*'." [1975: 270]

There may also be examples of non-homophonic names in the same language which are also naturally taken to denote the same dref. A possible candidate is Gillian Russell's example of 'Cassius Clay' and 'Muhammad Ali', for a description of which see footnote 10, above.

but rejects it when talking about him as a politician. Explanations appealing to differences in meaning run into trouble here because 'Paderewski' is the same word with the same meaning whether it is used to talk about Paderewski as pianist or as politician. If the drefs are the same in this case then their being different cannot be doing the work. This might also suggest that differences in dref may not be doing the work in the simple case either, since whatever is causing the trouble with Paderewski might also be able to cause the trouble with Phosphorus and Hesperus. As such, Cumming needs to say that the drefs are different even in the Paderewski case, and that is exactly what he does.

This means he needs a mechanism for generating multiple drefs for the same name (rather than just for indistinguishable names), so he says [§3.4] that when names are taught to people they introduce a new dref, and the new speaker's uses only start denoting the public dref after a while, when they synchronize their usage with that of the community at large.

This mechanism of introducing names with new drefs is partly motivated by a kind of phenomenon Cumming identifies where names are taught to people with indefinite constructions. Indefinite constructions always introduce new drefs.

> (C) Tampa was home to a serial killer named Bobby Joe Long. Long was known as 'the Classified-Ad Rapist'. [2007: 2]

This is somewhat fishy, because it takes a mention of a name and treats it like a use, but the use/mention distinction is not always as clear-cut in practice as Quine might like[22]. However, even if there are problems in

[22] For some examples of the distinction being less clear-cut than Quine might like, see Moore [1986].

other cases, there is a reading of (C) which respects the distinction entirely. The first sentence says something about who the name refers to so that the audience understands the second sentence. It is analogous to saying something like this:

> (D) 'Defenestrate' means to throw something out of a window. Yesterday I defenestrated a television.

In the first sentence of (D) we say something about a word we are about to use so people will understand us when we use it in the second sentence. We can explain (C) the same way. 'Long' in the second sentence of (C) can be anaphoric on uses of the name in previous discourses the speaker has been involved in, and the audience will understand that the dref attaches to a serial killer from Tampa. This model nicely accommodates Cumming's data.

The indefinite construction 'a serial killer named Bobby Joe Long' does introduce a new dref, also attaching to Mr Long. This is only to say something about the name though, and when you go on to use rather than mention the name it denotes the same dref it did in previous discourses. This means we do not have to introduce the name every time and then have new speakers co-ordinate their use with that of others after a while, merging their drefs with those of the community at large. The merging process is covered briefly in Cumming [2007: §3.4], but I don't think he needs it to deal with the case where the name is introduced by apparently mentioning it when describing someone introduced with a specific indefinite construction, as in (C). Instead we have the alternative I sketched, which still treats names as anaphoric. Cumming does however also use the creation/merging mechanism for generating extra drefs to account for Kripke's puzzle. I will argue that it does not ultimately solve the puzzle though, and since we do not need the mechanism to accommodate cases like the 'Bobby Joe Long' example either, we may not need it at all.

On Cumming’s account, with Peter the merging never properly happens. The community at large knows that Paderewski is a pianist and a politician and talks as if he is one person (‘Paderewski’ is not a homonymous stage-name), but Peter is still talking as if there were two. Cumming says that Peter has two drefs attaching to Paderewski which are only denoted in conversations in which he (or someone who defers to him) is a participant, and neither dref is the common-currency one:

> Furthermore, Peter does not possess the common-currency dref that refers to the musician-statesman (call it u_{pad}). Why not? We can narrate the situation as follows. At his first introduction to the name (under its politician guise) he acquired the dref u_{pad1} (remember, from §3.4, that one is always first introduced to a new dref, and only later, by learning the skill of coordination, acquires the one already in currency). The next time he heard the name, rather than connecting it with his old symbol, he forged a new symbol (thus betraying the ability he still lacked) and attached another new dref u_{pad2} to it.
>
> So long as his concept of Paderewski was ‘fractured’ (the terminology is Fine's) in this way, his ability to coordinate with others on the dref u_{pad} was compromised. For an interlocutor could not rely on Peter accessing the same mental symbol on successive uses of the name Paderewski. [Cumming 2007: 80]

Cumming seems to load two things into the example: an approximate symmetry between the strengths Peter’s acquaintances with Paderewski under the two guises, and a weakness in these acquaintances compared to those of ordinary members of the community. With these two features in place, Cumming’s solution looks in reasonable shape. It comes under more strain when we modify the example.

Someone (call her Penelope) might become very acquainted with Paderewski, easily satisfying the dref possession-conditions, but then also hear about him under a new guise, thinking it was someone else. Penelope first comes to know all about Paderewski as a politician, more than most in fact, comfortably acquiring the dref u_{pad}. (You don't have to know that someone plays the piano to understand their name.) Then she hears about him as a musician, and fails to make the connection. Cumming could say that Penelope loses u_{pad} and acquires two more drefs, or keeps u_{pad} and acquires one more. He can't say (as I will) that she either doesn't acquire a new one or acquires the old one a second time, because Penelope is now in a position to generate puzzle cases, and Cumming's explanation of puzzle cases is that there are two drefs attaching to the same object.[23]

Suppose we say Penelope loses or discards u_{pad} and acquires two new drefs. This means that the contents of assertions expressing her long-standing beliefs about Paderewski will have changed, since they no longer make the same demands on how the conversational scorecard is updated. I suppose the problem here is that the reason people lack the mob's drefs and have their personal ones instead is meant to be that they are in some way epistemically removed from Paderewski, but in fact Penelope knows more about Paderewski than most of the mob do, even if we ignore her knowledge about Paderewski as musician.

Coming to believe that there is a distinct pianist called 'Paderewski' does not stop her communicating as before, and effecting the same

[23] Cumming [2007: 80] anticipates in a footnote that he will be accused of false precision for talking about the merging process as if it was determinate which dref we grasped at any given time. I should stress that I'm not making this accusation or exploiting any indeterminacy here. It should become clear in the discussion to follow that the problem lies elsewhere.

changes in the beliefs of others as before[24]. Since she does not lose the ability explained by her grasp of the dref, it is unreasonable to say she loses the dref without offering another explanation of the ability. The acquisition of a personal dref should not explain it, since if it did there would have been no reason to postulate public drefs at all. I am arguing that we should solve the puzzles at the level of something individual rather than something public, and conceding that public drefs are an idle wheel in solving the puzzles would establish that.

Now suppose she keeps u_{pad} and acquires another personal dref. This is perhaps worse, because Penelope could go on to find out a lot about Paderewski as a musician, happily co-ordinating her use with the music buffs she talks to about him, such that she would have acquired u_{pad} if she didn't already have it corresponding to her notion of Paderewski the politician. We could say that the music buffs and the political analysts have different public drefs, but this is hard on people who know about all Paderewski's exploits political and musical, but only talk about his music to the music buffs and his politics to the political analysts.

Perhaps there is a danger that I have convoluted the example too much to make it realistic, and unrealistic cases are spoils to the victor. Given this, we can put the problem in a more general way. We need to decide how strong the possession conditions for a dref are going to be. If we make them stronger, people only superficially acquainted with public objects will have their private drefs, which reduces their usefulness for

[24] Note in particular that people talking to her will usually be able to tell that Penelope is accessing the same mental symbol (to speak in Cumming's framework) as before. Telling which dref is denoted by an occurrence of 'Paderewski' is no harder than telling which dref is denoted by an occurrence of 'he'. Penelope has a problem, but it is a problem about belief which mostly does not generate problems with communication. Note the title of Kripke [1979].

explaining communication and what is common about common knowledge or belief. If we make the conditions weaker, people will be able to acquire them more than once as in Penelope's case, so to describe their beliefs we have to look at their psychology and not at their drefs. I am happy with the latter course, because for the purposes of addressing puzzles like Kripke's I want to classify beliefs at the level of individual psychology and not the level of communication.

This point having been made, it is worth looking at how far it extends to other communication-based classifications of beliefs. Instead of talking about possession-conditions for drefs, we can talk about possession-conditions for concepts, or understanding-conditions for words. The strength/weakness dilemma is still going to arise, at least for most things you might have beliefs about. Exceptions might be *de se* beliefs, beliefs about sense-data, or beliefs involving logical constants, identity and other concepts which are simple enough that to possess them you need to know more or less all there is to know about them, and perhaps also complex concepts built out of these. What matters here is that beliefs about ordinary things are not exceptions. For *de re* beliefs about things like Venus and Paderewski we do not have the luxury of (the relevant kind of) nearly complete grasp of concepts. Instead we have something incomplete which allows the possibility of recognition-failure. This possibility means solving the puzzles at the level of psychology, and that leaves public content up for grabs. Note that beliefs could still be individuated by drefs, but with weaker possession-conditions for drefs and the Frege puzzles solved at the level of psychology. The important thing is just that the level at which successful communication gets people to have beliefs with the same contents cannot be the level at which the logical relations between contents mirror the rational relations between belief tokens.

This, at least, is one way of viewing the puzzles. Frege's puzzle shows that we can fail to recognize objects and form conflicting belief systems about them. One might try to explain this through a difference in words or concepts, but Kripke's puzzle shows that we can also fail to recognize words, concepts, drefs, or whatever other public objects we use to explain how two people can entertain the same thought contents. It is, however, worth considering a little longer exactly whether this is true, and if it is, why it is true and whether things could have been othwerwise. Kaplan [1968] introduces the notion of a *vivid name*:

> The notion of a vivid name is intended to go to the purely internal aspects of individuation. Consider typical cases in which we would be likely to say that Ralph knows x or is acquainted with x. Then look only at the conglomeration of images, names, and partial descriptions which Ralph employs to bring x before his mind. Such a conglomeration, when suitably arranged and regimented, is what I call a vivid name. [Kaplan 1968: 201]

Vivid names play the role which drefs play in the Frege puzzles, in that the puzzles arise from a person having two vivid names for the same thing:

> We can easily form two vivid names, one describing Bertrand Russell as logician, and another describing Russell as social critic, which are such that the identity sentence simply can not be decided on internal evidence. In the case of the morning star and the evening star, we can even form names which allow us to locate the purported objects (if we are willing to wait for the propitious moment) without the identity sentence being determinate. Of course Ralph may believe the negation of the identity sentence for all distinct pairs of vivid names, but such beliefs may simply be wrong. And the names can remain vivid even after such inaccurate non-identities are excised. It may happen that Ralph comes to change his beliefs so that where he once believed a non-identity between vivid names, he now

> believes an identity. And at some intermediate stage of wonder he believes neither the identity nor the non-identity. Such Monte Cristo cases may be rare in reality (though rife in fiction), but they are nevertheless clearly possible. They could be ruled out only by demanding an unreasonably high standard of vividness, to wit: no gaps, or else by adding an artificial and ad hoc requirement that all vivid names contain certain format items, e.g. exact place and date of birth. Either course would put us out of *rapport* with most of our closest friends. Thus, two vivid names can represent the same person to Ralph although Ralph does not believe the identity sentence. [Kaplan 1968: 205]

Vivid names are not public objects, but considering Kaplan's discussion can help us decide whether any public objects could play the role of concepts/drefs in explaining communication without allowing recognition-failure. There are two points to address here which Kaplan raises: the 'no gaps' condition and the 'format items' condition. The corresponding issues for concept/dref possession conditions are whether you need to know everything about the things your concepts are of, and whether concepts include specific uniquely identifying things you need to know to grasp the concept.

The first presumably cannot be met: we can talk about things and people we do not know everything about. The format items are more promising though, in the case of recognising particular concepts. While it would be *ad hoc* to say that there was some particular identifying format item which every concept had to incorporate to be a concept of a thing, it is less *ad hoc* to say that each concept has a format item which you must know in order to grasp the concept. (Different items for different concepts.) These concepts still need not be equivalent to those expressed by descriptions or rigidified descriptions, because while knowledge[25] of the format item could be a necessary condition for possession of the concept, there could also be other necessary conditions to do with causal acquaintance with the object, involvement in a linguistic practice with other users of the concept, or something else along those lines.

Since different concepts could have different format items, one might still be able to think of the same object under two such concepts. For example, if one concept required that the thinker know its object was the first heavenly body seen in the evening at such and such time in the astronomical cycle, and another concept required that the thinker know its object was the last one seen in the morning at such and such a time, these would satisfy the uniqueness condition but could still be of the same object without the thinker knowing. Some pairs of concepts will rule out their being of the same object though: the format item of a concept of Adam could entail being the first human and the term for a concept of Eve could entail being the second. These exclude each other. Concepts like this might not be susceptible to Kripke's puzzle. If there was a concept of Paderewski which you couldn't grasp unless you knew it was a concept of a person born at such and such an exact place and time, maybe you couldn't grasp that concept twice without either recognizing it or being subject to rational criticism.

This particular case assumes two people cannot be born at exactly the same place and time, and that the concepts of the places and times are themselves not subject to recognition failure. In general, the picture

[25] I have talked about knowledge here rather than belief, although perhaps there could be concepts of objects which required that thinkers believe the object satisfied the format item but not that they know it, or which even required that thinkers have a false belief about the object or at least an untrue gappy belief in the case of empty vivid names. Whether or not you find that picture appealing, similar arguments should still go through for concepts requiring mere belief.

needs the format items to be exclusive in the sense that no two objects could have the same item, and transparent in the sense that there could not be recognition-failure for the items themselves. These assumptions are substantive but I will not challenge them now. The second is probably more problematic than the first, since the format items could have a uniqueness condition built in.

Instead what we should say is that while the possibility (if it is a possibility) of such concepts is interesting, it is not an accurate description of the concepts/drefs which we actually use to communicate with one another, at least in many cases which generate puzzles. If there were cases where it was, then maybe a communication-based solution to Frege's puzzles might be appropriate for those cases, because the intrapersonal puzzles would not arise. The picture is probably more plausible for mathematical concepts than for concepts of concrete things. Consider whether Pierre could have got confused between 'nine' and 'neuf' without some kind of failure of rationality. Kripke [2008: 187] takes the view that 'nine' has a *revelatory sense*, which means 'one can figure out from the sense alone what the referent is'. It is not obvious whether the concepts under discussion are all and only the revelatory senses, but the two ideas are close. I expect there is room for somebody suitably sceptical about *a priori* knowledge and the analytic/synthetic distinction to resist the view that 'nine' has a revelatory sense, but if there are concepts for which Kripke's puzzle could not arise then mathematical concepts are promising candidates.

The issue is with the concepts though, rather than their objects, because the puzzles could presumably arise between two different concepts of the same number. Even setting aside the question of whether mathematical identities like '$e^{i\pi} = -1$' can be genuinely informative, not all concepts of numbers are mathematical concepts. One could have a concept of the Number of the Beast without knowing it was 666. I probably have a concept of Graham's Number although I have forgotten how to construct it, and perhaps never knew which number it is. (It's big.) The puzzles can obviously arise for descriptive concepts like that expressed by 'the number of the planets', but 'the Number of the Beast' and 'Graham's Number' do not clearly express descriptive concepts. They look like definite descriptions, but they could still be rigid designators, like 'The Statue of Liberty'[26]. Another example of a non-revelatory sense denoting a number might be that expressed by 'π', given the amount of effort which went into working out which number it was.

The fact that most of our concepts for concrete things are not transparent in this way is a non-trivial fact about communication and thought, but it seems to be supported by experience. If it wasn't, then the fictions and thought experiments in which cases like this are so rife would be psychologically implausible, rather than just historically so. If ways of thinking about things corresponded 1-1 with the meanings of words for them, that would be interesting and we would be missing out on some generalizations if we never classified singular thoughts according to the word-meanings they corresponded with. That is not how things are though, and if it was then communication would be different and perhaps harder. Kripke's puzzles arise from this failure of thought and communication to match. If we want to explain what Pierre, Peter and Penelope aren't getting wrong (it's obvious what they *are* getting wrong) then we have to look beyond communication.[27]

[26] Cumming [2007: §1.1] has an interesting discussion of the forms proper names can take, including examples which look like definite descriptions.

[27]Laurence Goldstein [2009] claims that the lesson of Kripke's puzzle is that we should be more careful about forming beliefs. We shouldn't, for example, form the belief that London is pretty on the basis of the way a few parts of it look, even if we are monolingual Londoners. This may be another lesson of the puzzle, but its being such doesn't conflict with anything I've said. It is sufficient for my purposes that Peter not be subject to rational criticism for having two beliefs of significantly different strengths that Paderewski is e.g. musical, even

2.2 A psychological solution to Frege's puzzle

2.21 The language of thought hypothesis

The easiest thing would be if there was something psychological which was structurally just like a language, with beliefs being like sentences, constructed out of things like names and predicates, and which allowed distinct names to sometimes co-refer. We could say that Pierre's beliefs were consistent in that there was an interpretation (in some sensible class e.g. first-order classical interpretations) on which they were all true, although they were inconsistent in that on the intended interpretation of the psychological things corresponding to names in a language the beliefs could not all be true[28]. The reason we do not fault Pierre's logical acumen is that his set of belief tokens is true on some sensible interpretation, and that is all logic demands.

This picture is the one put forward by Mark Crimmins and John Perry [1989] and by Jerry Fodor [2008: ch. 3]. On the Crimmins-Perry picture, beliefs are constituted by *ideas*, which correspond to predicates, and

though he uses the same dref to communicate these beliefs. If my purposes are not met in this regard, then Kripke's puzzle is not the argument for epistemic modesty which Goldstein takes it to be, so much as an argument for scepticism, and we have enough of those already.

[28] There is a small complication here when dealing with Frege puzzles involving predicates. We want to say that 'Cicero is a doctor' and 'Cicero is a physician' are inconsistent in the same way as 'Cicero is an orator' and 'Tully is an orator', but the definition we gave only looked at the intended interpretations of the names, not the predicates. To avoid problems of contingently co-extensive predicates, it might be best to extend the definition to doctor/physician cases by using a logic assigning properties to predicates and extensions to properties. Worldly consistency of beliefs would then be truth on some model which matched the intended model for assigning objects to names and properties to predicates.

notions, which correspond to names. Peter has two notions of Paderewski, one involved in beliefs involving his idea of being a pianist, and one involved in beliefs involving his idea of being a politician. It is fairly clear how this will go, and as far as I tell, it works.

Fodor already has independent reasons [see especially Fodor 1975, 2008] for thinking that propositional attitudes involve sentences in a language of thought, and he enlists this picture in solving Frege's puzzle. If you are already committed to a language of thought then perhaps you might as well use it to solve the puzzle if it works, and even if you are not committed to it for other reasons, its usefulness in solving them is a *prima facie* argument in its favour.

It should also be noted that the solution to Frege's puzzle offered by this picture can easily be extended to a treatment of beliefs corresponding to empty names, explaining why it could be rational to think that Vulcan is bigger than Santa, but not to think that Santa is bigger than Santa. We allow the constituents of beliefs corresponding to names to lack objects, and then treat them exactly as empty names were treated in chapter one. Essentially, if beliefs are like sentences then a treatment of names without referents can be used as a treatment of gappy beliefs, and we've got one of those already. Problem solved.

The snag is that the language of thought hypothesis is controversial and substantive. It is not the sort of hypothesis you can believe just to solve a puzzle, and if the brain does not work that way then we need another story. If the brain may or may not work that way, then we are still leaving a hostage to biology, and if it even might not have worked that way then we are leaving hostages to multiple realizability. Co-referring names in the language of thought cannot be the solution to the problem unless there is a language of thought with names in it. There are two obstinate strategies one could attempt. First, one could provide arguments for the language of thought hypothesis being the only plausible story about how the brain works, and as such dismiss the possibility that psychology does not provide us with suitable sentence-like structures. This may be appropriate within Fodor's project, but it would be good to have something with more general appeal.

You could instead say that beliefs being sentence-like is a part of folk psychology, so whatever the brain is like, if it realizes folk psychology then it realizes ideas and notions, so there is no hostage to empirical fortune. That is not clearly a hopeless strategy, but it is not quite mine. I will make some more minimal claims about folk psychology, which could be explained by the language of thought hypothesis' truth, but do not entail it. Showing that only the language of thought could explain it is a job for what Fodor [1975: preface] unabusively calls 'speculative psychology'. I have no real problem with speculative psychology, but I am not doing it here, and more crucially, I am not relying on any particular speculative psychological position. The picture I will put forward keeps what you need to solve Frege's and Kripke's puzzles and accommodate gappy beliefs but could still be true even if Fodor and his allies are wrong about how the brain works.

• 2.22 Dispensing with psychological sententialism

Beliefs have truth conditions. I do not mean this in any controversial sense. I have a belief token that Obama is American, and this belief is in some sense true iff Obama is American. If Obama is American then the belief state I'm in represents the world right, and if he is not or does not exist then the state represents the world wrong. Other things being equal, I try to be in belief states which represent the world right. Gappy beliefs can represent the world wrong too: a child's belief that Santa is jolly represents it wrong. There is room for gappy beliefs which are not existentially committing though, and these might represent the world right. An example might be a child's belief that Santa didn't write *Alice* *in Wonderland*. The situation is analogous to that with assertions containing empty names discussed in the last chapter.

Having truth-conditions in this hopefully uncontroversial sense is a feature beliefs share with utterances. With respect to truth-conditions, my belief that Obama is American is like an assertive utterance (under normal conditions) of 'Obama is American'. The utterance's truth conditions are explained by its syntactic structure. It is of subject-predicate form, and the subject refers to Obama and the predicate to being American, and these three facts combine to mean that it has the truth conditions it has. My belief is a more mysterious thing. We do not have such a straightforward story about why it has the truth conditions it has. It stands in some causal relations to Obama and America, but it is far from obvious that it has a constituent standing in the appropriate relations to each.

The simple and controversial explanation takes beliefs to be syntactically structured and have truth conditions explained in a manner analogous to those of utterances[29]. Following Mark Richard [1990], let's call this view *psychological sententialism* (hereafter PS). Will PS be true of all beliefs? It would have to be true of all singular beliefs to provide a general solution to Frege's puzzle, and there is

[29] I have been talking deliberately about utterances rather than sentences. Utterances are concrete tokens, and since I am talking about beliefs as concrete tokens, it makes sense to treat them as analogous to utterances rather than to sentences. Another benefit is that we do not need an account of the ontology of sentences here, which would experience similar problems to an account of the ontology of words, which are discussed in Kaplan [1990]. The ontology of utterances and beliefs seems more straightforward: utterances are actions and beliefs are states. There are general questions about the ontology of actions and states, but it is better to piggyback on the solution to a general problem than have to deal with the specific problem of what sentences are.

reason to think that it is not. One set of problems is all the reasons (except the bad ones) people argue against the language of thought hypothesis, but another problem is that sententialism is, as I will argue, not even true of all utterances. Some utterances have truth conditions which are not explained by their syntactic structure. Since language is supposed to be the paradigm, and sententialism is not even exceptionlessly true of language, this undermines the idea that it is exceptionlessly true of belief.

Richard [1990: 35] gives an example from George Pitcher [1964: 12] of a language in which “catamat” means that the cat is on the mat. Richard rightly points out that the example is underdeveloped, but it is to be expected that some words will have the same conventional meanings as syntactically structured assertions, but without having any discernible syntactic structure.

I do not want to lose half my audience by talking about language games, but consider some children playing on the street who warn each other of a car coming by just saying ‘car’ instead of saying ‘there’s a car coming’. ‘Car’ does not look syntactically structured, but maybe it is. Perhaps we could say that the context supplies the predicate and the word supplies the subject, fitting the assertion into a compositional semantics that way. If in that situation someone says ‘bus’, ‘polar bear’ or ‘Mr Wilkins’, that might well mean (examples of) those things were coming. To account for the fact (if it is a fact) that if they said ‘Antarctica’ it would not mean that Antarctica was coming, we can say the choice of noun phrase shifts the context to one that does not supply that predicate. The phenomenon of utterances shifting their own contexts to accommodate a charitable interpretation is familiar[30]. This sort of thing may have some mileage in it, but the point is that while language in fact has these systematic features for the most part, even in

[30] See Lewis [1979: 346-7] on rules of accommodation.

idiomatic constructions, they do not have to be everywhere. Some utterances could have rules of their own, communicating something complicated without having complicated structure. 'Fire!' may be a plausible example. Davidson's [1965] learnability argument for compositionality does not require compositionality to be exceptionless, since adding an unstructured idiom to a language does not make it unlearnable. Adding infinitely many would have this effect if the learnability argument is sound, but all I am suggesting is that linguistic sententialism is not exceptionless. It doesn't look exceptionless, and the arguments that linguistic sententialism is the rule do not show that it is an exceptionless rule, or even place a finite bound on the number of exceptions.

Similarly, beliefs would not have to all be syntactically structured to play the causal roles which make the beliefs' presences felt. Or if they would, this is not obvious. It is not obvious that beliefs would have to be syntactically structured in the normal case, and even less obvious that there could not be exceptions to the rule, even if PS was the rule. We shouldn't believe PS unless we have to, but maybe it seems that to solve the Frege puzzles we do. That would be a shame though, and I will argue that it is not the case: we can solve the Frege puzzles by making a weaker set of assumptions which would be unified by PS but which do not entail it.

Fine [2007: ch.3 esp. §§A, B] also criticises PS, not by claiming it to be incoherent or unable to account for tacit beliefs, but by saying that it is not clear that it must be true, and that it does not appear to be true in fact. There seem to be exceptions to it. He says:

> But suppose now that a thought signifies a proposition containing two occurrences of a given object, say the proposition that this man is the same as that man (it is better for the purposes of the example if the thought is not expressed in

> words, but is a "felt" identity). Then it is not clear that there must be two components of the thought, each responsible for putting its occurrence of the object into the proposition. Thoughts do not appear to have the same kind of clear syntax as sentences.
>
> This then creates a difficulty if we want to talk of co-ordination within a thought. For between what do we co-ordinate? What I would like to suggest is that it may still be correct to talk of a thought being of an object in a given occurrence or position in such cases, even though there may be no corresponding constituent of the thought. [Fine 2007: 73]

The talk of co-ordination relates to the idea Fine pushes throughout the book, which is (roughly: I'm not trying to sum up the book in a few lines) that semantic facts are not just about what the words mean, but about which occurrences are co-ordinated with which. The difference between 'Cicero is Tully' and 'Cicero is Cicero' is not a difference in the meanings of 'Cicero' and 'Tully', since there may be none. It is a difference in whether or not the subject and complement are co-ordinated. In 'Cicero is Cicero' they are; in 'Cicero is Tully' they aren't.

Yagisawa [1993] disagrees with this. He thinks that the semantic facts about a name are just that it refers immediately (as in not mediated by anything), and that's all. The fact that the subject and complement of 'Cicero is Cicero' co-refer is therefore not, for Yagisawa, a semantic fact. This isn't a crazy thing to think, especially in the light of Kripke's puzzle, since some utterances of 'Cicero is Cicero' might not have the subject and complement co-ordinated. One reason for disagreeing with Yagisawa is that one could say the same about 'Cicero' and 'his' in 'Cicero washed his socks'. If we allow co-ordination arising from anaphora, why not allow it with names? The framework of drefs can deal with both. However, Yagisawa could say about anaphoric pronouns what he says about names. There is a risk of a standoff here.

I think the best defence of co-ordination in language against Yagisawa goes in three steps. First we concede that it might be a widening of what counts as semantics to say that co-ordination is a part of it, but if co-ordination happens then that is enough: it doesn't matter whether we call it semantics or not. Second we say that refusing to theorize about the co-ordination of anaphoric pronouns misses out on connections which are there to be captured. We can say illuminating things by tying anaphoric reference to the drefs which relate speech acts to the dynamics of a discourse's shared assumptions, so we should. Third, now that we have widened our theorizing to include co-ordination, why not apply it to names? To understand this sentence you need to know which pronouns are co-ordinated with which:

> "John met his brother at five, but he had been waiting since four, and since he wouldn't admit his mistake he is now refusing to speak to him."

Exactly the same thing can happen with names though, it is just as much in need of explanation, and the same explanation will do. Quoting Kaplan:

> "My mother's primary care physician is Dr. Shapiro. He referred her to a specialist, another 'Dr. Shapiro' as it happened. My mother reported her gratitude to Dr. Shapiro for sending her to Dr. Shapiro and compared Dr. Shapiro's virtues to those of Dr. Shapiro in a blithe piece of discourse, clearly oblivious to be homonymy. I was racing to keep up (which I was strangely able to do). But from her point of view, she was quite properly using two different words to refer to two different people. Why should there be a problem?" [Kaplan 1990: 108]

Maybe you are tempted to think that Kaplan could follow what his mother said because he is so clever, and that it isn't a normal thing so we don't need to theorize about it. You do not have to be as clever as Kaplan to do it, though. In the sitcom *Frasier* it is a common comical device for Daphne to do the same thing, referring just as blithely to two characters as 'Dr Crane'. The viewer follows it but sometimes her interlocutor doesn't, with hilarious consequences. Note that one could pick up on the co-ordination relations even if one didn't know who either Dr Shapiro/Crane was, which shows that there is a fact separate from the references to pick up on. Fine thinks it is a semantic fact, but the important thing is that it is a fact.

Fine argues that relations of co-ordination obtain not just in linguistic communication, but also in individual thought, and between speakers and thinkers. Here we can exploit his idea of intrapersonal co-ordination of beliefs.[31] The last sentence of the passage quoted earlier is important:

> What I would like to suggest is that it may still be correct to talk of a thought being of an object in a given occurrence or position in such cases, even though there may be no corresponding constituent of the thought.

Folk psychology may not demand that PS be true, but it does demand that there be a difference between a belief that Helena loves Demetrius and a belief that Demetrius loves Helena. That difference might be explained by the beliefs' syntactic structures and it might not, but whatever the difference is, that is what thoughts being of objects in given occurrences or positions amounts to.

Now, what about co-ordination between different occurrences of objects of thought? This is the sort of thing which means that some

[31] It is Fine's idea so I don't take credit for the insight, but I don't want to attribute the way I'm explaining it to Fine since my purposes are not exegetical and he might not endorse my applications of it.

beliefs have to have the same object appearing in more than one position, such as a co-ordinated belief that Cicero likes Cicero, whereas some beliefs do have the same object but need not have, such as an unco-ordinated belief that Cicero likes Tully. It is also the sort of thing which is tied to rational requirements of consistency: Peter is not irrational to have his beliefs that Paderewski is and is not musical, because these beliefs are not co-ordinated with each other. I would be irrational to have contradictory beliefs about Paderewski, because all my beliefs about Paderewski are co-ordinated (I hope). If PS is true, this co-ordination might be due to belief tokens sharing constituents, like ideas and notions. Folk psychology might not say how the co-ordination happens, but it certainly says it happens. If it didn't say this, then Kripke's stories would sound stranger than they do.

It may be that folk psychology is bunk, or at least the fragment under consideration about singular thought is bunk. In that case Kripke's puzzle won't arise, because it is a puzzle about singular thought. You could deal with it as Lewis [1981] does by dropping the externalist typing of beliefs, or you could be entirely eliminativist about beliefs, in which case none of these problems about rationality and propositional attitudes arise. There are no contradictions or puzzles of this kind about neurons firing[32]. But if folk psychology is not bunk, and we do have singular thoughts, then their objects occur in various positions, and

[32] This seems to have some truth in it, but it may be worth questioning why the puzzles arise when we describe the beliefs in psychological or intentional terms but not in neural terms. Perhaps the idea is that the descriptions giving rise to puzzles are normatively loaded: we want to say what Pierre etc are doing right and what they are doing wrong. If that is true, eliminativism would only dissolve the puzzles if it eliminated the intentional descriptions in favour of something non-normative, but if it did that then presumably something would be lost. How much of a problem that would be would presumably depend on the kind of eliminativist you were.

they are co-ordinated such that some of the positions of some of a subject's beliefs have to be directed at the same object.

- ### 2.23 Problems with co-ordination

So far I have taken it for granted that there is a kind of irrationality such that Pierre isn't irrational to believe that *Londres* is pretty and London is not, but would be irrational to believe that London is and isn't pretty. The requirements not to be irrational in this way are closely related to co-ordination relations between thoughts. Now I'll pause to question whether the phenomenon under discussion really is a kind of irrationality at all, and if there is even a phenomenon in good standing there at all. Worries come from two places.

First, we have the anti-luminosity argument due to Williamson [2000: 93-113; 2008], which is meant to show that few if any non-trivial states will be such that if a subject is in the state they will always be in a position to know they are in the state. This is because most non-trivial states will have fuzzy boundaries, and if you are in the state but near the boundary, you will be close enough to not being in the state that you can't tell whether you are in it or not. If there are borderline cases of co-ordination, perhaps people are not always in a position to know which co-ordination relations obtain between their thoughts. Perhaps it is not irrational not to take co-ordination relations into account if you don't know they obtain; put another way, perhaps there can't be rational norms applying to people in virtue of cognitive states they don't know they are in.

Second, we have the possibility that co-ordination might not entail co-reference. Laura Schroeter [2007] discusses a 'slow switching' version of the Twin Earth thought experiment from Putnam [1975], where it seems that beliefs involving the same mental file (or however else we understand co-ordination) might not all be about the same object. If it is possible to have co-ordination without co-reference, it seems strange to say it is irrational to have differences in credence which take these possibilities into account, or perhaps even co-ordinated beliefs which jointly entail that they are about different things. These kinds of rational norms are supposed to rule out belief sets which cannot be all true, but if co-ordination does not entail co-reference, the norms would rule out belief sets which could be all true, and perhaps some belief sets which actually are all true.

I will deal with the two objections in different ways. In §2.321 I will argue that while we might have co-ordinated beliefs without knowing it, this does not undermine the rational norms governing them, since the same is true of beliefs themselves, which are uncontroversially governed by rational norms. In §2.322 I will argue that co-ordination must in fact entail co-reference (where there is reference at all), and that denying this risks leading to a vicious regress.

- *2.231 Anti-luminosity*

The anti-luminosity argument can only apply to co-ordination if there are apparent borderline cases. There are at least three types of possible example. The simplest of these is to try constructing an example by brute force, taking a clear case of co-ordination and transforming them molecule by molecule into a clear non-case of co-ordination. Somewhere in the middle of the series there may well be borderline cases.

A more psychologically interesting case is where two belief sets merge gradually, as the subject gets more confident of an identity belief, eventually just having one body of belief which includes the information that the object in question has two names. This is probably the process I went through with ‘maize’ and ‘corn’, and perhaps people went through the process with ‘Phosphorus’ and ‘Hesperus’ as the astronomical consensus was forming. It seems reasonable to say that there could be borderline cases in the middle of this process.

A third case relates to analyticity. Some true identity statements seem to be candidates for analyticity, not in Gillian Russell's sense of being true in virtue of reference determiner, but in that understanding them requires knowing they are true. 'Jacko is Michael Jackson' or 'John F. Kennedy is JFK' might be examples. 'The KLF are the Timelords', on the other hand, is definitely not analytic, although it is true. It is hard to pin down the exact difference; perhaps it is something like the difference between a variant and an alias. Is there any principled reason why there could not be analytic connections between 'John F. Kennedy' and 'JFK' if there can be such connections between 'bachelor' and 'married', as there paradigmatically can? Fine says this:

> 'One need not be a Quinean sceptic about the analytic/synthetic distinction to believe that the distinction has no clear application in the case of names.' [Fine 2007: 84-5]

He says this because even if there is identifying information associated with a name there would be no tenable distinction between the information which is constitutive of the name's meaning and that which is not. We could take this either as saying that the distinction is fuzzy, or as saying there is no distinction to be made. If there is a fuzzy distinction, anti-luminosity could set in. If there is no distinction, then at least this conception of analyticity will not apply to names. If there is, we have another possible kind of borderline case of co-ordination.

There is however a heavy to responding to possible borderline cases by saying the idea of co-ordination isn't in good standing and we shouldn't theorize about it. Borderline cases can be constructed for other mental states like belief and intention, and while their existence might affect our theorizing (e.g. borderline cases of belief might lead us to reject closure of implicit belief under multi-premise entailment), they don't make theorizing impossible. This should not be a problem once we are aware of it, though. It will however turn out to be helpful to introduce a concept to describe cases near the borderline which are not cases of co-ordination.

Where borderline cases are not quite co-ordination, they will tend to be cases of what we can call *pseudo-coordination*. We can say x and y are pseudo-ordinated iff they are not co-ordinated with each other, but they are co-ordinated with the two sides of an identity belief the subject has. So if Smith's belief that Phosphorus is big and Hesperus is round are not co-ordinated, they can still be pseudo-coordinated if Smith believes that Phosphorus is Hesperus, and the two Phosphorus positions are co-ordinated, as are the two Hesperus positions. Note that we cannot get by with just pseudo-coordination, because it is defined in terms of coordination. This will be important in the next section.

- *2.232 Slow switching*

A second kind of case threatens to undermine the notion of co-ordination. These are called 'slow switching' cases, and happen when the intentional object of a word or set of beliefs seems to switch slowly after the primary source of information relating to the word or beliefs shifts from one thing to another. The classic example of a word shifting its reference is 'Madagascar', discussed by Kripke [1980: 163] and Gareth Evans [1973: §3]. In that example something like the following happened: Marco Polo started using the name when he wanted to talk about the island, following the locals' use of the name. Actually the locals were using the name to refer to somewhere else on the mainland though, so maybe Polo's use, intended to follow the locals' usage, referred to that other place instead, at least at first. But by now, of course, the name refers to the island.

Laura Schroeter [2007] gives a slow switching Twin Earth example which is meant to be a slow-switching case of thought, but where some of the beliefs don't switch:

> Many years after his [unwitting] switch to Twin Earth, Peter is on vacation with his twin-family and he begins reminiscing about his childhood vacations at the ocean. He and his sister Jo, Peter recalls, used to love playing in the water. Glancing out the window, he notices the fastidious Jo who's unwilling to venture in the water despite general coaxing. Peter is suddenly struck with the juxtaposition of these two thoughts – his memory and his perceptual belief – and he begins to wonder how Jo could have changed so much. What exactly are the reference and truth-conditions for Peter's thoughts in this train of reasoning? Is he thinking about the Earthly things of his childhood or the Twin-Earthly things of his current environment, or both?
>
> The most natural answer, I submit, is that Peter's thoughts refer to different things. Peter's childhood memory is true just in case his biological sister Jo_1 liked playing in H_2O; his perceptual belief is true just in case her counterpart Jo_2 dislikes playing in XYZ. [Schroeter 2007: 606]

Schroeter's assessment of the case presents a direct challenge to the idea of co-ordination, and it does have some intuitive pull. The belief that Jo_1 loved playing in the H_2O was formed years ago, and has just been sitting there waiting to be recalled. Why should the move to Twin Earth suddenly make it a belief that Jo_2 loved playing in the XYZ? Peter never saw Jo_2 playing in the XYZ. On the other hand, the new beliefs must be about Jo_2 and XYZ, if we accept that slow switching can happen when you move to Twin Earth, just as it happened with 'Madagascar'.

Acting against these intuitions, we have co-ordination. If co-ordination is to play the role in justifying reasoning that it is meant to play, it will have to work at least almost all of the time. We have a conflict here between two forces determining what beliefs are about. One is that beliefs tend to be about the causal sources of the information which led the subject to form them. The other force is that co-ordinated beliefs tend to be about the same things. Which force wins?

You could say that co-ordination always wins, because it isn't really a force: the forces determine which thing a set of co-ordinated object positions will be about. There are forces pulling in both directions, but since the switching is deemed to have happened in some cases, it must have happened in all of them. While Peter's memory hasn't been accessed since before the move to Twin Earth, it got dragged with the rest when the switch happened. This is what I'm inclined to say.

The alternative is to say that co-ordination is a strong but defeasible force, and in Peter's case it is defeated. While I acknowledge this position's intuitive pull, it leads to a difficulty. If co-ordination is defeasible, it seems that rationally we should consider each case of co-ordination on its merits. Suppose I have co-ordinated beliefs that Venus is big and Venus is round, and I wonder how confident I should be that something is big and round. This is affected by my confidence that the object of one belief is the object of the other. But then we have something like pseudo-coordination except with no real co-ordination anywhere. If there is no co-ordination, how does this identity belief (that the objects of the beliefs are the same) get to be more infallibly connected to the beliefs about Venus than they were to each other? I don't say this difficulty can't be solved, but it's the reason I'm inclined to say co-ordination isn't defeasible.[33]

[33] A possible alternative is to say that while we know that co-ordination is defeasible, when dealing with our own thoughts it is always rational to treat particular cases as indefeasible, as part of some kind of anti-sceptical strategy. I won't pursue this here.

If we take this hard line, it would be good to say something to recognize the intuitions against it. One option is to use pseudo-coordination again. While our longstanding beliefs are co-ordinated, allowing us to integrate them into patterns of reasoning, when new information comes in it does so via unco-ordinated beliefs, often involving perceptual modes of presentation. These beliefs only get pseudo-coordinated with the longstanding beliefs when we identify something in front of us as something we have longstanding beliefs about.[34] So while Peter has co-ordinated beliefs about Jo_2 and XYZ, as he must to be able to use them both in his reasoning, he also has a pseudo-coordinated perceptual belief about Jo_2 and XYZ, and possibly also a quasi-perceptual one from his episodic memory about Jo_1 and H_2O.

2.24 A more formal framework for co-ordination

I have been arguing informally for some theses about folk psychology's treatment of singular thought. These theses are meant to be more innocuous than the language of thought hypothesis, although if something like that hypothesis is true then this would explain how our psychology got to have these features. Now we can put the features together into a more formal framework. First we'll recap the features, and then we'll present the formal framework.

- People have beliefs.
- Beliefs have truth conditions.
- Beliefs can be about things, and these things are involved in their truth conditions.

[34] This kind of picture is influenced by the one in Chalmers [2003], although he uses the temporary modes of presentation as ways of thinking about phenomenal properties, which we have while we are experiencing them, in virtue of experiencing them.

- The relation between beliefs and the things they are about has enough structure that the belief that Helena loves Demetrius is different from the belief that Demetrius loves Helena. This is what we mean by a belief being about an object in a certain position.
- Sometimes there are co-ordination relations between the objects of beliefs in certain positions, such that the *n*th object of B_x must be the *m*th object of B_y, where x and y may or may not be the same.

It is helpful to think of co-ordination as what Fine calls *strict co-reference*. A pair of beliefs can be intrinsically such that there has to be co-reference, if there is reference at all. This co-reference, if they refer at all, is *representationally required*, in Fine's terms. Schroeter [2007: 600] talks about *de jure* co-reference, distinguishing it from *de facto* co-reference where beliefs are not co-ordinated but co-refer anyway. This lets us think about co-ordination as corresponding to a kind of representational necessity, which we can represent with a box: $\Box_R$. $\Box_R\varphi$ will be true iff φ is true at all the representational possibilities: all the ways things could be while respecting the co-ordination relations, and whatever other representational requirements there are. A sentence φ is true iff true at the actual representational possibility, where beliefs are about what they are actually about. So $\Box_R$ obeys the modal axiom T: $\Box_R\varphi \rightarrow \varphi$.

When I say people have beliefs, I mean beliefs as particular states. No matter how similar your belief that Venus is round is to mine, they are different beliefs. Beliefs are states, and perhaps there is room for ontological qualms about referring to states, although it is fairly commonplace to refer to events. Nonetheless, the formalism will involve reference to beliefs, in order to say that they stand in relations to their objects.

We want to say that beliefs have objects. These can be either objects or properties and relations. I will represent the properties as their extensions, although we could have a *sui generis* domain of properties instead and talk about instantiation instead of membership. The use of extensions instead of properties is supposed to keep nominalists happy, since they will be used to paraphrasing set-talk. Beliefs have objects in different positions, and at most one object per position, so what we need is a function from <belief, number> pairs. We will have two functions, one for the objects and one for the properties and relations: O(B, n) is the thing B is about in nth object position, and P(B, n) is the property or relation B is about in nth property position. These functions will be partial, to allow for gappy beliefs. Since beliefs are not in general representationally required to have the objects they actually have, these functions can take different values at different worlds. However, if a belief is representationally required to take a particular object, perhaps if it is a mathematical or identity belief, we can express it as in these examples:

$$\Box_R[O(B, 1) = \pi]$$
$$\Box_R[P(B, 2) = \{<x, y>: x=y\}]$$

Beliefs have truth conditions, which we can represent in terms of the functions from beliefs to their objects. Suppose I have a belief *B* that Helena loves Demetrius. *B* has two object positions and one relation position, and is true iff the first object stands in the relation to the second object. We express this thus, where $\equiv_R$ stands for necessarily$_R$ iff:

$$T(B) \equiv_R <O(B, 1), O(B, 2)> \in P(B, 1)$$

In fact O(*B*, 1) is Helena, O(*B*, 2) is Demetrius, and P(*B*, 1) is {<x, y>: x loves y}, so it follows, with the T axiom, that T(*B*) ↔ Helena loves Demetrius. That's the right result. We can deal with intentional failure –

cases like the belief that Santa is coming – in the same way we dealt with it in the previous chapter. Beliefs might be existentially committing or not. Suppose a child has an existentially committal belief B_1 that Santa is not coming (because they have been naughty), and I have an existentially non-committal belief B_2 that Santa is not coming (because he doesn't exist). We have these:

$$\mathrm{T}(B_1) \equiv_{\mathrm{R}} \mathrm{O}(B_1, 1) \in \mathrm{P}(B_1, 1)$$

$$\mathrm{T}(B_2) \equiv_{\mathrm{R}} \neg[\mathrm{O}(B_2, 1) \in \mathrm{P}(B_2, 1)]$$

$$\neg\exists x[x = \mathrm{O}(B_1, 1)]$$

$$\neg\exists x[x = \mathrm{O}(B_2, 1)]$$

$$\mathrm{P}(B_1, 1) = \{x\text{: } x \text{ is not coming}\}$$

$$\mathrm{P}(B_2, 1) = \{x\text{: } x \text{ is coming}\}$$

Evaluated according to a negative free logic, motivated in the same way as in chapter one, B_1 is false and B_2 is true, which is how things should be. (If you decided that different pessimistic truth values were the way to go in chapter one, then we get corresponding truth values here, and that is how things should be instead.) Now we can deal with a slightly more complicated example, showing how co-ordination is involved in deduction. Suppose I have a belief B_3 that Venus is big and a co-ordinated belief B_4 that Venus is round. I can infer that something is round. We start with these, expressing their truth conditions and the co-ordination relation:

$$\mathrm{T}(B_3) \equiv_{\mathrm{R}} \mathrm{O}(B_3, 1) \in \mathrm{P}(B_3, 1)$$

$$\mathrm{T}(B_4) \equiv_{\mathrm{R}} \mathrm{O}(B_4, 1) \in \mathrm{P}(B_4, 1)$$

$$\Box_{\mathrm{R}}[\mathrm{O}(B_3, 1) = \mathrm{O}(B_4, 1)]$$

Now I want to form a new belief B_5 (that Venus is big and Venus is round) which is guaranteed to be true if B_3 and B_4 are both true. This belief will have two object positions, co-ordinated with each other and with the object positions of B_3 and B_4, and two property positions, one co-ordinated with each of the property positions of B_3 and B_4.

$$\mathrm{T}(B_5) \equiv_\mathrm{R} [\mathrm{O}(B_5, 1) \in \mathrm{P}(B_5, 1) \ \&\ \mathrm{O}(B_5, 2) \in \mathrm{P}(B_5, 2)]$$
$$\Box_\mathrm{R}[\mathrm{O}(B_5, 1) = \mathrm{O}(B_5, 2) = \mathrm{O}(B_3, 1) = \mathrm{O}(B_4, 1)]$$
$$\Box_\mathrm{R}[\mathrm{P}(B_5, 1) = \mathrm{P}(B_3, 1)]$$
$$\Box_\mathrm{R}[\mathrm{P}(B_5, 2) = \mathrm{P}(B_4, 1)]$$

From all these truth conditions and representational requirements, it follows that it is representationally required that if B_3 and B_4 are both true, B_5 must be true:

$$\Box_\mathrm{R}[[\mathrm{T}(B_3) \ \&\ \mathrm{T}(B_4)] \rightarrow \mathrm{T}(B_5)]$$

This shows how co-ordination justifies the integration of beliefs into deductions. We can generalize this sort of thing, by defining notions of *deductive irrationality*, *deductive licensing*, and *deductive forbidding*:

$$\Sigma \text{ is deductively irrational} =_{\mathrm{df}} \Box_\mathrm{R}\exists\beta[\beta\in\Sigma \ \&\ \neg\mathrm{T}(\beta)]$$
$$\Sigma \text{ deductively licenses } \alpha =_{\mathrm{df}} \Box_\mathrm{R}[\forall\beta[\beta\in\Sigma \rightarrow \mathrm{T}(\beta)] \rightarrow \mathrm{T}(\alpha)]$$
$$\Sigma \text{ deductively forbids } \alpha =_{\mathrm{df}} \Box_\mathrm{R}[\forall\beta[\beta\in\Sigma \rightarrow \mathrm{T}(\beta)] \rightarrow \neg\mathrm{T}(\alpha)]$$

There is room for debate over the exact normative force of these kinds of facts. The issues are parallel to issues about the normative role of logic. For orientation in that debate see MacFarlane [MS]. Some of my own views on the subject appear in my [MSa, MSb]. A reasonable first pass is the following:

If Σ is deductively irrational, Σ is jointly impermissible.
If Σ deductively licenses α, and Σ is permissible, α is permissible.

We don't have a special norm for deductive forbidding, because it can be defined in terms of deductive irrationality. Σ;α is deductively irrational iff Σ deductively forbids α, although the two notions might not be interdefinable in a development of the system based on a peculiar logic in which $\Sigma, \alpha \vDash$ did not entail $\Sigma \vDash \neg\alpha$. If we adopted a logic like that, we might need an independent concept of deductive forbidding and a separate norm for it.

2.3 Conclusion

The principal aim of this chapter has been to give an account of how rational norms governing consistency and valid deduction can apply to gappy beliefs, without any special pleading for them. To this end, §2.1 showed how an account of rational norms which is independent of the beliefs' propositional contents can be independently motivated, by considering versions of Frege's puzzle as it applies to beliefs, and especially Kripke's puzzle. The lesson we drew from them is that logical relations between interpersonally accessible propositional contents are ill-suited to explain the rational norms governing beliefs. Instead we should explain them at the level of individual psychology. We need to invoke a notion of co-ordination between the object positions of an individual's belief tokens. These co-ordination relations could be explained by a language of thought, but however they are realized, folk psychology is committed to the co-ordination relations. I considered the objections that co-ordination was not a notion in good standing, either because we could not always know when our beliefs were co-ordinated, or because it is possible in principle for co-ordination to happen without co-reference. I conceded that co-ordination relations were not transparent, but they are no worse off in this regard than beliefs, which are uncontroversially subject to rational norms. I argued that while in some cases we may seem to have co-ordination without co-reference, this would lead to a problematic regress. I concluded that the concept of co-ordination is in good shape.

§2.2 showed in a more formal way how we can use the co-ordination relations to define the notions of consistency, deductive licensing and deductive forbidding which we need to talk about rational norms. These relations were defined independently of the beliefs' actual contents, so they do not fall foul of Frege's and Kripke's puzzles, and they can still apply when there is intentional failure. They were instead defined in terms of the beliefs' potential objects, as constrained by co-ordination relations. None of this relies on any commitment to gappy propositions, because the co-ordination relations are about what beliefs have the potential to represent, not what they actually represent.

Now we have a reasonable account of how names could be meaningful, and how the corresponding beliefs could be involved in deductions suitable for non-trivial rational appraisal, even when the names are genuinely empty and the beliefs suffer from intentional failure. In chapter one we argued that the sentences and beliefs in question should be assigned pessimistic truth values, at least in non-intentional contexts. Pessimistic truth values, however, have less plausibility in intentional contexts, like ‘Leverrier believes Vulcan is a planet’ and ‘Fred worships Zeus’. The next chapter will try to give a more suitable account of this kind of case.

The next chapter will also pick up a loose end which was discarded in this chapter. There I argued that no level of content could be both public enough to explain communication and fine-grained enough to explain the rational norms governing individuals’ beliefs. For the problems of this chapter, we had to go for an individualistic solution. In the next chapter we will be looking at public-level content again.

Chapter 3 – Contents without Constituents

3.0 Introduction

Chapter one dealt with empty names introduced in the context of mistakes and lies, where there is little pressure to be charitable about the truth values of utterances containing them. We allowed that the names in this kind of case were meaningful, or at least grammatical, but said that atomic predications containing empty names were false, or at least untrue. We can get different truth-evaluations of compound sentences containing empty names depending on how we understand the untruth of the atomics and how we understand the compounding, and I argued that a negative free logic without truth value gaps is probably the best option.

Chapter two discussed the corresponding systems of propositional attitudes: if a child thinks ‘Santa’ refers they will acquire corresponding gappy beliefs. That chapter offered a way of understanding the rational relations between beliefs, both in Frege-puzzle cases and gappy belief cases, which did not rely on the logical relations between the actual propositional contents of the beliefs. One consequence of this is that we have more freedom in giving an account of their propositional contents if any, and of the corresponding attitude ascriptions. These things are the subject of this chapter.

This chapter deals with a use of empty names where the speaker is not making a mistake: attitude ascriptions. We could talk about attitudes the way chapter two did, but this is not how we usually do it, and it does not make interpersonal generalizations about propositional content. Normally we ascribe attitudes using the empty names the speakers would use in making assertions expressing those attitudes. We describe Leverrier as believing that Vulcan was a planet, using the name ‘Vulcan’, even though if everyone had known what we know about the solar system then the name would not have been introduced into the language. Since our use of the name in the attitude ascription does not arise from any mistake of ours, there is some pressure to give charitable truth values to sentences like ‘Leverrier believed that Vulcan was a planet’.

Where empty names are not involved, a semantics for attitude ascriptions usually takes them as referring to a proposition, although they may also be referring to something else too[35]. In the simplest picture ‘S believes that P’ is analysed as a binary predication, asserting that the referent of ‘S’ stands in the belief relation to the proposition referred to by ‘that P’, which will be the proposition ordinarily expressed by ‘P’[36]. There can be some variation from this simple model,

[35] Frege [1952] and Church [1950, 1954] have done much to influence the view that attitude ascriptions involve reference to propositions. Frege said that in direct quotation sentences refer to themselves and in indirect quotation they refer to the thoughts they customarily express. The attempt to have sentences in indirect quotation and attitude contexts referring to themselves was embarrassed by Church’s translation test, since the German for e.g. ‘“the sky is blue”’ is ‘“the sky is blue”’, and having the English expression appear in the German translation of an English attitude report looks terrible. I agree with the spirit of Church’s point even if I have misgivings about the letter, mostly relating to considerations about the individuation of linguistic expressions of the type raised by Kaplan [1990]. As such I have been trying to avoid metalinguistic analyses of discourse not overtly about language where possible. I will continue avoiding them in this chapter.

[36] From a syntactic point of view, it is probably unsatisfactory to treat ‘that’ clauses as noun phrases, as argued in Huddleston and Pullum [2002: 1014-22]. In the semantics I give in the appendix to this chapter they will be in a special category called *proterms*. From a semantic point of view we can however get the results we want by treating proterms as referring to propositions, while having a different syntactic category allows us to hold that

but typically some appeal will be made to the propositional content of the embedded sentence. Here we have a problem when ‘P’ contains an empty name. We might think that normally if a sentence contains a name, the referent of the name will be a constituent of the proposition the sentence expresses, either directly or by being the *res* of a *de re* Fregean sense which is a constituent of the thought, in the manner of McDowell [1984]. Where ‘P’ contains an empty name we want ‘that P’ to name the proposition ‘P’ expresses, but *prima facie* that would mean reifying a proposition without reifying all its constituents. That looks bad, but it would be a shame to have to abandon the semantics for attitude ascriptions which takes them as referring to propositions, unless we have to.

Solutions to this problem can vary a lot in their details, but in view of the way we have set it up they can be divided into three general strategies:

- Reify the constituents.
- Avoid reifying the constituents by not reifying the propositions either.
- Reconcile the reification of the propositions with the rejection of the constituents.

§3.1 will raise some issues which might help us choose between these options, relating to the theoretical role of propositions, intuitions about truth values and the validity or invalidity of some arguments, the connection between propositional and objectual attitudes, and the theories’ ontological commitments. §3.2 will discuss the first strategy, where the constituents are either taken to be non-existent concreta or existent abstract objects. §3.3 will discuss some versions of the second

they cannot be freely interchanged with names while preserving grammaticality.

strategy, either taking the attitude ascriptions in question to be false, paraphrasing them, or understanding speakers as only pretending to ascribe attitudes. §3.4 will discuss two versions of the third strategy, where the contents are either David Braun's ontology of Russellian propositions which can have gaps (which we first met in §1.1), or Fregean propositions which can have non-denoting modes of presentation as constituents. I will come down on the side of this last option. I present a semantics for attitude ascriptions embodying the Fregean view, and show how the proposal copes with the issues discussed in §3.1. An appendix presents the semantics in a more formal way.

3.1 Desiderata for a solution

With a few solutions to the problem of contents without constituents on the table, we need to find some criteria for deciding between them. In this section I will discuss four issues a solution should deal with:

- The theoretical role of propositions for expressing generalizations about communication and thought.
- Truth intuitions about attitude ascriptions: we want an account of attitude ascriptions which either makes the ascriptions true which we think are true, or explains why we get them wrong.
- A theory of attitude ascriptions is best if it links up nicely with a theory intentional[37] transitive verbs, such as ‘admires’, ‘worships’ and ‘seeks’.
- The ontological commitments of our chosen solution should not be too implausible.

- ### 3.11 The role of propositions

In §2.11, I said this:

> One way of thinking about communication is that I believe something, express my belief with an assertion, and then you end up believing what I believe. This kind of communication aims to get the hearer to resemble the speaker in some way, and this resemblance can be described as us having beliefs with the same contents. We can also say that assertions have the same

[37] There is a tendency in the literature [e.g. Richard 2001, Forbes 2013] to call these ‘intensional transitives’, with an S. This is misleading. It prejudges the question of whether they are extensional, and suggests they are not hyperintensional. The question is how to deal with verbs expressing intentional attitudes, and if we want to phrase the question in terminology neutral between different answers, we should call them intentional transitives, with a T.

> contents as the beliefs they are used to transmit. I will argue that no level of content which can play this role in understanding communication can also explain the rational relations between belief tokens.

In chapter two I was looking at the role of propositions in explaining rational relations between belief tokens, but now we can focus on the other role, in expressing generalizations about thinkers and communication between them. The basic picture is that thought contents express important resemblances between thinkers, including how they relate in similar ways to the world. Utterances have contents transmitting these resemblances. There are two important things to note about this picture as applied to the problems of the present chapter. First, we need to make sure we individuate the contents in a way that captures the important generalizations. If two children's beliefs that Santa is coming are importantly similar to each other and different from a third child's belief that the Tooth Fairy is coming, then it would be good to say that the first two children's beliefs had the same content and the third's had a different one. This issue is about generalizations about thought. The second issue is about the link between thought and communication. When an utterance expresses a belief token, in much the same way as saying 'ouch' expresses a pain token, the content of the utterance and the content of the belief should be the same. This gives sense to the idea that people verbally express their thoughts, rather than saying one thing because they think something else altogether. This issue will also become important in §4.432, when we look at intentional attitudes involving fictional names.

• 3.12 Truth intuitions about attitude ascriptions

We make attitude ascriptions, and have intuitions about their truth values. Some of these intuitions are firmer than others, but insofar as we have truth intuitions about attitude ascriptions, a theory is better if it predicts that these intuitions are right. Where it predicts that they are wrong, it would be good to have an explanation for this. Here are some examples of intuitions about attitude ascriptions we might want to uphold:

- *Disquotation*: If a normal English speaker, on reflection, sincerely assents to 'p', then they believe that p. [From Kripke 1979: 439]
- *Biconditional disquotation*: A normal English speaker who is not reticent will be disposed to sincere reflective assent to 'p' iff they believe that p. [Also from Kripke 1979: 439]
- *Non-substitutivity:* A person can believe that *n* is *F* and not believe that *m* is *F*, even if (in fact, unbeknownst to them,) *n* is *m*. [From Frege 1952]
- *Positive quantifying in*: If *n* is a *G* and a person believes that *n* is *F*, then there is a *G* that they believe is *F*. [From Sider 1995: §8]
- *Negative quantifying in:* If *n* is a *G* and a person does not believe that *n* is *F*, then there is a *G* that they do not believe is *F*. [Also from Sider 1995: §8]

All these principles give rise to puzzle cases, and some solutions to the cases might give some of them up, or at least restrict them. However, it would be good if our theory of attitude ascriptions containing empty and/or fictional names did not create any new problems with respect to our truth intuitions about attitude ascriptions.

3.13 Intentional transitives

There is a temptation to think that once you've got a theory of attitude ascriptions that links up nicely with a theory of the embedded sentences appearing as assertions, you're done. But you're not really done, because of intentional transitives. We don't just express intentional attitudes using sentences of the form 'x [attitude]s that φ'; we also express them with sentences of the form 'x [attitude]s y'. Some examples:

- Jack admires Jill.
- Jill worships Zeus.
- Everyone loves a sailor.

You could just take these as a separate problem, claiming that nothing about attitude ascriptions directly commits you to anything about intentional transitives. Alternatively, you could say, drawing on Larson et al [1997], that intentional transitives can be paraphrased as propositional attitude ascriptions like 'Jack thinks Jill is admirable', and then say that the semantics for attitude ascriptions will therefore suffice. This strategy is a bit of a promissory note and its use of paraphrase is slightly unsatisfactory in the way uses of paraphrase tend to be, even if extensionally adequate paraphrases could be given.

If we don't take a theory of attitude ascriptions to be an automatic theory of intentional transitives, it is dangerous to try treating the problems separately. That is because there are connections between our propositional and objectual attitudes. If Jack admires Jill, that is a reason to think Jack thinks various things about Jill. Substitutivity issues arise for intentional transitives as well as propositional attitudes: Lois fancies Superman but not Clark, and believes Superman is brave and Clark isn't. Moreover, it makes sense that she fancies the one she thinks is brave and doesn't fancy the one she thinks isn't. Existential commitment issues arise as well: you can worship Zeus even if Zeus doesn't exist, although *pace* Parsons [1980: 217] you probably can't rationally worship him if you know he doesn't exist. Lastly, intentional transitives raise issues about the validity of arguments involving quantification, just as propositional attitude ascriptions do. For some examples of valid or at least nearly valid argument forms involving intentional transitives, see Richard [2001: 106-7].

Part of the issue is that similar problems arise for both sorts of attitude ascriptions, so we would be missing a trick if we didn't at least see whether the same machinery could deal with both. It is however at least as important that an account should be unified enough to make the connections natural, and let us express those connections by saying things like ‘Lois admires everyone she believes is brave’, which should be inconsistent with ‘Lois thinks Clark is brave but doesn’t admire Clark’. This will be hard to explain if our analysis of ‘Lois thinks Clark is brave’ is unrelated to our analysis of ‘Lois admires Clark’.

- **3.14 A plausible ontology**

As with many debates in philosophy, one way of choosing between competing theories is to look at their ontological commitments. We don’t want our theory to commit to too many things. If it does, it is good if they are things we are committed to already for some other reason. If they are not, it is good if the new things are not too strange. In all cases, it is good if there is a satisfying explanation of how the considerations at hand give reasons to think that there are such things, rather than that people mistakenly think there are such things. If it sounds like these platitudes are being used to stitch up the Meinongians, that is because in a way they are. However, they really are platitudes, people really do reject Meinongianism on the basis of them, and part of the project of this thesis has been to provide a viable and ideally preferable alternative to Meinongianism. I’m not dismissing Meinongianism out of hand or saying it is incoherent, but that doesn’t mean I can’t draw attention to its costs.

3.2 Reifying the constituents

3.21 Meingongianism

A Meinongian solution to the problem of contents without existent constituents says they have non-existent constituents. The big, obvious problem with this is that the ontology is implausible to a lot of people, and as I've just said, probably with good reason. However, it shouldn't go unnoticed that the Meinongian proposal is quite neat. If you can come up with a Meinongian ontology which doesn't give rise to paradoxes, the solutions it offers to the problems of attitude ascriptions are quite straightforward. You can effectively just graft it onto your preferred theory of attitude ascriptions. If you are a Russellian, then you can have people believing propositions with non-existent objects as constituents. If you are a Fregean, you can have people believing propositions with modes of presentation of non-existent objects as constituents. Intentional transitives can also be dealt with straightforwardly, with thinkers related either to non-existents or to modes of presentation of non-existents.[38]

Meinongianism makes it easier to give an account of the semantics of intentional attitude ascriptions not involving existents. This feature might appeal to our laziness, but what would really speak in its favour is if there was something Meinongianism could do that its competitors couldn't do, even with difficulty. In fact, there may be. Meinongianism allows you to reconcile the following:

[38] If we go for the second option, this won't mean that 'Fred worships Zeus' says that Fred worships a mode of presentation of Zeus. It will mean, approximately, that Fred worships the thing that mode presents, via that mode. This allows us to reject the substitution of identicals *salva veritate* in intentional transitive contexts, since different modes of presentation could present the same (non-existent) object, and a subject could worship that object via one mode but not the other.

- No existential commitment: it is possible that n does not exist and S believes that n is F, where these claims are taken at face value rather than paraphrased.
- Non-substitutivity *salva veritate* of names without existent referents: it is possible that n does not exist, and m does not exist, and S believes that n is F, and S does not believe that m is F.
- Referential transparency: if n is m, and S believes that n is F, then S believes that m is F.

The first two features are fairly uncontroversially desirable. Referential transparency is less clearly desirable; it isn't consistent with the non-substitutivity and biconditional disquotation desiderata from §4.12, given that people sometimes competently assent to 'n is F' and not 'm is F', even when n is m. There are however things that can be said in favour of referential transparency, especially if we don't tie it to existential commitment and substitutivity *salva veritate* of names without existent referents. You can construct examples which seem to speak in favour of referential transparency, such as those due to Jennifer Saul [1998] and Sider and Braun [2006] which will come up in §4.41. Referential transparency also makes it easier to give a semantics validating positive and especially negative quantifying in, although the semantics I will give manages to validate both of these without transparency. Finally, transparency can seem especially plausible in the case of intentional transitives. Here is Kripke:

> What about [Church's Fregean] analysis? Applying it here [to intentional transitive verbs] seems to me to be beset by various difficulties. First, it implies that the verb 'worship' is intensional, in the sense of not being subject to ordinary substutivity of identity. But this seems to me not to be so. And similarly for 'admires'. Suppose Schmidt admired Hitler. If Hitler was the most murderous man in history, then it seems to me that Schmidt did admire the most murderous man in history. And it does not seem to me that the

> latter statement is ambiguous as between an intensional, or opaque, and a transparent one. There is not one sense in which he did admire the most murderous man in history and another in which he didn't. If he himself would deny that Hitler can be so characterized, then it is true that he didn't admire Hitler *as* the most murderous man in history. But it still is true that he admired the most murderous man in history…
>
> …Now even when 'worships' is followed by an apparently empty name we can make such substitutions. Suppose the Greeks worshipped Zeus, and Zeus is the tenth god mentioned by Livy. Then the Greeks did worship the tenth god mentioned by Livy. [Kripke 2013: 68-69; emphasis in original]

Kripke doesn't argue for transparency for all intentional verbs, and his biconditional disquotation principle (from Kripke [1979]) conflicts with transparency for propositional attitude ascriptions. There do seem to be some reasonably strong intuitions on his side in the passage just quoted, though. In §3.42 I will explain how my preferred treatment can meet him halfway on his examples, but it is worth pointing out that reconciling his intuitions are a big difficulty for most accounts while giving Meinongianism no trouble at all.

• 3.22 Abstract artefcacts

Most people don't want to reify non-existent objects, and in §3.14 I gave some reasons for this stance. A natural alternative is to reify some existent objects to play the role instead. One option for reifying fictional characters as existent objects is taking them to be *abstract artefacts*, which means they are abstract objects which exist in virtue of our practices. Fictional characters could be created by our literary practices, but we could also have objects like Vulcan and Zeus existing in virtue of mistakes and lies. Defenders of such a view, e.g. Nathan Salmon [1998: §VI] sometimes distinguish these from fictional objects by calling them *mythical* objects. There will be much more on fictional characters as abstract objects in the next chapter, particularly §4.121 and §4.3, but for present purposes two things are important. First, the objects exist, in the same way numbers, properties, sets and whatever other abstract objects there are exist (assuming they do). Second, the objects are abstract, and as such are nothing like the way the people making the mistakes think they are. If Vulcan is an abstract object, then it is not a planet, not spatially located, does not orbit the sun, and so on. While a Meinongian can say that Vulcan is a non-existent planet orbiting the sun, the abstract artefact theorist cannot, because if another planet orbiting the sun existed then we would know about it.

Both of these divergences from the Meinongian position lead to difficulties. The first, that they exist, leads first to an obvious problem with negative existentials. When we say 'Urbain believes than Vulcan exists', we should be attributing Urbain a false belief, but on the view under consideration we would not be. There could also be problem with properties besides existence where the abstract object is coincidentally the way the subject thinks the object of their belief is. We can either respond to this with paraphrasing, or an error theory, or a combination of the two, but whichever way we go we will end up with something less neat than the Meinongian proposal.

The second issue is that we are attributing subjects some quite wacky recognition failure, in saying that people's beliefs, rather than being about nothing, are radically false beliefs about abstract objects. We are saying that someone who thinks Vulcan is a planet has the wrong end of the stick in the same way as you might if, for example, you heard someone say 'I saw *Così fan tutte* last night' and thought *Così fan tutte* was a person. One might think that recognition failure this extreme is not possible, and instead results in referential or intentional failure. On this view you don't believe *Così fan tutte* is a person; you have a gappy belief referring to nothing at the object position, although you do believe '*Così fan tutte*' is the name of a person. Imogen Dickie [2011] offers such a treatment of cases like this, and gives a principled way of deciding when we have recognition failure and when we have a gappy belief, which can be applied to cases like this[39]. Whether we follow Dickie on this or not, the case is weird, and won't come up very much.

Both of these problems make the resultant theory ugly, but they probably do not make it untenable. A third problem may be more decisive. If we adopt the abstract artefacts view, then we may have to sever the link between the contents of the attitudes we ascribe people and the contents of their utterances, which undermines the theoretical role propositions were meant to play in the first place. If we do not, then we can't use the proposals given in chapter one. Furthermore, we wouldn't be able to use the proposals for understanding gappy beliefs in chapter two, unless we said that gappy beliefs can be analysed in two ways: as gappy beliefs when explaining the cognition at an individual level, and as about abstract artefacts when ascribing them shareable contents. If abstract artefacts were the only acceptable option on the table, then we would just have to live with that, but we will see later than other options are available, and adopting them does less violence to the hopefully well-motivated machinery of the first two chapters.

[39] She considers our judgements of when reference fails and when it succeeds in a variety of cases, and argues that to succeed we have to be somehow tuned in to the possible behaviour of the thing we are trying to refer to. She calls this the *Governance View*. The possible behaviour of an abstract artefact is radically different from that of a person, so her view is likely to rule that reference will fail in this kind of case.

3.3 Rejecting the contents

- ### 3.31 Error theory

As a general methodological principle, it is usually good idea to be open to the possibility that the discourse we are trying to understand is mostly false. We can separate the questions of whether mathematical discourse commits us to abstract objects and whether there are such objects, or whether moral discourse commits us to objective values and whether there are such values. To say that the commitments of a discourse are false in this way is to embrace an *error theory*. Some error theories are very plausible; for example some religious discourse almost certainly commits to there being things of a kind which there are not. Interpretive charity must come to an end somewhere, especially if we have an explanation for people making the error in question.

As such, rejecting a Meinongian ontology does not immediately rule out a Meinongian analysis of discourse involving fictional names. Marga Reimer [2001] proposes just this package. We analyse sentences involving empty names as the Meinongian does, except that wherever the Meinongian sees reference to a non-existent, we see failure of reference. Even attitude ascriptions are then bad cases which can be treated along the lines of chapter one, because it is a mistake to be committed to the content the believing of which we are ascribing. There are no non-existents, so there are no contents with non-existents as constituents, and so nobody can believe them, and we shouldn't say they do.

Reimer notes that even if a theory of gappy propositions like Braun's (see §4.41) can account for all our intuitions about the truth of sentences involving fictional names, it can't account for our intuitions about their content. Following A. P. Martinich [1996: 184], she says that non-philosophers think that when we use empty names we are talking about their bearers. If a kid says ‘Santa is coming’, or if I say ‘Santa doesn’t exist’, we say they are talking about Santa. These content intuitions commit us to there being a Santa, even if there doesn’t have to be a Santa to evaluate what the child says as false and what I say as true. And if what we say about contents commits us to a Meinongian ontology, then it is reasonable to take us as (implicitly) committed to one, and to give a Meinongian analysis of discourse involving empty names. This leads to either a Meinongian ontology or an error theory, and Reimer’s sense of reality is robust enough to recommend the latter.

Reimer’s challenge is a serious one, and she is right to separate the commitments to the Meinongian analysis and the Meinongian ontology. We can argue that her position is uncharitable, but as with other error theories, charity has only so much weight. Another way of arguing against her is to point out that the case for non-Meinongian analysis of different kinds of discourse involving apparently empty names doesn’t only rest on the implausibility of the Meinongian ontology.

In chapter four we will distinguish two kinds of discourse involving fictional names, one which is about things as real as musical works, and one which is pretence and so does not commit us to anything. In §3.42 I will offer a non-Meinongian analysis of the attitude ascriptions. Even if we stand by the case for those analyses, we could still remain error theoretic about the content intuitions, although hopefully we won’t have to even do that.

We can of course question how strong and how resilient the intuitions are, but if we can’t explain them away, we can offer a non-Meinongian analysis of the content intuitions. We could take ‘x is about y’ not to be extensional in the y position, and analyse it the same way we will analyse intentional transitives in §3.42. Alternatively, we could follow Andrew Bacon [2013] and adopt an independently motivated positive free logic for ascriptions of reference and aboutness. The first option is closer to my own position, but Bacon's proposal shows that Reimer's is not the only other game in town when it comes to ascribing reference and aboutness.

An error theory would be appropriate if we found non-Meinongian analyses of attitude ascriptions unsatisfying, but still found the Meinongian ontology implausible. I agree that the Meinongian ontology is implausible, but I don't agree that non-Meinongian analyses are unsatisfying. The case against Reimer rests in large part on the positive case for the analyses I will give later on, in §3.42 for attitude ascriptions, and in the next chapter for discourse involving fictional names.

3.32 Pretence and paraphrase

Reimer's error theory is one way of rejecting the contents because there aren't the constituents: she says we're committed to them and we're wrong. A more charitable way is to say that we know full well there aren't the contents, and so we don't presuppose that there are. We might sound like we are presupposing that there are the contents, but we are speaking either non-seriously or non-literally. There are a few ways of cashing this out. On the non-serious side, we could say that the attitude ascriptions are pretence. We could either be pretending to attribute Meinongian contents, or pretending to attribute contents involving the abstract artefacts if we actually want to commit to those[40]. On the non-literal side, we could paraphrase the apparent attitude ascriptions as somehow conveying information about thinkers' knowledge of the make-believe games, or the texts, or something like

[40] In the terminology of §4.42, pretending to attribute Meinongian contents without committing to there being such contents would be *unanchored* pretence, while pretending to attribute contents involving objects which we do commit to would be *anchored* pretence.

that. Walton [1990: 396-419] proposes something like this, where the paraphrases are unsystematic, but in each particular case taken in context it will be reasonably clear what information is being conveyed.

My objection to the pretence and paraphrase accounts also applies to the error-theory. Certainly we do sometimes speak non-seriously, non-literally or falsely, in order to convey true information to one another. There is no particular reason in principle why attitude ascriptions involving fictional names could not be a case in point. The problem is that it isn't plausible to say that this non-serious, non-literal or false discourse doesn't convey some true information, and if we admit that it does, our theoretical work isn't done. Saying that it isn't part of the literal content and thus consigning it to what Kripke [2011d: 328] called the 'pragmatic wastebasket' doesn't really help.

Recall the role of propositions for expressing generalizations about thought and communication. We don't commit to propositions just to make ourselves feel better about our actual propositional attitude ascriptions; we commit to propositions either to explain or at least describe resemblances between the way thinkers cognitively engage with the world, and the way these resemblances propagate themselves through communication. The question about whether attitude ascriptions involving fictional names commit us to a level of propositions is not just about giving a semantics for the attitude ascriptions; it is about whether the resemblances we are conveying information about are properly modelled by a level of propositional content. I will suggest in §3.42 that they are.

Once we have done the work, investigating whether it is reasonable to invoke a level of propositional content here, we may find that a semantics for the attitude ascriptions drops out of it. This will be the case if two conditions are satisfied: the resemblances must be describable in terms of propositional content, and the resemblances so described must match up systematically enough to the sentences we use to convey the information that a semantics can be given. If it can, this will undercut the motivation for treating the attitude ascriptions as non-serious, non-literal or false. It follows then that treating the attitude ascriptions in this way only leaves the job half done, and until we have done the other half we can't know whether we did the first half right.

3.4 Gappy propositions

• 3.41 Russellian gappy propositions

We first met David Braun's theory of Russellian gappy propositions (GPs) in chapter one[41]. Then we were considering it as a way to justify evaluating the truth values of sentences containing empty names according to a negative free logic. We saw that it wasn't the only way to justify this, but it was one way. As well as having sentences express GPs, however, we could also let embedded sentences in attitude ascriptions refer to GPs. So, on that view, 'Vulcan is a planet' expresses <∅, {Planethood}>, and 'Urbain believes that Vulcan is a planet' says that Urbain believes <∅, {Planethood}>.

The first worry about Russellian GPs is metaphysical. They are strange things, in that we know which possible state of affairs corresponds to <{Mars}, {Planethood}> but it is mysterious which corresponds to <∅, {Planethood}>. We can see it as an abstraction from all the propositions that something is a planet, but perhaps that doesn't help. This metaphysical worry is a problem if we want propositions to play a robust explanatory role, but if we only want them to play a descriptive role we can just define their descriptive role and leave it at that. From chapter two, we have a notion of beliefs having a representationally required structure to their truth conditions, and being about objects

[41] To recap: an ordinary Russellian proposition is an entity structured similarly to the sentence expressing it, but instead of being composed out of words it is composed out of the objects and properties words refer to. We can represent them as ordered n-tuples, e.g. the proposition that Jack is tall is represented as <{Jack}, {Tallness}> and the proposition that Jack loves Jill as <<{Jack}, {Jill}>, {Loving}>. Russellian GPs are like this except they can have gaps where the objects would go in an ordinary proposition. We represent the gaps with the empty set, so the proposition that Vulcan is a planet is represented as <∅, {Planethood}>.

and properties. (In chapter two we represented the properties as extensions for simplicity, but now we should use properties.)

In the notation of chapter two, we can say that if someone believes the (gapless) proposition represented as <{n}, {P}> iff they have a belief *B* such that:

- True(*B*) $\equiv_R$ O(*B*, 1) instantiates R(*B*, 1)
- O(*B*, 1) = n
- R(*B*, 1) = P.

Similarly, we can say that they believe the GP represented as <∅, {P}> iff they have a belief *B* such that:

- True(*B*) $\equiv_R$ O(*B*, 1) instantiates R(*B*, 1)
- ¬∃x[O(*B*, 1) = x]
- R(*B*, 1) = P.

These definitions generalize to other GPs. This doesn't say anything about what GPs or propositions in general are, but if we are only using them for the purposes of description then this probably does not matter. If the definitions in terms of ordered n-tuples were taken as read, then we could probably use the n-tuples themselves, but since in everyday discourse these definitions are not taken as read, we can take people's descriptions to presuppose an ontology of GPs structurally similar to the n-tuples. How happy we are with this will really depend on what we think about the role of abstract objects in general.

Now we have a way for the Russellian GP theorist to determine who is going to count as believing which Russellian GPs. We can also give a semantics for attitude ascriptions bearing it out in terms of the sential semantics in the appendix to this chapter[42]. This makes the proposal precise, so now we can see whether it does what we want. First let's

[42] The restriction on models we need to apply is to have each object denoted by only one object sense, and have only one non-denoting object sense.

look back at the criteria we considered in §3.12. We have disquotation, but not biconditional disquotation. This goes with the fact we have referential transparency, rather than non-substitutivity. That conflicts with some intuitions, but there are arguments against those intuitions based on some examples we will look at shortly. On the sential semantics we will have both positive and negative quantifying in, unless the quantification is taken as existentially committing, in which case we presumably shouldn't have either. Like the Meinongian we don't have existential commitment, but unlike the Meinongian we have substitutability *salva veritate* of non-referring names in belief contexts. This isn't a bad showing for the Russellian GP proposal. Now let's look at some objections.

The first is that it doesn't carry over very well to an account of intentional transitives. We could just treat them as extensional according to a negative free logic, which means you can't admire Thor because there is no Thor. Alternatively, we could say (again using the sential semantics) that if someone has an attitude towards *n*, and 'n' is empty, then they have that attitude towards all *m* where 'm' is empty. That is not very satisfying, but it preserves the link between objectual and propositional attitudes, since the classification of the propositional attitudes is correspondingly unsatisfying. Alternatively, we may be able to paraphrase objectual propositional attitudes ascriptions, but this does no better than the previous strategy, since empty names will again be interchangeable. This is not really an extra problem though, since if we can get used to Leverrier believing Thor is a planet, we can probably get used to the Vikings worshipping Vulcan.

This leads us into the second objection, though: can we get used to Leverrier believing that Thor is a planet? This problem about the interchangeability of empty names is closely related to Anthony Everett's objection, discussed in §1.2. That objection was that the Russellian GP theorist should be able to say what 'Santa Claus doesn't exist' and 'Father Christmas doesn't exist' have in common with each other that they don't have in common with 'Hamlet doesn't exist'. I argued in §1.2 that there are things you can say in response: first that the objection risks unfairly playing two incompatible intuitions off against each other, and second that we can cash out the difference in terms of reference-fixing rather than propositional content. Nonetheless, the Russellian GP account does make a lot of positive attitude ascriptions come out surprisingly true, and their negations come out surprisingly false.

Millians [e.g. Salmon 1986: ch. 8] have been defending themselves against this line of attack for a while, since they are also surprisingly permissive about attitude ascriptions like 'Lois believes Clark can fly' and 'Lois believes Clark is Superman'. Similarly, they must reject 'Lois does not believe Clark can fly', although they accept 'Lois believes Clark cannot fly', and Lois would reject 'Clark can fly'. The Millian can say that Lois believes the propositions but not under the guises associated with the embedded sentences 'Clark can fly' and 'Clark is Superman'.

One way of bolstering the Millian response is to find situations where we would make the strange attitude ascriptions the Millian accepts, so we can say that they are strictly speaking true although normally we wouldn't say them. Jennifer Saul gives an interesting example the Millian can use:

> The well-known failures of substitutivity only tell half the story. Sometimes, substitution of co-referential names *does* seem to guarantee sameness of truth value. The following provides some indication of this: Suppose I am discussing what people tend to think of Bob Dylan's singing abilities, and the person I'm talking to knows him only as 'Bob Dylan'. I've been told (truthfully) that Glenda, a childhood friend, who knows him only as 'Robert

> Zimmerman', believes that he has a beautiful voice. Specifically, someone I trust has uttered sentence (6):
>
> > (6) Glenda believes that Robert Zimmerman has a beautiful voice.
>
> I may report this with sentence (7)[43]:
>
> > (7) Glenda believes that Bob Dylan has a beautiful voice.
>
> (7) seems true, even though Glenda would never assent to it. To know that (7) is true, moreover, we don't need to know anything at all about how Glenda thinks of her childhood friend Robert Zimmerman. All that matters is his identity, and the fact that she liked his voice. Substitution inferences, this suggests, are sometimes perfectly acceptable. Since we sometimes find them unacceptable, we need an account which can reflect the fact that our intuitions about the legitimacy of substitution inferences vary with context. [Saul 1998: 366; original emphasis.]

This presumably isn't game over for the Fregean, but it does suggest that if the Fregean is right then the amount a name's associated mode of presentation contributes to an attitude ascription's assertability conditions (and perhaps truth conditions) varies with context. Saul's example makes use of the fact that the person having the attitude ascribed to them is only familiar with Dylan under one name, and their audience was only familiar with him under a different name. We can elicit a similar intuition if the person ascribing the attitude only knows him under one name. Suppose Dylan met Glenda back home and said 'hey, do you know a lot of people think I'm the best songwriter in the world?'. Glenda could then truthfully say this to her friends: 'a lot of

[43] This '7' is a '2' in the original, which I have assumed to be a typo.

people think Robert Zimmerman is the best songwriter in the world'. It works both ways, and both kinds of example support the Millian.

However, this support for the Millian doesn't look like it will carry over to support the Russellian GP theorist. They need cases where people would say 'Urbain thinks Thor is a planet', and you can't construct those in the same way. You can of course try persuading people of Braun's theory on other grounds and then assert that Urbain thinks Thor is a planet, but that kind of data is hardly admissible. Even if the Millian can talk people round over substituting co-referring names, the Russellian GP theorist will probably have to just bite the bullet over substituting empty names.

Sider and Braun [2006] also defend Millianism by arguing that our non-Millian intuitions about the truth values of attitude ascriptions clash with our logical intuitions about the validity of arguments, such as positive and negative quantifying in. This argument doesn't carry much weight against the proposal I'm adopting in the next section, because it validates positive and negative (non-objectual) quantifying in without being a Millian. Even if you don't adopt my solution, however, note that in the context of empty names you can only make positive and negative quantifying in work anyway if the quantification is not existentially committing. If the quantification is objectual, then the argument form 'S believes that n is F; therefore ∃x[S believes that x is F]' isn't valid, since Urbain believes Vulcan is an intramercurial planet but there is nothing Urbain believes is an intramercurial planet. Empty names thus undermine the case for Millianism based on quantification intuitions. Combined with the problems with substitution intuitions and intentional transitives, we should consider an alternative view.

- ### 3.42 Fregean gappy propositions

One advantage of Russellian (non-gappy) propositions is that their constituents are ordinary objects. If, perhaps unlike Frege[44], you can stomach the idea of ordinary concrete things being the constituents of thought contents, there isn't a further mystery about what the constituents are like. They are ordinary objects, and we already know what those are like. If you want to have Fregean propositions, however, distinguishing between propositions that Cicero is bald and that Tully is bald, then you need to say something more about what the constituents are like. The Fregean answer is to have the constituents be senses, which are modes of presentation of objects, but there is work to be done saying what these senses are like.

Russellian GPs have a similar advantage. If you can stomach the idea of propositions with gaps in, there isn't a further question about what the gaps are like. The gaps left by the emptiness of 'Thor' and 'Vulcan' are the same: they are just gaps. Consequently, the propositions that Thor is angry and that Vulcan is angry are the same. If you want to distinguish them, as I argued there is some pressure to do, then we need to say more about the contributions of 'Thor' and 'Vulcan'. They don't contribute an object, but they don't just leave a gap either. The natural Fregean answer is that they contribute a non-denoting sense, which gives us what we can call a Fregean GP. Susanna Schellenberg [2011: 27-9] already argues that we should understand hallucinations as having Fregean GPs as their propositional contents, where the non-denoting senses are modes of presentation which can be subjectively

[44] Frege wrote this in a letter to Philip Jourdain: 'Now that part of the thought which corresponds to the name 'Etna' cannot be Mount Etna itself; it cannot be the reference of this name. For each individual piece of frozen, solidified lava which is part of Mount Etna would then also be part of the thought that Etna is higher than Vesuvius. But it seems absurd that pieces of lava, even pieces of which I had no knowledge, should be parts of my thought.' [Frege 1980]

indistinguishable from perceptual modes of presentation which do denote. The idea here is to extend that account to cover the contents of attitudes ascribed using empty names. It gets you the propositions you want, but it leaves some work to do on the metaphysics, explaining what the Fregean senses are and why the ones we need can be non-denoting. I will have to leave some of that for further work, although I will have some preliminary things to say about it. First, however, I will show why the work is worth doing, by setting out a version of the proposal in a more detailed way, and considering some of the problems the it solves.

To get a fixed version of the proposal, we need a formal semantics for attitude ascriptions which can validate Fregean non-substitutivity intuitions, and which allows non-synonymous non-denoting names. Some work has been done in this area both others, such as Richard Montague [1973], and Richmond Thomason [1980] which builds on Montague. The semantics I will use is my own, from Bench-Capon [MSc]. A formal presentation of that semantics is given in an appendix to this chapter; now I will present it informally and show to what extent it meets the desiderata laid out at the beginning of this chapter. That way we will have a concrete version of the semantic part of the Fregean GP proposal, and we will be able to see how it helps. The metaphysical part, establishing what the potentially non-denoting senses are really like, is left for further work; at this stage the senses are more or less black boxes, although we will talk some more about how the further work might go at the end of this section.

The central idea is to have two value functions, one from linguistic expressions to senses, and another from senses to references. We allow non-denoting senses by having the function from senses to references be partial, but we rule out meaningless expressions by having the function from expressions to senses be total. To get the results we wanted involving quantification, the basic kind of quantification is sential, not objectual, in that it uses assignments which assign senses to expressions, not objects. The senses in the ranges of these assignments are from a privileged subset of the senses, called *object senses*. Object senses can only denote objects (or nothing), and names can only express object senses. Objectual quantification can be defined in terms of sential quantification and identity. In non-intentional contexts the logic is a negative free logic, and this includes identity, so where x, y, or both are assigned non-denoting senses, x=y is false.

We can exploit the senses of expressions to deal with intentional contexts, both for objectual attitudes (intentional transitives) and propositional attitudes. In non-intentional contexts, truth values are determined by the objects denoted by the senses expressed by the expressions, whereas in intentional contexts the senses can make a difference to the truth value without differing with respect to reference. The two kinds of intentional context here are intentional transitive verbs and a THAT operator for dealing with propositional attitudes. The extension of an intentional transitive verb (i.e. the extension denoted by its sense) will be a set of ordered pairs of objects and senses, rather than just objects and objects, such that Fxy is true iff <V(S(x)),S(y)> is in the extension of F. The THAT operator is a sentential operator, such that the sense of THATφ denotes a proposition corresponding to the sense of φ. There are some constraints on the function from sentences' senses to propositions to ensure that propositions' truth values track those denoted by the senses, but aside from that we can treat propositions as black boxes, like the senses. (It is left open whether propositions *are* senses.) Now that we have the THAT operator, we can have attitude verbs like BELIEVES(x, THATφ), which is true iff the object denoted by the sense of x believes the proposition denoted by the sense of THATφ, which will be the proposition corresponding to the sense of φ. Sentences in the scope of a THAT operator can contain free variables, and this lets us quantify into propositional attitudes.

The appendix presents the foregoing material more formally, but this presentation should give us enough to work with. Now we can see how it deals with the desiderata set out in §3.1. These were: doing justice to the theoretical role of propositions, preserving truth intuitions about particular kinds of sentences and the validity of some kinds of argument, linking up propositional and objectual attitude ascriptions, and not making implausible ontological commitments.

One of the places where the proposal is strongest is in giving propositions their proper theoretical roles. The proposition referred to by THATφ just is the proposition expressed by φ, so the propositional contents we ascribe to people's beliefs are the same as the contents they assert when they assertively utter the embedded sentence. This goes for sentences involving empty names too: if we are willing to commit ontologically to Fregean GPs, then this semantics makes full use of them. Unlike the treatments involving Meinongian objects or abstract artefacts, it also does justice to the idea that apparently empty names don't refer to anything, which makes it fully compatible with the machinery of chapters one and two.

Now we can look at our intuitions about truth values and validity in particular cases. First we can satisfy Fregean intuitions about non-substitutivity of co-referential names in intentional contexts. Lois doesn't believe Clark can fly, or that Clark is Superman, but does believe Superman can fly and Superman is Superman. This is because THAT[Fc] and THAT[Fs] need not refer to the same proposition, even if c and s refer to the same object, if they have different senses. The same goes for THAT[c=s] and THAT[s=s]. If we do want a name to appear in an intentional context transparently and with existential commitment, we can simulate this as follows:

Opaque: Believes(Lois, THAT[Flies(Superman)])
Transparent: ∃x[x = Superman & Believes(Lois, THAT[Flies(x)])]

We probably can't have transparency without existential commitment, but it is difficult for any non-Meinongian proposal to do that, since it involves having identity without an entity, given that only Meinongians have non-existent entities. As well as non-substitutivity, we can satisfy the disquotation principle, and almost get biconditional disquotation too. Biconditional disquotation does break down in Kripke-puzzle cases, however, since in some contexts Peter will reject 'Paderewski is musical', but he still believes that Paderewski is musical, in virtue of his other belief. This is because 'Paderewski' has the same sense in both contexts, unbeknownst to Peter. I argued in §2.13 that individuating public modes of presentation finely enough to deal smoothly with Kripke puzzle cases isn't really viable, so this restriction on biconditional disquotation is probably necessary. We still do much better on biconditional disquotation than the Millian.

We can also allow both positive and negative quantifying in, without being Millian. (This is actually the puzzle, from Sider [1995], that motivated the semantics in the original paper.) These would be formalized as follows:

Positive:

$$\frac{\text{Believes}[a, \text{THAT}\varphi b]}{\exists x[\text{Believes}[a, \text{THAT}\varphi x]}$$

Negative:

$$\frac{\neg\text{Believes}[a, \text{THAT}\varphi b]}{\exists x[\neg\text{Believes}[a, \text{THAT}\varphi x]}$$

The conclusion of each is true when the premise is, because the open sentence is true on the assignment assigning the sense of 'b' to x. We don't have existential commitment, however, in that the argument from 'S believes that n is F' to 'n exists' isn't valid. This combination is possible because the quantifier is sential and thus does not carry existential commitment. (The sense of ‘b’ could be non-denoting.) We can express that there is something existent which S believes is F, but this involves the existence predicate ‘E!’ (which is definable in terms of the identity predicate in the normal way, assuming the models don’t allow non-existent objects)

Non-committal: $\exists x[\text{Believes}(S, \text{THAT}[Fx])]$
Committal: $\exists x[E!x \;\&\; [\text{Believes}(S, \text{THAT}[Fx])]]$

The proposal also smoothly incorporates a semantics for intentional transitives which does not treat them as extensional. This is because we already have the modes of presentation in the models, so we can allow intensional or hyperintensional predications which are true when the mode of presentation stands in a certain relation, rather than when the object it denotes does. This relation will be, approximately, admiring (say) the object presented, via the mode of presentation in question. (This issue was addressed earlier, in footnote 38.) This satisfies intuitions about non-substitutivity and existential commitment, and makes it simple to express connections between objectual and propositional attitudes such as ‘Lois admires everyone she believes can fly’:

$$\forall x[\text{Believes}(\text{Lois}, \text{THAT}[\text{Can fly}(x)]) \rightarrow \text{Admires}(\text{Lois}, x)]$$

One problem is that the non-substitutivity intuitions about intentional transitives are not universal. We saw in §3.21 that Kripke is among the dissenters, and I promised to meet him at least halfway. Specifically, Kripke thinks these two arguments are valid:

(1) Schmidt admired Hitler.
Hitler was the most murderous man in history.

Schmidt admired the most murderous man in history.

(2) The Greeks worshipped Zeus.
Zeus is the tenth god mentioned by Livy.
____
The Greeks worshipped the tenth god mentioned by Livy.

(1) is easier to deal with. We can treat ‘the most murderous man in history’ in Russellian fashion, rather than a as referring expression, and have it appear outside the scope of the attitude. We can analyse the argument as follows:

(1*) $\mathrm{Adm}(s, h)$
$\exists x[\forall y[x=y \leftrightarrow y \text{ is a most murderous man in history}] \;\&\; x = h]$
____
$\exists x[\forall y[x=y \leftrightarrow y \text{ is a most murderous man in history}] \;\&\; \mathrm{Adm}(s, x)]$

Kripke says he doesn’t think that (1) has a reading on which it is invalid. And in fact, you can’t formalize it into our semantics in a way that makes it invalid, unless you treat ‘the most murderous man in history’ as a simple referring expression. We could have a device for turning definite descriptions into referring expressions, and then (1) would have an invalid reading, but if Kripke is right then perhaps we shouldn’t have one. I don’t think Kripke would be right to say that there was no invalid reading if we replace the description with a name. He doesn’t say this, and perhaps this points to a genuine difference between names and definite descriptions. Note that if we replace the definite description with an indefinite description like ‘a dictator’ then our intuitions easily fall into line with Kripke’s and there is little or no pressure to treat the description as referring and put it in the scope of the attitude.

(2) is harder, partly because it is unclear exactly what our intuitions should be, and partly because we have a definite description applying to an apparently empty name. To explore our intuitions, let's see if we really want referential transparency by trying it with two co-referring names:

(3) The Babylonians worshipped Hesperus.
Hesperus is Phosphorus.
____
The Babylonians worshipped Phosphorus.

I think there's a reasonably strong non-substitutivity intuition here, especially when you contrast 'the Babylonians worshipped Hesperus' with 'the Assyrians worshipped Phosphorus'. (As far as I know nobody actually worshipped either.) Others' intuitions may differ. But one way of having (3) come out invalid and (2) come out valid is to say that false gods exist in the way that fictional characters exist, and there is a special usage of 'worship' which we use to express the relation between people and the false gods that exist in virtue of their erroneous religious practices. (This could perhaps be integrated into the account of fictional characters given in the next chapter.) It won't be the normal use of 'worship' for when the object of worship exists, because substitutivity fails for that usage. I don't think that's a terrible solution. It is hard to know exactly what to say about the case though, because it is hard to know what the intuitions are that we want to satisfy. We would need to try it out with other attitude verbs, other referring expressions and other kinds of entity, and ideally use informants who aren't the people who will have to systematize the intuitions. I will have to leave that for further work.

Now we have seen that Fregean GPs and the associated semantics do quite well with the problems we were talking about. This motivates doing the metaphysical work to see if plausible and suitable entities can be found to play the role of modes of presentation, or object senses. I can't do all of that here, but I can say some things about what a proposal could look like.

We have a choice about how to classify object senses: metaphysically or semantically. A semantic classification would look at their meanings, or perhaps their meanings in different epistemically possible scenarios. Chalmers [2011] uses two-dimensional intensions in this role, which can be seen as representing the different intensions referring expressions might have in different possible scenarios.

A metaphysical classification of senses looks at the mechanisms presenting objects or determining meanings, rather than at the consequences of those mechanisms. Whether our semantics ultimately uses the semantic classification or not, we will probably need to think metaphysically about modes of presentation anyway to give the semantics respectable foundations. One example of a mechanism could be drefs, which do well when our main acquaintance with an object is via referential deference to better-informed users of its name. Another example could be a perceptual mechanism, when we are acquainted with an object through perceiving it, or when a group of people are acquainted with an object they can all perceive. (These modes of presentation are good candidates for being necessarily denoting when they exist.) A third example could be something corresponding to reference via definite descriptions. Perhaps there are mixed cases, where we have co-ordinated beliefs about an object justified by evidence from a variety of sources.

The main thing to worry about here is keeping our account general enough. We want to be able to ascribe beliefs to people outside our linguistic community, without existential commitment or referential transparency. We want to allow for non-denoting senses, but we may also want some senses to be *de re* and thus necessarily denoting.

Schellenberg's potentially non-denoting experiential modes of presentation look good for our purposes, but they are hostage to issues in the philosophy of perception quite far removed from what I have been doing here. There is work to do but there is reason for optimism: people do think about things and are acquainted with objects in more or less similar ways. This has to happen somehow, we already have some things to say about the mechanisms involved, and people are working on it.

3.5 Conclusion

This chapter was about attitude ascriptions, but that wasn't all it was about: it was about the kind of information we communicate using attitude ascriptions. That is information about what people are thinking, described in terms which often class different people as thinking the same thing. A standard way of thinking about this is to classify people has having the attitudes towards the same thought contents, and we do this using attitude ascriptions involving a word for the attitude and a 'that' clause referring to the proposition. Empty names cause a problem here, because it is usual to think of the objects a proposition is about as somehow entering into the proposition, but the propositions in question seem not to be about anything real. We had three options: keep the propositions and find some objects for them to be about; reject the propositions, and explain what is going on some other way; or keep the propositions without saying they are propositions about things.

§3.1 looked at some issues we might use to decide between the competing views. Propositional attitude ascriptions are supposed to communicate information about generalizations between people, both in what they think and how this links up to what say. We use propositions to understand these generalizations. Our account of empty names should do justice to this. There are also some intuitions about arguments involving substitution and quantification in attitude ascriptions, and we want to get the intuitive truth values and validities if we can. Also, a theory of attitude ascriptions should be able to smoothly accommodate objectual attitudes, ascribed using intentional transitives. Finally, we don't want an account that leaves us with implausible ontological commitments.

§3.2 looked at reifying the objects. This can proceed in two ways. We can have the objects be concrete, and much as people mistakenly believe them to be, except for being non-existent. This does quite well with the intuitions about validity and truth values, and can incorporate intentional transitives. It falls down on the plausibility of its ontology, and also could only do justice to role of propositions by demanding a revision to the treatment of chapters one and two, where we had referential failure, rather than reference to non-existents. The use of existent abstract objects is similar, although it scores better on ontological plausibility and worse on truth intuitions, especially with negative existential beliefs. It also holds that people are radically mistaken about the objects of their thoughts, which might sit badly with our view on how reference works, if it is anything like Imogen Dickie's.

§3.3 looked at the most ontologically parsimonious option: avoid reifying the constituents by rejecting the contents. We can do this with a pretence theory, an error theory or a paraphrasing strategy. This deals with attitude ascriptions themselves in a way which people may find more or less satisfactory, but it doesn't really get to the bottom of the problem. Attitude ascriptions involving empty names do seem to convey some true information about thinkers and how they cognitively relate to the world, and if this information is of the right kind then there is pressure to say that it is propositional. If we need the propositions anyway, then it makes sense to include them in our analysis of attitude ascriptions.

§3.4 looked at reifying the propositions but not the objects the propositions are about. We could opt for Braun's Russellian GPs, but this did not really capture enough of the information that seems to be conveyed, and Millian defences in the case of co-referring names did not straightforwardly carry over. I argued that it would be better to opt for an ontology of Fregean GPs, perhaps understanding the contents as metarepresentational, i.e. classified in terms of how they try to represent things rather than in terms of what they represent. There is more work to do on this, but early indications are promising, and if it can be made to work then it is probably what we should go for. Provided the metaphysics of Fregean GPs can be sorted out, we can understand the semantics as laid out informally above and formally in the appendix to follow. This semantics gives intuitive results about truth and validity, incorporates intentional transitives, and does justice to the theoretical role of propositions.

At this point we have a fairly self-contained proposal for treating names which are empty because they were introduced in the contexts of mistakes or lies. In non-intentional contexts we assign what I've called pessimistic truth values, preferably according to a two-valued negative free logic. In attitude ascriptions the truth values are less pessimistic. Nonetheless, the treatment we have so far would take the following sentences as untrue, while the man in the street will assent to all of them:

- Sherlock Holmes is famous.
- Sherlock Holmes lives at 221B Baker Street.
- Arthur Conan Doyle created Sherlock Holmes.
- Holmes is cleverer than any real detective.
- According to Doyle's stories, Holmes lives at 221B baker Street, but actually nobody does.

The sentences don't involve names from mistakes or lies. They involve names from fiction. The pessimistic truth values can't be accepted for sentences like these, at least not without some explanation. The next chapter will try to deal with fictional names in a more optimistic way. It will also show how to fit them into an account of attitude ascriptions, without contradicting what we have already done.

Appendix: Sential Semantics for Attitude Ascriptions

The system is adapted from the semantics in Bench-Capon [MSc]. The presentation has been streamlined, and predicates expressing intentional transitive verbs have been added. The basic elements are that expressions have senses relative to assignments, and there is a function from senses to objects. Sential quantification is primitive and objectual quantification can be defined in terms of it. A THAT operator exploits a function from the senses of sentences to the propositions the sentences express. The extension of an intentional transitive verb Ftu is a set of ordered pairs <x, y>, where x is the object denoted by the sense of t, and y is the sense of u. That is the basic shape; technical details follow.

We have a language L containing names, variables, quantifiers, (extensional) n-place E-predicates, (intentional) 2-place I-predicates, identity, truth-functional connectives, punctuation, a sentential operator 'THAT', and a predicate 'BEL' relating believers to the propositions they believe. Here is L's syntax:

- Names and variables are terms.
- If F is an n-place E-predicate or I-predicate and $t_1, \dots, t_n$ are terms, then $Ft_1 \dots t_n$ is a formula.
- If φ is a formula then ¬φ is a formula.
- If φ and ψ are formulas then (φ&ψ) is a formula. (Other connectives are defined in the usual way.)
- If φ is a formula then THATφ is a proterm.
- If t is a term and p is a proterm then BEL(t,p) is a formula.
- If t and u are terms then t=u is a formula.
- If φ is a formula and x is a variable then ∃xφ is a formula. (∀ is defined in the usual way.)
- If φ is a formula with no free variables, then φ is a sentence.

For the semantics, first we have a total function S from expressions of the language and assignments to the expressions' senses relative to those assignments. Sense are members of a set Σ. An assignment is a function from variables to senses in a subset α of Σ. If $S_A(x)$ is the same on all assignments, S(x) takes that value. Otherwise S(x) is undefined.

- Where t is a term, $S_A(t)$ must be a member of a privileged subset α of Σ. Note that α can have members which are not the senses of any name in the language. The members of α will be called *object senses.* α contains no n-tuples.
- Where t is a variable, $S_A(t)$ is A(t).
- Where t is a name, $S_A(t)$ is the same for all A.
- $S_A(F)$ is not in α or an n-tuple when F is a predicate, and it is the same for all A.
- Where F is an E-predicate, $S_A(Ft_1...t_n)$ is <ε, n, $S_A(F)$, $S_A(t_1)$, ..., $S_A(t_n)$>
- Where F is an I-predicate, $S_A(Ft_1t_2)$ is <ι, $S_A(F)$, $S_A(t_1)$, $S_A(t_2)$>
- $S_A(\neg\varphi)$ is <N, $S_A(\varphi)$>
- $S_A(\varphi\&\psi)$ is <C, {$S_A(\varphi)$, $S_A(\psi)$}>
- $S_A(t=u)$ is <I, $S_A(t)$, $S_A(u)$>
- $S_A(\text{THAT}\varphi)$ is <θ, $S_A(\varphi)$>
- $S_A(\text{BEL}(t,p))$ is <β, $S_A(t)$, $S_A(p)$>
- $S_A(\exists v\varphi)$ is <γ, {$S_B(\varphi)$: B(x) = A(x) unless x=v}>.

Σ therefore contains object senses, which are the senses of names; the senses of predicates, which we can call predicate senses, and set-theoretic constructions out of these and the place-holders ε, ι, N, C, I, θ, β, γ, and positive integers, which stand for extensionality, intentionality, negation, conjunction, identity, THAT, BEL, quantification and n-place predications.

S represents the function from expressions to senses. Now we represent the function from senses to references with a function V from Σ into a set $D \cup E \cup \{T, F\} \cup P$. D can be any set but will include the things names refer to. D plays the role which on an ordinary semantics is played by the domain of quantification. E is the set of n-tuples of members of D, i.e. the union of the sets D^n for each $n \geq 1$[45], so it can contain the extensions of predicates. T is truth and F is falsity. P is a set of propositions. V is allowed to be partial, to allow for non-referring names and possibly non-denoting predicates: those expressions will have sense but not reference.

- Where defined, $V(x) \in D$ if $x \in \alpha$. For every object x in D, $V(y) = x$ for some y in α.
- Where defined, $V(S(F)) \subseteq D^n$ where F is an n-place E-predicate.
- Where defined, $V(S(F)) \subseteq D \times \alpha$ where F is an I-predicate.
- $V<N, x>$ is T iff $V(x)$ is F; otherwise $V<N, x>$ is F.
- $V<C, \{x , y\}>$ is T iff $V(x)$ and $V(y)$ are both T; otherwise $V<C, \{x , y\}>$ is F.
- $V<\varepsilon, n, y, x_1, ..., x_n>$ is T iff $<V(x_1), ..., V(x_n)> \in V(y)$; otherwise $V<\varepsilon, n, y, x_1, ..., x_n>$ is F, including when either $V(y)$ or some of the $V(x_i)$ are undefined[46].
- $V<I, x, y>$ is T iff $V(x)$ is $V(y)$; otherwise $V<I, x, y>$ is F.
- $V<\theta, x>$ is the proposition represented by x (see below for this representation relation).
- $V<\beta, x, y>$ is T iff $V(x)$ believes $V(y)$; otherwise $V<\beta, x, y>$ is F.
- $V<\iota, x, y, z>$ is T iff $<V(y), z> \in V(x)$; otherwise $V<\iota, x, y, z>$ is F, including when $V(x)$ or $V(y)$ is undefined.

[45] We only really need to include D^n where there are n-place predicates in the language.

[46] This generates a two-valued negative free logic for empty names outside of intentional contexts.

- V<γ, Γ> is T iff V(x) is T for some member x of Γ; otherwise V<γ, Γ> is F.

I have used a notion of senses representing propositions, which may seem dodgy at first because the senses of formulas are supposed to be propositions, rather than just represent them. For this reason the S function need not quite be seen as a function from L to senses as traditionally understood. Since it is unclear how the proposition expressed by a sentence is determined by the senses of its parts, we can duck the question by representing propositions with set-theoretic constructions out of object and predicate senses, along with some placeholders N, C, I, θ, β, ι, ε and the natural numbers. We can say that when S(φ) represents a proposition *p*, φ expresses *p*. There will be some constraints on which senses represent which propositions:

- <ε, 1, y, x> represents an atomic proposition true iff the value of x, if any, instantiates the property (if any) corresponding to y, and false otherwise.
- <ε, n, y, x_1, ..., x_n>, where n is greater than 1, represents an atomic proposition true iff the values of x_1, ..., x_n (if any) stand in the relation (if any) corresponding to y, and false otherwise.
- <ι, x, y, z> represents a proposition true iff the value of y (if any) stands to the object sense z in the appropriate attitude relation (if any) corresponding to x, and false otherwise.
- <N, x> represents the negation of what x represents. (It is left open whether <N, N, x> and x represent the same proposition.)
- <C, {x, y}> represents the conjunction of the propositions represented by x and y. (We include the case where x=y.) This means a pair of propositions will only have one conjunction, so the order is not important. Distinguishing between the propositions that P&Q and that Q&P would complicate things but not in any important way. As with double negations, we leave open whether other equivalent truth functional compounds

express distinct propositions. It is also open whether <C, {x}> represents the same proposition as x (i.e. whether (P&P) expresses the same proposition as P).

- <I, x, y> represents an identity proposition between the values of x and y.
- <β, x, y> represents a belief proposition with the value of x as believer and the value of y as proposition believed.
- <γ, Γ> represents a quantificational proposition true iff one of the propositions represented by the members of Γ is true.

Note that although atomic sentences containing empty names in non-intentional contexts are false, BEL(a, THATφ) can still be true even if φ contains an empty name, as can Ftu where F is an I-predicate. It is also worth pointing out that the identity relation still holds between objects rather than senses, so it is possible that V(S(t=u)) is T even if S(t) is not S(u). Co-referring names will not in general be substitutable *salva veritate* in intentional contexts, though in non-intentional contexts they will be. Synonymous names, if there are any, will have the same senses, so they will be substitutable everywhere. (If we want Russellian propositions we constrain models by having only one non-denoting sense and only one sense for each object, which will make co-referring names substitutable *salva veritate* everywhere. If we want Russellian GPs too then we can have only one non-denoting object sense.) Finally, if we have an existence predicate the value of whose sense is D (the domain of objects, or if we are Meinongians then its existent subset), then that will give the right answers for positive and negative existentials.

That concludes the exposition of sential semantics and quantification. Now we can show how to use the sential quantifier to define an objectual quantifier, i.e. one which is existentially committing and ignores distinctions of mere sense, even in intentional contexts. We have to add something to secure the existence requirement, and substitute something for the BEL predicate and I-predicates in the scope of the quantifier which ignores distinctions of mere sense. The other predicates already ignore the distinctions. In general, $\exists_o x\varphi x$, where $\exists_o$ is an objectual quantifier, can be taken to abbreviate this:

$$\exists x(E!x \mathbin{\&} \varphi'x),$$

where 'E!' is an existence predicate, and $\varphi'x$ is φx except with all occurrences of BEL(t, THATψx) replaced by $\exists z(z=x \mathbin{\&} \text{BEL}(t, \text{THAT}\psi z))$, and all occurrences of Ftx where F is an I-predicate replaced with $\exists y(y=x \mathbin{\&} Fty)$.

We may want to express a relation like identity that also applies to non-denoting senses, at least relating non-denoting senses to themselves. This would allow us to simulate existentially non-committal numerical quantifiers. We can partition α into equivalence classes $[x]_R$ determined by this relation and have a non-extensional predicate '≈' to express it:

- If t and u are terms then $t \approx u$ is a formula.
- $S_A(t \approx u)$ is $<Z, S_A(t), S_A(u)>$
- $V<Z, x, y>$ is T iff $[x]_R$ is $[y]_R$; otherwise $V<Z, x, y>$ is F.

$<Z, x, y>$ represents a proposition true iff $[x]_R$ is $[y]_R$ and false otherwise.

Chapter 4 – Fictional Names

4.0 Introduction

Suppose that in an ordinary context I say ‘Eddy Merckx was Belgian’. To find out whether this is true, you can examine the list of all the Belgians who have existed. If my usage of ‘Eddy Merckx’ refers to someone on the list then the utterance is true, and otherwise it is false. This is consistent with the two-valued negative free logic proposed in chapter one, so ‘Santa was Belgian’ comes out false when uttered by someone who erroneously thought ‘Santa’ was the name of a Belgian. However, suppose that in a similarly ordinary context I say ‘Poirot was Belgian’. If you list all the Belgians who have existed, none of the people on the list is referred to by my usage of ‘Poirot’ either, so treating my assertion in the straightforward way will make it come out false too.

Perhaps it is not the end of the world if we say it is false. After all, the stories about Poirot from which we get the information that he was Belgian are not true stories. Something should be said though, to account for the fact that people saying ‘Poirot was Belgian’ may do so in apparent seriousness, without making a mistake. Also, if they say in the same mode ‘Poirot was French’, they are making a mistake, but it is not the same kind of mistake which a child might make if they said ‘Santa is coming’. It is, on the face of it, closer to the kind of mistake someone might make if they said ‘Eddy Merckx was Dutch’. The problem with ‘Poirot was French’ is not one of referential failure; it is just one of getting the facts wrong. If we get the result that the sentence is false just from the referential failure of ‘Poirot’, we have the wrong explanation. We need an explanation which gets the right results for the right reasons. Even if we ultimately decided that my utterance of ‘Poirot was Belgian’ was not true, we would still need to show how to distinguish it from ‘Poirot was French’. In large part, this amounts to distinguishing the good assertions from the mistakes, and since mistakes like ‘Poirot was French’ do not always spring from referential failure, the treatment of the preceding chapters is inappropriate.

Those chapters treated empty names introduced in the contexts of mistakes and lied as similarly as possible to cases where reference succeeds, because that is what the speakers and thinkers are trying to do. We tried to work out what happens to the treatment designed for successful reference when reference fails. In discourse within and about fiction, however, it seems that sometimes we are not even trying to refer in a straightforward way. This goes some way towards explaining why we should not be surprised that the straightforward treatment gets the wrong results. There are three main strategies for getting the right results.

One strategy says that the utterances can be taken straightforwardly after all. The problem with the account of ‘Eddy Merckx was Belgian’ and ‘Poirot was Belgian’ was that it told us to look at the list of Belgians who have existed. If we looked at the non-existent Belgians too, according to this strategy, we would find the referent of ‘Poirot’ among them, and he would be Belgian and not French, so ‘Poirot is Belgian’ is true and ‘Poirot is French’ is false. Drop the prejudice in favour of the existent and we can take the utterance at face value. This strategy is associated with Meinong [1960], and it is defended more recently in its most unadulterated form by Terence Parsons [1980].

The second strategy takes utterances like ‘Poirot was Belgian’ to be asserting a different content from the one they seem to have at face value. There is more than one way of doing this. We could take the sentence to have a tacit fictionality operator: ‘*According to Agatha Christie’s stories*, Poirot was Belgian’. This is associated with David Lewis [1978]. Alternatively, we might be happy to reify Poirot as an abstract object but baulk at the idea of accounting for fictional discourse by reifying additional Belgians, instead saying that ‘Poirot is Belgian’ asserts that some relation other than instantiation holds between Poirot and being Belgian. This is associated with Peter van Inwagen [1977], Ed Zalta [1983, 2000] and Amie Thomasson [1999].

A third strategy takes ‘Poirot is Belgian’ not to be an assertion at all. We are only pretending to assert. This may involve pretending that ‘Poirot is Belgian’ has a semantic content when actually it does not, which saves us the trouble of finding a content for it, so we do not need to include Poirot in our ontology as a constituent of that content. The pretence strategy is associated with Kripke [2011b, 2013], Kendall Walton [1978a, 1978b, 1990] and Gareth Evans [1982: ch. 10].

The acceptance of an ontology of fictional characters can be incorporated into any of these strategies, although it is more integral to some than others. Not all unmistaken discourse involving fictional names is like ‘Poirot is Belgian’, though. This can be taken as an utterance *within fiction*, being the kind of sentence which might appear in a novel. We also have utterances *about fiction*, such as ‘Poirot was created by Agatha Christie’. Possible intermediate cases include utterances asserting relations between fictional characters and real things: ‘Poirot is shorter than Obama’ and ‘some real detectives admire Poirot’. These can be mixed in various ways. Different kinds of utterance generate different problems and lend themselves more readily to different treatments. I will argue for two different analyses, one primarily for uses about fiction, and one primarily for uses within fiction. For mixed cases we will have to mix the analyses.

§4.11 considers an argument from van Inwagen [1977] for an ontology of fictional characters, and defends it against some objections due to Takashi Yagisawa. I conclude that van Inwagen’s argument is a good one, but that it only shows that there are fictional characters, and does not show what they are like. In particular, it does not show that they are the way van Inwagen thinks they are. The §4.12 and the whole of §4.2 are primarily exploratory in character, looking at the kinds of views which other people have put forward, and looking at the problems which my positive view should be able to solve. I will evaluate and criticize positions along the way, but my positive view is laid out in §4.3 and §4.4. §4.12 outlines some different accounts of what fictional characters might be, with particular reference to Amie Thomasson, Ed Zalta and Terence Parsons. §4.2 outlines several problems which any ontology of fictional characters must deal with. In §4.3 I give my positive view of discourse about fiction, by sketching a systematic fictional ontology which attempts to address these problems. In such discourse, fictional names will refer to objects from this ontology. §4.4 presents a different, pretence-theoretic account of discourse within fiction, and proposes a different semantics for this use of fictional names which develops the account of truth in fiction given by Lewis [1978]. At the end of the chapter I show how my view of fictional names behaves in attitude ascriptions, and argue that the machinery I develop for fictional names as used within fiction would not unproblematically extend to the names from mistakes and lies, and the treatment of those should be left much as it was in the first three chapters.

4.1 Fictional Ontologies

- 4.11 Van Inwagen's Argument

Van Inwagen [1977] attempts to establish two main conclusions, which should be kept separate. First, that there are fictional characters, or equivalently that there are such things as fictional characters. Second, that these fictional characters exist. To establish these conclusions, he argues for the conditional that if there are fictional characters then they exist, and then that there are fictional characters. We will examine the arguments separately.

- *4.111 'If there are fictional characters then they exist'*

Van Inwagen argues that if there are fictional characters then they exist, on the grounds that everything exists. He is explicit [1998] that his stance on metaontology is heavily influenced by Quine [1948], and following Quine he takes 'there are *F*s' and '*F*s exist' to be equivalent on the only readings he understands. This makes the claim that there are non-existent fictional characters either contradictory or unintelligible. Lewis [1990] adopts a similar position towards the Meinongian 'noneism' defended by Richard Routley [1980] and subsequently Graham Priest [2005]**.** Lewis says that anti-Meinongians like himself should take noneists as holding that all the things they say there are exist, even though they say they do not. Arguments pleading incomprehension occupy a strange position dialectically. On the one hand, van Inwagen is not going to convince anyone who thinks they understand a consistent reading of 'there are non-existent fictional characters'. On the other, if he cannot understand his opponents' claim as anything other than contradictory, even after making reasonable attempts to do so, this gives him reason to believe they are wrong, and gives other people in the same position reason to agree with him. For present purposes, we will take the Meinongian claim that it is consistent for there to be non-existent things seriously. This means having a logic which distinguishes between existence and being. We can take ‘there are’ as equivalent to ‘∃’, and not existentially committing, and translate ‘exists’ as a monadic predicate, whose extension need not be the whole domain. In this language, the debate between Meinongians and their opponents becomes a debate over whether ‘there are non-existent things’ is true. This claim is translated as $\exists x(\neg Exists(x))$. Van Inwagen may think this expresses something logically impossible, but since its negation is not a theorem of the system in which we are conducting the debate, that argument will be inadmissible. With these rules of debate in place, we can try to argue that the Meinongian position is unmotivated even if it is intelligible. This is what we will do, although the van Inwagen/Lewis position is still open if the arguments we put forward here are found wanting. We now turn to Van Inwagen’s argument that there are fictional characters, which does not violate the rules of the debate we have set up, and can be used by Quineans and Meinongians alike.

- *4.112 ‘There are fictional characters’*

‘Are there such things as fictional characters?’ is a question about ontology. Van Inwagen takes it that we have a fairly well established method for answering questions about ontology, which is mostly due to Quine. We take our best theories about the world, and paraphrase them into a canonical first-order language. Then we see what must be in the domain of quantification for these theories to be true. For example, if our canonically paraphrased theory contained the sentence $\exists x$(x is a dog), this could only be true if there were dogs, and we would thus be committed to there being dogs.

We need some constraints on what counts as an adequate paraphrase, or we could just paraphrase our whole theory as a sentential constant and not commit ourselves to anything. Much could be said about this, but instead of arguing about methodology in ontology here we will address van Inwagen's argument on his own terms. He places the following constraint on an adequate paraphrase:

> (LC) An adequate paraphrase must not be such as to leave us without an account of the logical consequences of (the propositions expressed by) the paraphrased sentences. [1977: 304]

The way he sees this going is that, in the canonical language, formal consequence (i.e. consequence in a system such as first-order logic) and logical consequence will line up. To see how this works, consider two candidate examples of logical consequences which are not formal consequences:

Phosphorus is bright.
____
Hesperus is bright.

Fred is a bachelor.
____
Fred is a man.

The first inference is not formally valid because two different names are used, so although the premise and conclusion would express the same Russellian proposition, this is not guaranteed just by uniformly interpreting the non-logical vocabulary in the argument. The canonical language could sort this out by only having one name for each object. The second inference is not formally valid for a similar reason, but in the canonical language we could translate 'bachelor' as 'unmarried man', and then it would be formally valid. There is room for arguing over both these particular cases, but they illustrate the idea. Within this methodology, van Inwagen argues that there are fictional characters, on the grounds that we can make inferences like that from (1) to (2):

> (1) There are characters in some nineteenth century novels who are presented with a greater wealth of physical detail than is any character in any eighteenth century novel.
>
> (2) Every female character in any eighteenth century novel is such that there is some character in some nineteenth century novel who is presented with a greater wealth of physical detail than she is. [van Inwagen 1977: 302-3]

Van Inwagen thinks that systematic paraphrases of (1) and (2) which did not quantify over fictional characters would be messy if they were possible at all, which he doubts. He says that the most promising paraphrases would quantify over the names of fictional characters, although he does not see how to do it. We will discuss two such attempts at paraphrase later in this section, of which one is Yagisawa's and one is new.

We do perhaps have some grasp of the relations which the classes of eighteenth and nineteenth century novels would have to stand in to each other for (1) and (2) to be true. As such, we could coin words to express these relations. He suggests 'dwelphs' and 'praphs', so the class of nineteenth century novels dwelphs and is praphed by the class of eighteenth century novels. (We could also use plural reference to put it in a way which eliminates talk of classes: the nineteenth century novels dwelph* and are praphed* by the eighteenth century novels[47].) These

[47] This is actually slightly different in truth conditions because one or both of the centuries might be novel-free. Strictly we should paraphrase it as 'there are some nineteenth century novels and either they dwelph* and are praphed* by the eighteenth century novels or there are no eighteenth century

words could be used to give unsystematic paraphrases of (1) and (2), but this leaves it unexplained why (1) logically entails (2). It seems the only way to explain it is to give paraphrases of ‘dwelph’ and ‘praph’ in terms of fictional characters. If the paraphrases revealing the ontological commitments of our assertions should also reveal the logical relations between those assertions, as LC demands, then this would mean (1) and sentences like it commit us to the existence of fictional characters. Since some sentences like (1) are true, there must be fictional characters.

This argument did not involve any mention of fictional names, but it can be extended to cover them. Consider these sentences:

(3) Gulliver is a character in an eighteenth century novel.

(4) Some character in some nineteenth century novel is presented with a greater wealth of physical detail than Gulliver.

(1) and (3) jointly logically entail (4), and if ‘Gulliver’ was not the name of one of the fictional characters (1) quantifies over and whose existence van Inwagen argues for, then they would not. So if van Inwagen’s argument is right, ‘Gulliver’ is not an empty name in (3). It refers to a fictional character. Perhaps it does not always refer to one, but in (3) and similar sentences it must to secure the entailment, and this should be enough to secure van Inwagen’s conclusion.

Van Inwagen’s view is that fictional characters are abstract objects, of the same kind as plots, meters, rhyme schemes and so on. He calls these

novels. See Boolos [1984, 1985] for more on plural reference and its uses in nominalist paraphrasing. It is also worth noting that van Inwagen [1990: §2] is on board with plural reference too. See Lewis [1991: §3.2] for an influential discussion of the innocence of plural reference and quantification.

things *theoretical entities of literary criticism* [1977: 302-3]. He also says they could not be merely possible objects, because they are actual [1977: n. 11]. Gulliver actually is a fictional character.

Nothing in his argument requires that fictional characters be actually existing abstracta, although if we took them to be mere possibilia we might still have trouble paraphrasing (1)-(4), since mere possibilia are not ordinarily taken to be in the domain of quantification for sentences outside the scope of any modal operator. Aside from this possible caveat, they could be anything, and in particular they could be a quite different kind of thing from the other things he counts as theoretical entities of literary criticism. Fictional characters could still be either concrete or abstract, and the argument leaves open whether they are created by their authors or whether they always existed and their authors selected them by choosing to write the stories they did. We will look at some competing accounts of what fictional characters are in the following three sections, but first we will consider some objections to his argument.

- *4.113 Yagisawa's objections*

Takashi Yagisawa [2001] objects to a few things that creationists about fictional characters tend to say, but two of his arguments are particularly pertinent to van Inwagen's position. One (§6) is very simple: being fictional entails not existing. Van Inwagen could avoid this problem by saying that there are fictional characters but they do not exist, but van Inwagen explicitly makes this Meinongian response unavailable to himself, and in any case if the view that there are fictional characters can only be accepted in conjunction with Meinongianism then we should be aware of it.

Yagisawa also examines two non-Meinongian replies and finds both wanting. As Yagisawa notes [2001: n. 40], van Inwagen mentions both of the replies [1977: n. 11] and takes one to be a more precise version of the other. Yagisawa takes them to be separate and treats them separately, and we will do the same.

The first reply paraphrases ‘Poirot does not exist’ as ‘nothing has all the properties ascribed to Poirot in the stories’[48]. Yagisawa rejects this because even if someone did have all these properties, ‘Poirot does not exist’ would still be true. We will grant him this for now, although the issue is discussed in more depth in §4.22 in relation to Kripke’s argument that (in a sense) fictional characters could not have existed and fictional kinds could not have been instantiated.

The second reply paraphrases ‘Poirot does not exist’ as ‘there is no such man as Poirot’. In general, the strategy says that fictional characters will come with a sortal property, and when we assert that they do not exist we will be saying that they do not fall under it. Yagisawa [2001: 169] rejects this because we cannot provide a suitable sortal to use in paraphrasing ‘boojams do not exist’, and in any case it does not extend in any obvious way to ‘no fictional individual exists’ and the like[49]. The boojams problem seems soluble, if in a fairly unsatisfying way: it seems reasonably clear that boojams are meant to be animals, so we could say ‘there are no such animals as boojams’. Perhaps this will not do, and perhaps a better example could be found. In any case, Yagisawa’s problem of paraphrasing ‘no fictional individual exists’ is more acute.

[48] ‘Ascribed’ is a technical term which van Inwagen introduces [1977: 305] to refer to the relation between stories or parts of stories, fictional characters and the properties those characters have in those (parts of) stories. We will discuss ascription some more shortly.

[49] ‘Boojam’ (sometimes ‘boojum’) is a word used in Lewis Carroll’s poem *The Hunting of the Snark*. A boojam is supposed to be a kind of thing, but there is not much information in the poem about what kind.

More precisely, the suggestion is that we paraphrase ‘no fictional *F* exists’ as ‘there are no such *G*s as fictional *F*s’, where *G* is some sortal which *F*s all fall under. For the paraphrase to be true, no fictional *F*s may be *G*s. So ‘no fictional individual exists’ must be paraphrased as ‘there are no such *G*s as fictional individuals’, where *G* is some sortal which all individuals fall under. But fictional individuals are individuals, since everything is. This means that there will be no suitable choice of *G* available.

Consider also the case of fictional fictional individuals: the characters in fictions which stories refer to but do not fully recount.[50] One might try to evade the problem by saying that fictional fictional individuals are really just fictional individuals, and it does not matter whether the fictions they appear in turn up within fictions or not. This probably will not work. A fictional fictional individual will be ascribed having properties ascribed to it. Double ascription is not the same as ascription, because while an object will always have a property either ascribed to it or not, there will be double ascription gaps. Gridley Quayle might be neither ascribed being ascribed baldness nor be ascribed not being ascribed baldness.

Since fictional fictional individuals are fictional individuals, we will want a true paraphrase of ‘no fictional fictional individuals exist’. According to the suggestion under consideration, this will be paraphrased as ‘there are no such *G*s as fictional fictional individuals’, where *G* is a sortal which fictional individuals fall under. Since fictional fictional individuals are fictional individuals, they are *G*s too, and this

[50] An example would be Gridley Quayle, the hero of a series of detective stories written by one of the characters in Wodehouse [1976]. A more common example is Gonzago, the victim in a play within a play in *Hamlet.* However, one might complain that the play is fully performed during the performances of the play. It is however definitely true in Wodehouse’s story that the Gridley Quayle stories contain more detail than Wodehouse tells us about.

paraphrase will be false. There are probably some unsatisfying ways of dealing with attributions of non-existence to fictional characters[51], but a satisfactory solution will have to wait until §4.431.

We now move to Yagisawa's second objection [2001: §4]. Van Inwagen wants us to take sentences like (1)-(4) at face value, as assertions of literary criticism. However, he also wants us not to take sentences like 'Gulliver visited a flying island' and 'Poirot has a moustache' at face value. They are just as much platitudes of literary criticism, but taken at face value they are false. For Yagisawa this is fine: he holds [2001: 163-4] that literary criticism aims at improving our appreciation of literature rather than accurately describing the world and we should not expect its claims to be true. However, if we take that attitude towards literary criticism, we need not take (1)-(4) at face value either, and this means we have no platitudes from which to infer that there are fictional characters.

Yagisawa's point about the function of literary criticism should not be swallowed uncritically. At first pass it seems he might be on to something: maybe literary criticism and literature form a nice self-contained symbiotic package which does not impinge on the real world. Literary criticism is just there to supplement the literature, and factual accuracy is equally unimportant in a novel and in a work of literary criticism. Is that right? Well, perhaps Yagisawa is right that literary

[51] I suggest three. (1) We take the first paraphrase, that nothing has all the properties ascribed to any fictional object, and stipulate that fictional characters all have non-actuality, or non-identity to each actual object, ascribed to them. (2) We take the first paraphrase and say that, contrary to Yagisawa's intuition, an actual person having all the properties ascribed to Poirot is precisely what Poirot existing would amount to. (3) We deny the premise that fictionality entails non-existence, taking van Inwagen (standing on Quine's shoulders) to have discovered that they do exist, just like everything else.

criticism exists to increase our appreciation of literature, or primarily for that. It does not follow from this that it does not mean to state facts, though. Consider a book on car maintenance. That exists to help people maintain their cars, but it still manages to state facts about how cars work and what happens when you do things to them. Is this a mysterious coincidence, that when someone writes a book that helps people maintain their cars they find themselves stating facts? No. The book helps people maintain their cars because the things it says about cars are true. Perhaps literary criticism works the same way. Suppose a critic writes that a novel has a romantic subplot, or has two heroes. Why do these increase our appreciation of the novel? The obvious answer is that they are true, and readers will understand the book better with the information than without. But if the critic's statements are true, then van Inwagen's argument applies. Perhaps there is another way to explain why literary criticism achieves its aims, but the obvious explanation is that it does so by stating facts. Or at least, enough of it does to run van Inwagen's argument. Perhaps some of the more literary literary criticism does not work by stating facts, but CliffsNotes and the like probably do.

Yagisawa would need to do more to dismiss literary criticism as false, then. But even if Yagisawa seizes the wrong horn of his dilemma, both horns create *prima facie* trouble for van Inwagen. If we accept the platitudes of literary criticism as true, van Inwagen's theory is false, since it denies the ones like 'Gulliver visited a flying island'. If we do not accept them, van Inwagen loses his data.

I do not think Yagisawa is properly addressing van Inwagen's argument, however. Van Inwagen says that it is a truth of logic that (for example) (1) entails (2), and despairs of finding a paraphrase which preserves the inference. As such, he thinks we should take (1)-(4) at face value. He does not despair of finding a paraphrase of 'Gulliver visited a flying island', so we do not have to take that at face value. The paraphrase he offers uses a relation of *ascription* between properties, fictional objects and pieces of literature. He takes ascription as primitive, but thinks we have a fair grasp of it: fictional objects are ascribed the properties they have in the stories they come from. Pieces of literature – he calls them *places* – include works and parts of works. Real objects appearing in fictions do not have properties ascribed to them according to van Inwagen: ascription just holds between fictional objects, the stories to which they are native, and the properties they have in those stories. As such, it is not just a relation holding between a property, a thing and a story according to which it has that property, since real characters can have properties in stories too[52]. Ascription is a *sui generis* relation grasped by anyone who understands 'Poirot is Belgian'. I agree that we have a fair grasp of the relation and that van Inwagen is justified in taking it as primitive for the purposes of his discussion, though it may admit of further analysis in principle, and clearly there is room for debate as to exactly what properties are ascribed to which characters in which places. Now, 'Gulliver visited a flying island' is paraphrased as:

FG $\exists x$[Ascribed(visiting a flying island, Gulliver, x)]

Playing according to van Inwagen's rules, this is an adequate paraphrase only if it satisfies LC. So, what are the logical consequences of 'Gulliver visited a flying island'? One obvious candidate is 'Gulliver visited an island'. That is paraphrased as follows:

IG $\exists x$[Ascribed(visiting an island, Gulliver, x)]

[52] These properties will presumably not always be purely qualitative: Holmes is ascribed living in London; but London, being a real city, is not ascribed being home to Holmes or anything else. As such we will have to cash out 'Holmes lives in London' as the ascription to Holmes in Doyle's stories of the monadic object-involving property of living in London, not the two-place ascription to Holmes and London of the dyadic qualitative relation of inhabiting.

IG is not a formal consequence of FG, so van Inwagen needs 'Gulliver visited an island' not to be a logical consequence of 'Gulliver visited an island' either. Is it? Well, it is an open question whether ascription is closed under (multi-premise) logical consequence. The notion of logical consequence involved here might be non-classical to deal with inconsistent fictions, but pretty much any notion of logical consequence will have visiting an island being a consequence of visiting a flying island. Certainly it seems like any story in which someone visited a flying island would have them visit an island, at least if the story did not explicitly say that they did not. We must be careful however to distinguish two questions. One is whether ascription is in fact closed under logical consequence, and another is whether its closure under logical consequence is a truth of logic. Only the latter causes van Inwagen a problem, because only the latter makes 'Gulliver visited an island' a logical consequence of 'Gulliver visited a flying island'.

To see the difference, suppose we set up a machine which systematically prints out the logical consequences of a set of first order axioms one by one, and leave it running for ever. We could make sure it missed none out, by having it run through an effective enumeration of finite sequences of sentences of the language and check whether they are proofs from the axioms according to a sound and complete axiomatic proof procedure. The output of the machine will be closed under logical consequence, but that it is so closed will not be a truth of logic; it will be a contingent truth dependent on the way the machine was set up[53].

[53] In case some readers are worried, I should point out that this sort of machine does not contradict Gödel's incompleteness theorem or Church's undecidablity theorem. It would not violate incompleteness because for some sentences φ the output theory need not contain either φ or ¬φ, and it would not violate undecidability because the machine would never be finished.

Now, it is possible that ascription is closed the way the output of the machine is closed, not as a matter of logic but because of the conventions governing the interpretation of literature. Is it likely that it is closed under consequence as a matter of logic? The answer to this question depends on the metaphysical story we accept about fiction. If we thought, with the Meinongians, that for 'Gulliver visited a flying island' to be true someone had to visit a flying island, we might well think that as a matter of logic someone would have to visit an island too[54]. If we thought, with van Inwagen, that for 'Gulliver visited a flying island' to be true certain literary practices have to take place but no flying islands need get involved, then whether it would have to have 'Gulliver visited an island' as a consequence would depend on the conventions governing those practices. Fine [1982: 116] says that there could be a practice of *inert literature* whose conventions were that anything true in the work was stated explicitly. If he is right, van Inwagen is safe, and if Fine is wrong for some reason other than logical necessity, then van Inwagen is still safe.

Van Inwagen's story looks coherent, then. He can coherently maintain that if ascription is closed under some kind of consequence then that is a contingent fact dependent on how our literary practices work. If Gulliver's visit to a flying island entails a visit to an island, that is because we have decided, collectively and implicitly, that logic holds in his world, not because independently of our decisions logic holds in

[54] Meinongians do not have to think this though: they may hold that logical completeness and consistency are constraints applying only to the existent. Parsons [1980: 19] holds this (see §3.123, below). Zalta [1983] holds that abstract objects must be consistent and complete in the properties they exemplify by not in the properties they encode, and Gulliver only encodes visiting a flying island. He does in fact encode visiting an island too, but this is not a logical consequence of his visiting an island. See §3.122 for more on Zalta's theory.

ours. Van Inwagen's paraphrase does not fall foul of LC on those grounds.

There are however some other kinds of consequence of ascription sentences which van Inwagen will need to preserve. For example, from 'Gulliver visited a flying island' and 'Gulliver is a character in an eighteenth century novel', we get 'a character in an eighteenth century novel visited a flying island'. Van Inwagen has no trouble with this. Here are the paraphrases:

> ∃x[Ascribed(visiting a flying island, Gulliver, x)]
>
> Character in an Eighteenth Century novel(Gulliver)
>
> ∃x[Character in an Eighteenth Century novel(x) & ∃y[Ascribed(visiting a flying island, x, y)]]

Ultimately we should not be so surprised that van Inwagen's paraphrase gets the right entailments if we think that his story about fictional characters is a credible one. If assertions like 'Gulliver visited a flying island' are about literature and not about islands, van Inwagen's paraphrase is more perspicuous. It gets closer to the truth, and so it should get closer to the right logical consequences. Of course, if we do not accept his metaphysical story then we might well not accept his paraphrase either. In general, we can expect paraphrases to fall foul of his condition LC when they are trying to eliminate ontological commitments by brute force, but not when they are trying to get at what is really being talked about.

So Yagisawa's dilemma should not trouble van Inwagen. Yagisawa [2001: 165-7] has another response to the argument, however, which is to offer paraphrases of van Inwagen's data, i.e. sentences like (1)-(4), which do preserve the logical inferences. Following van Inwagen's suggestion, he understands the paraphrases as quantifying over the terms used to refer to fictional characters. If these paraphrases satisfied condition LC, much of the force of van Inwagen's argument would be lost, but perhaps some would remain. He could hold that the paraphrases in terms of names of fictional characters are implausible, and that even if they can be made to give the right truth values, they do so only artificially and by changing the subject. A related response to a paraphrase is to say that even if when you paraphrase something you provide a different way of saying it, the original is still true. Here is Kripke:

> In ordinary language, we very often quantify over fictional characters. Perhaps such quantification could be eliminated if it were always possible to replace the original (quantified) sentence with a sentence describing the activities of people. [Footnote: Nevertheless, it is true that there are fictional characters with certain properties, and anyone who denies this is wrong.] [Kripke 2011b: 63]

I am quite sympathetic to Kripke's attitude here, and if a viable paraphrase was produced then a proper defence of it would be worth investigating. That style of argument does however risk resting heavily on intuitions, which would need to be balanced against other inputs to our reflective equilibrium, such as intuitions about the ontological queerness of fictional characters. Van Inwagen's position certainly seems stronger if there is no competing paraphrase in the game. In view of this, let us examine Yagisawa's. To get a sense of how he envisages the paraphrases going, here is his paraphrase of (1):

> There be (apparent) singular terms, t_1, t_2, …, t_k ($1<k$), in some 19th-century novels such that for any (apparent) singular term t_m in any 18th-century novel the accompanying predicates for t_1, t_2,

> …, t_k exhibit a greater wealth of physical detail than the accompanying predicates for t_m. [Yagisawa 2001: 165]

What is wrong with this? Well, one charge which metalinguistic paraphrases always have to answer is that of changing the subject. Do statements about the detail with which fictional characters are described entail the existence of predicates and referring expressions? Is this entailment logical? Perhaps the case can be made more plausibly here than it can in metalinguistic analyses of attitude ascriptions, for example the one Church [1950, 1954] attributes to Carnap [1947: §§13-15] and criticizes using his translation test, which I described in §1.13. Copies of novels and the sentences in them are linguistic tokens, and perhaps Yagisawa can claim that discourse about fiction really is discourse about language. Since literary translation is a tricky topic, perhaps we could also say something to spare Yagisawa's paraphrases embarrassment at the hands of the translation test.

A problem which is probably less tractable is that characters can be referred to with more than one singular term (orthographically individuated); and singular terms, even proper names, can be used to refer to more than one character in the same work. *Wuthering Heights* is a particularly nasty case (spoilers[55]), but the phenomenon is ubiquitous and stops us replacing fictional characters either with singular terms or equivalence classes of singular terms. It would be foolhardy to say categorically that Yagisawa's proposal could not be patched up without committing to characters or something just as

[55] 'Mr Earnshaw' refers to three people, 'Mr Heathcliff' to two, and 'Linton' to three, one of whom is a Mr Heathcliff; two people are sometimes called 'Catherine' and sometimes 'Cathy', there are two people called 'Mrs Heathcliff', and three called Mrs Earnshaw. One Catherine/Cathy is variously a Mrs Heathcliff and a Mrs Earnshaw. All are of course referred to using various pronouns, and a system of nested narrators ensures that several different characters are also referred to as 'I'.

unpalatable, but as it stands the proposal will not do, and there are reasons to think the prospects are bleak.

We would presumably want to individuate the (apparent) singular terms in some non-orthographic way, perhaps along the causal-historical lines in Kaplan [1990], but without reifying the causal chains and histories themselves. The problem is that we want the singular terms to be really individuated, because we are really quantifying over them, but we do not want to reify the things individuating them. Of course it would be quite extreme to reify the causal chains and histories involving fictional characters, where one points at another and says 'let's call him NN', since none of this pointing goes on. (It is fiction.) The problem goes deeper, though. We do not even want to reify things like novelists' artistic creations, and maybe not even literary practices corresponding to fictional characters, since in doing so we would be reifying theoretical entities of literary criticism, which would either be van Inwagen's creatures of fiction themselves or stand-ins which save nothing in ontology and just make the paraphrases uglier. Van Inwagen is not reifying flying islands or giants or anything like that; he is just committing to the objects with reference to which we describe literary practices.

There is a reply open to Yagisawa. Whatever van Inwagen says is true about fictional characters, we can duplicate it without ontological commitment by using a positive free logical[56] theory parasitic on van Inwagen's theory. Van Inwagen thinks, presumably, that the way the fictional characters are supervenes on the way the concrete part of reality is. We can say that, rather than generating a realm of fictional entities, this creates some singular terms and determines the truth values of sentences containing them. These singular terms need not be

[56] A positive free logic is one where atomic formulas containing empty names need not always be false.

created *ex nihilo*; the generation process could cause some objects from elsewhere in the ontology, e.g. sets, to count as singular terms. This option is available to anyone who holds that there are enough things (of whatever kind). Then we can understand the quantification substitutionally. This ought to successfully mimic the results van Inwagen gets, without committing to fictional characters. We can even modify van Inwagen's results slightly if we like, for example by having '*n* exists' be false where *n* is one of the new singular terms.

An immediate response is that if we were allowed to do this sort of thing then we could do it everywhere and never have to commit ontologically to anything. This isn't right, though. We still need to be realistic about the determining base facts, or there is nothing to generate the language and determine the truth values of the sentences. The reason we can be parasitic on van Inwagen's theory is that the nature of the fictional realm is determined by that of the concrete realm, so we can extract something equivalent to the whole theory while only committing ontologically to a part of it. The strategy does generalize, but only to other theories where one part is determined by another. For example, if the truth of the continuum hypothesis (and all the other mathematical undecidables) are not determined by the base, this strategy cannot be applied to mimic realism about sets without commitment. When the base determines the rest, however, we can always make do with just the base, assuming this strategy is legitimate. And maybe that is right. Strategies in this vein are offered by Rayo [2007, 2008], Williams [2012], Linnebo [2012], Cameron [2010], and Melia [1995, 2008]. On the other hand, this kind of radically minimal paraphrasing strategy has to say something about ordinary talk about tables and chairs, because ordinary talk is not going away, and (what is at least as important) ordinary language is useful because it corresponds in some way to how things really are. If a paraphrasing strategy cannot distinguish in a principled way between the reality of fictional characters and that of medium sized dry goods, then fictional characters will probably remain respectable enough.

- **4.12 Categorizing Fictional Ontologies**

Kit Fine [1982: 97] draws three distinctions between different views of non-existent objects: Platonism/empiricism, literalism/contextualism, and internalism/externalism. Platonism holds that it is necessary which non-existent objects there are, independent of us or anything else, while empiricism does not. Literalism holds that the objects really have the properties ascribed to them in the literary contexts, while contextualism does not. Internalism individuates the objects according to the properties they have in the literary contexts, whereas externalism does not. That is quite simple. Perhaps some views could try to straddle one or other of the distinctions, but most do not.

The three distinctions give rise to eight positions, some of which fit together less naturally than others. We could also make finer distinctions, and will make two. Platonism/empiricism can be divided into two distinctions: whether what non-existents there are is necessary or contingent and whether they depend ontologically on human activity or not. These come apart if what fictional characters there are is contingent on something else. Fine's distinctions are also designed for classifying theories of non-existents rather than fictional characters, so to capture van Inwagen's view about fictional characters we would have to add the Meinongian/Quinean distinction, with Meinongians holding that fictional characters do not exist and Quineans holding that they do. Van Inwagen is on the Quinean side. More distinctions will be possible, but we will stop here. These five distinctions give rise to thirty-two views, of which some will be quite peculiar. Rather than consider all thirty-two, we will look at three categories: dependent abstracta, Platonic abstracta, and Meinongian concreta. This will let us orient ourselves within the space of positions we have mapped out, and situate the theories people actually put forward within it.

- *4.121 Dependent Abstracta*

Thomasson [1999] puts forward an account of what fictional characters are which is close to the spirit of van Inwagen's position. The basic idea is that fictional characters are *abstract artefacts*. Sculptures and screwdrivers are concrete artefacts, and fictional characters are like that but abstract. Authors initiate literary practices, and these practices give rise to various abstract objects, such as novels, poems and characters. They are the same type of thing as other non-concrete artistic creations such as musical works. More generally, they are the same type of things as other things which are not concrete but exist in virtue of the activities of humans, such as political institutions and games. John Searle [1995] has explained in more detail how we might understand these social entities arising. If we don't like that view, then with some ingenuity of the kind discussed in Lewis and Lewis [1970] we can often find concrete things to identify with these things. Perhaps the House of Commons is a building or a group of people; perhaps nations are pieces of territory and their contents; perhaps games are events (or at least mereological fusions of events). If we do accept an ontology of dependent abstracta to serve as social entities, however, it is not much ontological profligacy to accept that fictional characters are among them too.

Kripke [2011b] takes a similar view, and while unpublished his work has influenced the development of the position[57]. He explicitly says that they exist contingently, drawing the comparison with nations:

[57] In particular his 1973 John Locke Lectures, now published as Kripke [2013].

> It is important to see that fictional characters so called are not shadowy possible people. The question of their existence is a question about the actual world. It depends on whether certain works have actually been written, certain stories in fiction have actually been told. The fictional character can be regarded as an abstract entity which exists in virtue of the activities of human beings, in the same way that nations are abstract entities which exist in virtue of the activities of human beings and their interrelations. A nation exists if certain conditions are true about human beings and their relations; it may not be reducible to them because we cannot spell them out exactly (or, perhaps, without circularity). Similarly, a fictional character exists if human beings have done certain things, namely, created certain works of fiction and the characters in them. [2011b: 63]

Thomasson and Kripke hold that fictional characters exist contingently, depend ontologically on human activity and do not literally have the properties ascribed to them in the stories (except sometimes coincidentally). It is also in the spirit of the view of fictional characters as dependent abstracta to hold that if two independent literary practices ascribed a character the same properties, there would be two different characters. Thomasson takes this view explicitly [1999: ch. 5], which places her on the externalist side, although perhaps truth in fiction works in such a way that this never happens. (Obvious candidates for duplicated fictions involve symmetrical universes and simple stories, although symmetrical universes might well ascribe different object-involving properties even if they ascribed the same qualitative properties. Our Holmes lives in our London, while the other half's Holmes lives in London's duplicate.) But Thomasson's externalism also involves holding that if an author had told their story differently then the same characters would have been ascribed different properties. That comes up in real situations much more.

- *4.122 Platonic Abstracta*

Zalta [1983] puts forwards a systematic ontology of abstract objects, and identifies fictional objects with some of these. The view is supposed to supply all the abstract objects we need: mathematical objects, properties, Fregean senses, possible worlds, and even Platonic forms [1983: 41-7] and Leibnizian monads [1983: 84-90] if we want them. One advantage of finding a place for fictional characters in this ontology is that it is ontologically parsimonious: Zalta can commit to fictional characters without committing to anything he was not committed to already. Another feature which may be an advantage is that our knowledge of a systematic ontology of abstracta will presumably be *a priori*. Since abstracta appear not to be causally efficacious it can be difficult to see how we could have *a posteriori* knowledge of them. There are problems with *a priori* knowledge of abstracta too, but they are different problems and perhaps they have solutions, at least if the abstract ontology is systematic and independent of what goes on in the concrete part of reality. This is the view taken by Linsky and Zalta [1995: especially §V].

There will be more than one way of developing an account of Platonic abstracta which could include fictional characters among them. The features distinguishing such accounts will generally be that the objects are discovered rather than created, they are somehow plenitudinous, and they do not literally have the properties ascribed to them in the stories. Zalta's account (with one caveat[58]) has all these features, and we will take it as a representative way of implementing the general project and look at some of its details.

[58] The caveat is that Zalta holds that there are two kinds of instantiation and the copula is ambiguous between them, so while 'Holmes is a man' doesn't say the same about Holmes that 'Obama is a man' says about Obama, it does still say it literally. It is however still contextualism in the sense that the objects do not literally exemplify the properties they have in the stories.

Zalta [1983: 11ff] distinguishes two kinds of instantiation. The familiar and uncontroversial kind is *exemplification*, so I exemplify being human, the Eiffel Tower exemplifies being tall, and Mercury exemplifies being a planet. However, following an idea he credits to Meinong's student Ernst Mally, he distinguishes another kind of instantiation, which only abstract objects do. He calls this *encoding*. For many conditions on properties, there will be exactly one abstract object encoding just those properties. This will mean that abstract objects can be inconsistent or incomplete with respect to the properties they encode. Non-contradiction and excluded middle apply to exemplification but not to encoding.

He symbolizes exemplification in the normal way, like *Fa* for '*a* instantiates *F*', and encoding as *aF*. Properties can also be expressed by lambda terms, so the property exemplified by everything satisfying an open formula φx is expressed by the term [λx.φx]. He introduces a plenitude schema for abstract objects, which is supposed to generate all of them[59]:

$$\exists x(A!x \,\&\, \forall F(xF \leftrightarrow \varphi)), \text{ where x is not free in } \varphi. \text{ [1983: 34]}$$

'A!' is a predicate meaning 'is abstract'. Except for the instance generating the object encoding all the properties and the object encoding none of them, F will be free in φ, so φ picks out a condition on properties, just as an open sentence with one free objectual variable picks out a condition on objects. For example, we know there is a property encoding all and only Obama's properties, because of this instance of the schema:

[59] From Linsky and Zalta [1995: 552]: 'One reason that Platonized Naturalism is simple is that a single, formally precise principle asserts the existence of all the abstract objects there could possibly be.'

$$\exists x(A!x \,\&\, \forall F(xF \leftrightarrow F(\text{Obama})))$$

He also defines a third order relation of property identity holding between any pair of properties encoded by all the same objects [1983: 13], so we can have instances of the schema which enumerate the properties encoded by an object:

$$\exists x(A!x \,\&\, \forall F(xF \leftrightarrow [F{=}\text{Round} \vee F{=}\text{Square}]))$$

This generates Meinong's [1960: 82-3] notorious round square[60]. Unsurprisingly given the history of naïve set theory, Zalta needs to put some constraints on φ in order to avoid paradoxes. We discuss how successful this is in §4.234.

He takes it that encoding is a kind of instantiation, and that the natural language copula is ambiguous, so '*a* is *F*' is ambiguous between an encoding and exemplifying reading. He holds that fictional characters are the objects encoding all and only the properties ascribed to them in the stories to which they are native, so 'Sherlock Holmes is a detective' has a true (encoding) reading and a false (exemplifying) reading. Assuming that the stories leave open whether or not Holmes likes broccoli, Holmes will neither encode liking broccoli nor encode not liking broccoli. 'Holmes does not like broccoli' will thus be triply ambiguous between (using obvious symbolizations) $\neg LBh$, which is (of course) true, $\neg hLB$, which is true, and $h[\lambda x.\neg LBx]$, which is false. Holmes

[60] This gives us a round square, but perhaps Meinong had a different one in mind. This round square encodes only roundness and squareness, not their consequences. Perhaps Meinong thought his round square had straight sides as well as being square. The round square presumably does not have the properties it encodes closed under classical or necessary implication though, because then it would encode every property and be indiscernible with the round triangle. We will not explore the exegetical point.

and Watson will both be abstract objects and encode properties involving each other. Holmes lives with Watson and Watson lives with Holmes, so we have h[λx.Lxw] and w[λx.Lxh].

An important feature of the account is that abstract objects can and often will stand in various relations to concrete objects like us: we think about them, draw pictures of them, admire them and so on. Zalta [2000] shows how his theory can accommodate a lot of the things we want to say about fictional characters. It is impressive if it works, although we will examine a possible problem with it in §4.234.

It is worth situating Zalta's account within the five distinctions outlined earlier, and seeing how much variation across these distinctions would affect the account's spirit. It is Platonist rather than empiricist, in Fine's sense: it is not meant to be contingent what abstract objects there are. It is contingent which of them count as fictional characters because it is contingent which of them get written about, but whether we write about them or not they will still exist and be intrinsically unchanged. It is internalist, in that the objects are individuated by the properties they encode. It is contextualist in that e.g. Holmes is not a human in the same way that I am a human. Zalta does however take 'Holmes is human' to be literally true, because of his view about the ambiguity in the copula. The theory is Meinongian, but it does not make a special case for fictional characters: it holds that no abstract objects exist. The view could however be modified without changing much, instead saying that the concreta, contingent non-concreta and abstracta all necessarily exist. (Linsky and Zalta [1994: §4] offer just such a modification.)

The issue of whether the objects depend ontologically on us or anything else is slightly vexed. Some objects encode object-involving properties. Holmes encodes living with Watson, which involves Watson, but he also encodes living in London. One might think that these properties depended on Watson and London respectively, and that the objects depended on the properties they encoded, making Holmes and Watson co-dependent and both dependent on London. Within Zalta's framework there is no danger of this making them exist contingently, since he does not think it is contingent whether there is such a thing as London. He thinks that mere possibilia are contingently non-concrete, along the same lines as Williamson [1998, 2002]. Zalta holds *contra* Williamson that non-concrete things are non-existent, but as we said this can be modified without disrupting much else.

Even if we agree about it not being contingent what there is, we might still hold that objects encoding object-involving properties depended on those objects, even if there could be no non-trivial modal dependence. This might create some worries about reciprocal dependence, such as that between Holmes and Watson. Perhaps we should not worry about this in the case of abstract objects, but something should at least be said. If we take ontological dependence seriously, as is becoming fashionable[61], then saying it is not contingent what there is will not get us off the hook.

If we hold that it is contingent what there is, as most people do, then the problems could go beyond those of reciprocal dependence. If Holmes essentially depends on London[62], and it is contingent that there is such a thing as London, then it is contingent that there is such a thing as

[61] For orientation on this development, see Bennett [forthcoming] and Correia [2008].

[62] By 'essentially depends' I mean that he depends on London and could not exist without depending on London. There are plausible candidates for things which inessentially depend on others, for example an object might depend on its parts even though it could have had different parts and depended on those instead. The case of Holmes and London looks like essential dependence though: Holmes essentially encodes living in London and living in London essentially involves London. If encoding and object-involvement entail dependence, then Holmes essentially depends on London.

Holmes. This would put the view on the empiricist side of Fine's distinction. It should however be noted that the resultant view still has fictional characters depend on the concrete world in a very different way from how they do in Thomasson's framework. Contingentist Platonism has the domain of individuals generate a plenitude of abstracta whatever those individuals are like, whereas Thomasson's view has the abstracta depend on what the individuals do. This kind of issue suggests that Fine's Platonist/empiricist distinction does not capture all the distinctions we want to make. We should keep the issues of necessity and dependence separate, and note that even dependent abstracta can vary in how Platonist or empiricist they are in spirit.

- *4.123 Non-existent Concreta*

At least in its non-Meinongian form, there is nothing especially pre-theoretically jaw-dropping about Zalta's theory. We seem to be talking about some things so we introduce some abstract objects to be the things we were talking about, and then explain how to cash out what we say in terms of them. Taken in the right way, the views of Thomasson, Kripke and van Inwagen should also not irritate the ontological scruples of anyone already on board with social entities of the kind Searle argues for. Maybe our sense of reality is robust enough to get offended by the postulation of anything that cannot be kicked, but the view of fictional characters as abstract objects is not ontologically extravagant in a way that other uses of abstract objects are not. Whether they give an adequate account of discourse within and about fiction is a different issue, but the accusation of craziness is unlikely to stick.

There is another approach to fictional characters, however, which can look a little crazy. Within Fine's distinctions, this is the literalist approach. Literalists hold that creatures of fiction are creatures of flesh and blood. Hamlet really is a thinking, conscious, indecisive prince. Faust really made a pact with Mephistopheles. Harry Potter is a real wizard and Hogwarts a real school. You have to say something in mitigation to get a view like this taken seriously. The main options are distinguishing fictional characters from us by saying they do not exist or are not actual. Lewis [1978] does not quite identify fictional characters with concrete possibilia, but a development of it which comes closer to making this identification is propounded, though not really endorsed, by Frederick Kroon [1994]. Here we will examine Parsons' [1980] Meinongian view that fictional characters are non-existent concreta, because that is the most straightforward literalist position. In particular, Kroon's suggestion takes there to be a plurality of Holmeses, and suggests a supervaluational treatment of sentences like 'Holmes is cleverer than any (actually) existent detective'. Parsons takes Holmes to be one non-existent man, who may well be cleverer than any existent detective.

Parsons has a principle of plenitude for objects which, like Zalta's, constructs them out of properties. There is no encoding/exemplification distinction: the objects just straightforwardly have the properties they are constructed from. However, according to his theory there is no golden mountain that exists, no possible round square, no television thought about by Socrates and nothing complete but not organic or inorganic. (*Complete* is a technical term Parsons uses for objects which, for every property, have either that property or its negation.) This means he needs to impose restrictions on how to construct objects out of properties.

Parsons distinguishes between *nuclear* and *extranuclear* properties, and holds that for every set of nuclear properties there is an object which has just those properties. Existence, possibility, being thought about by Socrates and completeness are all extranuclear properties. They are respectively ontological, modal, intentional and technical, which are the four categories of extranuclear properties Parsons identifies [1980: 23].

To get round Russell's famous problem of the existent golden mountain and the like, Parsons [1980: 42-4] says that extranuclear properties also have watered-down nuclear versions. The existent golden mountain has existence, but does not exist. This was Meinong's solution. Russell was unimpressed, so was Quine [1948], and if I may report an intuition, so am I. Going beyond intuitions, we can put the problem like this. In the stories, Holmes is a human and he exists. Parsons says that Holmes is a human just like me, but he does not exist just like me. I exist in the watered-down way and in the neat way, but Holmes only exists in the watered-down way. But this raises the question of what being a human is supposed to amount to. Why not say that I am a human in the neat way and Holmes is only human in the watered-down way? Well, it turns out that to make the theory work you need two kinds of existence but you can get away with only one kind of humanity. This makes the distinction *ad hoc* and obscure though. Is humanity like the watered-down properties or like the neat properties? It seems to me to be like the watered-down versions, because in itself it does not have any existent-world consequences. But that's unsatisfying, because my humanity seems no less watered-down than my existence. This line of objection is impressionistic, but the distinction it objects to is obscure. More needs to be said, and it is not clear to me that anything much illuminating can be said.

Fine [1984] also raises some technical and some philosophical objections to Parsons' view. In response to the technical objections he makes some suggestions as to how the view could be fixed in the same spirit, while the philosophical objections are to the spirit. Parsons' view is necessitarian, Platonist, internalist, literalist and Meinongian, whereas Fine favours a view which is contingentist, empiricist, externalist, contextualist and Meinongian. I won't go through all the technical difficulties and suggested fixes, but suffice to say that Fine thought Parsons' theory needed some work.

It may be that a literalist theory can be put forward which does not need watered-down properties and does not experience the technical troubles Fine finds in Parsons' theory. Yagisawa's [2010] theory may be a candidate, although he may not see it as Meinongian: he holds that all the possibilia and impossibilia are (absolutely) real and concrete, but that their existence is relative to worlds and times [2010: 49-61]. If Yagisawa's view is not strictly Meinongian then it is presumably structurally similar to one which is, and the distinction may be merely verbal. In any case, my strategy will try to sidestep the issue of whether a version of concrete Meinongianism can ultimately be made to work. Instead we can try to undermine its motivation by presenting a combination of pretence for discourse within fiction and abstracta for discourse about fiction, which includes a straightforward but non-Meinongian treatment of negative existentials. That will be the goal of §§4.3-4.4.

4.2 Problems for Fictional Ontologies

- ### 4.21 Creations or Discoveries?

If you think there are fictional characters, you need to say something about their relationship with their authors. There are three natural positions here:

1) Authors create fictional characters: before Doyle wrote his stories there was no Holmes, but now there is, as a result of Doyle's activity.
2) There are some things already out there which authors turn into fictional characters by their activities. Holmes was already there, but it is as a result of Doyle's activity that he is a fictional character.
3) Authors discover fictional characters: the fictional characters are all already out there, and Doyle discovered Holmes by writing his stories.

Thomasson holds the first position, that fictional characters depend ontologically on the literary activities of concrete beings like us. For Thomasson, our activities bring them into existence, and under some conditions they can go out of existence too. For example, if the Earth had been destroyed before we started broadcasting radio and TV shows into space, all traces of our literary practices would have been destroyed, and all the fictional characters depending on those practices would have been destroyed with us. Some fictional characters were probably destroyed when the library of Alexandria burned, too, or if not immediately then soon after when their readers had all forgotten them or died. The creationist position has some support from the things we ordinarily say: we often talk about authors creating their characters, and we seldom talk about them being discovered. Thomasson's position allows us to take this as literally true. There is room for holding that fictional characters can be created but not destroyed or (bizarrely) *vice versa*, but I will ignore this distinction here.

Zalta holds the second position: there is a plenitude of abstract objects, and when an author writes a story according to which there is something which is a certain way, the abstract object encoding the relevant properties becomes a fictional character. This makes fictional characters stand to their authors in the kind of relation Marcel Duchamp's famous fountain stands to Duchamp. Before Duchamp's artistic activity the object was just a urinal, and his activity turned it into an artwork. He did not cause there to be any objects that were not there before, but he did cause there to be an artwork where before there was no artwork. Parsons [1980: 188] mentions this suggestion and credits it to David Kaplan, without indicating that either of them endorses it.

The third position, that fictional characters are already fictional characters before their authors discover them, is not so popular. There is not much which could make us pick it ahead of the second, and the second probably accords better with ordinary talk. This is a symptom of the shallowness of the difference between them, though, at least from a metaphysical point of view. Perhaps considerations could be brought to bear from the details of our literary activities. Interesting as that sort of thing might be, I will leave it to other people. The metaphysical question is about what there is and what it is like. The second and third positions agree about these, except for the extension of 'fictional character' (with respect to various times), which makes the disagreement between them look merely verbal. For present purposes, we will look at the disagreement between positions like Thomasson's, which we can call *creationist*, and positions like Zalta's, which we can call *selectionist*, because authors select objects to turn into fictional characters.

As I mentioned earlier, creationism seems to be most in line with ordinary talk. We say Doyle created Holmes; we don't say he found him and decided to write about him. Perhaps we're wrong to say this, or our talk of creation is not meant literally, but insofar as ordinary talk is on either side, it is on the creationist side. Or so it seems. Presumably what goes for Holmes goes for all fictional characters, but ordinary talk can be more capricious than our metaphysics should be. It seems to me that the best candidates for fictional characters which we might say *prima facie* were discovered are the characters in jokes. Jokes abound with characters: rabbis, bishops, Englishmen, Irishmen, and the amusingly named callers in knock knock jokes. Jokes are often simple enough that people might well think of them independently, and the author's input can often seem more like the discovery of a pre-existing near-homophone than the *ex nihilo* creation of a fictional universe and its inhabitants. Insofar as there is a pull towards saying jokes are discovered and not created, there is some pull towards saying the same of their characters. Perhaps the pull is not strong, but if the point is conceded for jokes, we can push it to novels, since the case is hard to make that they are different in kind. Good novelists discover that if words are arranged in a particular order then a good novel results. The possibility of so arranging the words was already there, and perhaps the characters were there too.

Examination of ordinary talk can shade into discussion of aesthetic reasons for saying that characters are created rather than discovered. One might think discovery did not do justice to artistic creativity. It is perhaps more common to argue in this way about musical works. Here is Jerrold Levinson:

> The first objection to the view that musical works are sound structures is this. If musical works were sound structures, then musical works could not, properly speaking, be created by their

> composers. For sound structures are types of a pure sort which exist at all times…
>
> But why should we insist that composers truly create their compositions? Why is this a reasonable requirement? This question needs to be answered. A defense of the desideratum of true creation follows.
>
> The main reason for holding to it is that it is one of the most firmly entrenched of our beliefs concerning art… [Levinson 1980:7-8, discussing a point he credits to Wolterstorff 1975: 138]

The same sort of thing could be said about fictional characters. Maybe if Shakespeare merely worked out that you could write a play about someone just like Hamlet that detracts from his achievement. I do not see how though, really: writing the play would be just as difficult either way. The hard part is writing the right words in the right order, whether metaphysics works such that this constitutes creating Hamlet or discovering him. To paraphrase Davidson [1971: 23], Shakespeare never did more than move his body, and the rest was up to nature. Rearranging the furniture of the world can be as artistic as adding to it, as with flower arranging. (Doubtless some people will say that calling flower arranging mere rearrangement does not do justice to its creativity either, but considering concrete artworks at least puts the issue about fictional characters in some perspective.)

It is however conceivable that our appreciation of literature would be improved if we believed that authors created their characters rather than discovering them. Should this affect what we believe? There are a few reasons for thinking it should.

First, truth gives a *pro tanto* reason to believe a proposition, but it is probably sometimes rationally permissible to believe in opposition to the evidence and arguments, for people who can manage it. (Standard examples involve powerful beings who will punish or reward you based on your beliefs, as in one version of Pascal's wager.) Maybe aesthetic concerns could provide such non-epistemic reasons to hold a particular position about the metaphysics of fictional characters.

Second, if the metaphysics of fiction is itself fiction, then literature might be better off for metaphysics contributing one picture rather than another, and it is the job of writers of fiction to improve literature. This possibility is a live one: fictionalism is a common view all over metaphysics. Perhaps fictionalism is not an option in the metaphysics of fiction on pain of some kind of regress, but perhaps it is. I will not be addressing that question here because it is more a meta-metaphysical issue than a metaphysical one and discussion of it would take us too far afield.

Third, it is not uncommon to classify works of art in one genre or another according to what genre they would be a good example of, and this may even contribute to determining which genre a work actually falls under. (For example, Ben Caplan [2011] argues partly on this basis that the movie *Fight Club* is a romantic comedy.) If there is room in our metaphysics for both created and discovered characters, we might be correct to classify characters one way or another depending on how it contributes to our appreciation of the works they appear in.

The first reason can be ignored, because I am only interested in telling people what to believe insofar as that matches up with telling them where the evidence and arguments point. The second reason will be ignored here too, because it is weird and only applies if metaphysics is fiction anyway. The third seems only to apply if we adopt a theory according to which there are two genres of fictional work, one discovering characters and one creating them. I will not defend such a theory, and know of nobody who does. Given these considerations, I will ignore the aesthetic reasons for picking one theory over another, and look at considerations of a more technical nature.

A style of argument not resting on aesthetic considerations would be one like this: suppose that, whether fictional characters are created or discovered, it is always possible to create or discover arbitrarily many of them, even into the transfinite. To argue for this, suppose for *reductio* that K is the limit. We tell the following story: 'As I was going to St Ives, I met a man with 2^K wives'[63]. By Cantor's theorem, this creates/discovers more than K fictional wives. So there is no limit: however many there are, we could create/discover more. If they are discovered, this is a contradiction, since you cannot discover more things than there are. This means they are created.

A robust line to take in the face of the problem is to say that actually there is an upper bound on how many fictional characters we could populate our stories with, because of limits on the stories it is metaphysically possible to tell. Certainly we could write stories containing very large multitudes, but there could be a limit somewhere. The limit is however presumably not set-sized, since people do tell stories according to which there are more things than would fit in a set. It is perhaps orthodox to believe there are that many abstracta, and Daniel Nolan [2004] describes worlds in which there are that many concreta.

If we allow that there might be too many fictional characters to form a set, the argument seems to break down, since 2^K is less well defined where K is that large. Even if the argument could be made to work round this issue though, the creationist should probably not rely on it

[63] In case you are worried that spacetime does not have enough room for so many wives, consider them living in parallel universes overseen by God, who married him to them all at (for the groom) the same time. For further discussion of transworld romance, see Sinhababu [2008].

anyway, for fear of seeming to settle questions about the size of the set-theoretic universe by telling silly stories like 'As I was going to St Ives, I met a man with an inaccessible cardinality of wives.' It would probably be sensible to say that, at least in some cases, multitudes do not have as many members in reality as they have according to the story. This is what Terence Parsons (1980: §7.5) does. According to the story the man may have 2^K wives, but this does not mean there are 2^K fictional women who are married to him according to the story. Some intuitively true sentences will come out false on this view:

> There are at least 2^K characters in 21st century stories described with less physical detail than is any heroine of a nineteenth century novel.

It is thus a cost to a theory to solve the problem of the multitude this way, but this would not be the first bullet people have bitten over contradictions in some formulations of set theory. Not much ordinary literary critical talk would have to go, and we have an explanation for the falsity in the part which would. The explanation would be similar to that given for me not having the property of non-self-instantiation, even though I do not instantiate myself (there is more on this paradox in §4.234).

It is possible that the creation/discovery debate can be defused by making space for systems of things structurally like what each side thinks fictional characters are. A token fictional character will fall under a type, and the types are discovered, even if the characters are not. A particular literary practice relating to the type is created, even if the character is not. We can sort out the metaphysics of both the types and tokens, piggybacking on the systems already put in place by the different sides for accommodating fictional characters. Then we can leave the debate about where exactly fictional characters fit into the picture for another day, secure in the knowledge that whether they are created or discovered they will be able to fit in somewhere. If created they will track the authorship events, and if discovered they will track the character types.

A related point is made by Zalta [2000: §6], in which he suggests that pretence theorists paraphrase their talk in terms of possible patterns of pretence behaviour, so they have a more systematic semantics using only referents to which they are already committed. Whether or not Zalta's specific proposal works, the discovery/creation distinction may boil down to a type/token distinction, in which case we had better be able to cope with both.

It does make a bit of difference whether the referents of 'Poirot' and so on are the types or the tokens, for example the types will in general be more famous than the tokens. It may well be, however, that usage does not settle the matter one way or the other. We could try to settle it by examining intuitions about the modal properties of fictional characters, but I suspect this will not work. If we talk about whether the characters would have existed if the stories had not been written, this could be explained by a selectionist as saying the thing which is the character would not have been a character. It might still have existed. If we say that the same character could have been given different properties by its author, this can be explained counterpart-theoretically, or by treating the relevant referring expressions as non-rigid. We have the same problem with 'If I were you' or 'If Gandhi had been a woman[64]',

[64] It is not actually clear that necessity of origins entails necessity of actual original sex. We can imagine a sperm having its sex chromosome removed and replaced with the father's other one, producing a person of the opposite sex from the same sperm and egg, which the same genetic parents. There is at least very little indeterminacy in the genomes of someone's closest counterparts of the opposite sex (as far as high-school biology goes). 'If I were you' provides a more solid example of an everyday counterfactual whose

and we do not draw conclusions about whether ‘I’ and ‘Gandhi’ refer to types or tokens. We could have a counterpart relation relating characters in different worlds according to the authorship events they correspond to, rather than according to the properties they have in the stories. This option would to an extent dissolve Fine’s question of whether fictional characters should be individuated internally or externally. The positions would still disagree over individuation of fictional characters within a world, but for transworld individuation we could have it both ways.

4.22 Kripke and Unicorns

Kripke [1963] says that there could have been things that actually there are not. This is effectively equivalent to denying that instances of the Barcan formula are always true. The Barcan formula says that for any condition φ, if there could have been a φ, then there is something which could have been a φ. Symbolically, it is this schema:

BF $\Diamond\exists x\varphi x \rightarrow \exists x\Diamond\varphi x$[65]

Some people accept the Barcan formula, such as Williamson [1998, 2002]; Linsky and Zalta [1994, 1996]; and Bolzano as interpreted by Schnieder [2007], who hold that there are contingent non-concreta, and if my parents had not met then I would have been one. Nonetheless, the consensus is still probably with Kripke. He gave Sherlock Holmes as an example of something which does not exist but could have done: ‘Holmes does not exist, but in other states of affairs, he would have existed’ [1963: 65].

antecedent is impossible if taken at face value. (Or if you follow Caspar Hare’s [2009] egocentric presentist semantics for ‘I’, then ‘if Gandhi was you’.)

[65] It is named after Ruth Barcan Marcus, who used its necessitation as an axiom schema in Barcan [1946: 2].

It seems undeniable that there could have been people whose parents were Quine and Margaret Thatcher, but there is plausibly nothing that could have been a person with them as parents. There will of course be combinations of views on personal identity and material constitution entailing that some gerrymandered fusions of particles could have been such people, but there seems nothing *prima facie* wrong with denying the Barcan formula, and it seems an odd thing to settle questions about persons and constitution on the basis of it. In any case, there could be other examples; for example, the universe could presumably contain more elementary particles than it actually does, in which case some of them would have to be non-identical to any actual ones, and it is hard to think of anything else which could plausibly have been an elementary particle. David Armstrong gives a further example:

> [I]t seems very hard to deny that it is possible that the world should contain more individuals than it actually contains. There is no mouse in my study. Nevertheless, it is possible that there should be one. But why does the mouse have to be one of the world's mice? Why not an additional mouse? And, if additional, why not made up of particles (assume a materialist theory of mice) which are additional to the world's particles? [Armstrong 1989: 57-8]

Later, Kripke [1980: 156-8] changed his mind. He still maintained that there could have been things that actually there are not, but he no longer held that Sherlock Holmes is one of them[66]. He also holds that there could have been things belonging to natural kinds that nothing

[66] Kripke [2011b] holds that there is such a fictional character as Sherlock Holmes, but this is a different (though of course related) usage of the name, referring to an actual artistic creation, not a non-actual but possible person. The point at issue is whether the person could have existed, not the abstract creation. Trying to avoid Kripke's conclusion by pointing to the possibility of the abstracta is a blind alley.

actually belongs to, but there could not have been such things as unicorns. In discussing his argument people have mostly focused on the brief remarks in Kripke [1980], although he also discusses it in Kripke [2011b, 2013] and in Dummett et al [1974b][67].

It seems at first a puzzling claim that Holmes and unicorns could not have existed. Kripke [1980: 23] said that his argument 'doesn't ever convince anyone', although some people have been convinced since he wrote that. Reimer [1997] and Yagisawa [2010] are examples, but there are many others and the view is now fairly mainstream. It seems puzzling though, because it looks like there is no impossibility in the Holmes stories or in the stories about unicorns. Even if there is, the stories could be tidied up or simplified to remove the inconsistencies. Some stories, like that in Priest [1997], are meant to be inconsistent, and some stories presumably contain accidental inconsistencies which could not be removed without doing violence to the point of the story, but stories like those about Holmes and unicorns are not meant to be like that. They recount events which did not happen but could have done, or so it seems. These stories certainly appear to say that unicorns or Holmes existed, so it is puzzling to say that they could not have done. We can set it up as an inconsistent triad:

- Things could have happened as the stories say.
- The stories say that Holmes/unicorns exist.
- Holmes/unicorns could not have existed.

[67] Dummett et al [1974b] is a transcript of the discussion of a presentation of the paper eventually published as Kripke [2011b]. It does not get cited much and seems not to be well known, so as a point of both historical and intellectual interest it is worth drawing attention to it. The participants in the discussion were Davidson, Dummett, Gilbert Harman, Kaplan, Kripke, David Lewis, Charles Parsons, Barbara Partee, Putnam, Quine and Sellars. Dummett et al [1974a] has the same people discussing Quine's indeterminacy of translation.

One can just pit Kripke's arguments and the intuitions to the contrary against each other, take a side and leave it at that. I will not do this. The problem is to my mind a deep one which would tell us a lot about the way fictional reference works if we could get to the bottom of it. In this section I will present Kripke's argument and some choices people make in response to it. In §4.4 I will give my own explanation of what I think is going on.

- *4.221: Kripke's argument*

Considering that its conclusion is so strange, Kripke's argument is quite simple. In view of the sketchiness of his remarks, I will present a version which is along the same lines but may not be quite the same as Kripke's intention in the details. It begins from the observation that the stories leave a lot open. Insofar as the events of the stories could have happened at all, they could have happened in many different ways. In these different ways that things could have gone, someone would have played the Holmes role, or some species would have played the unicorn role. Even if we think no actual person or species could have played these roles without changing the story, lots of different non-actual people or species could have done. If you dislike this formulation because it seems to quantify over possibilia, these are more innocent:

- For some incompatible properties F and G, it is consistent with the stories that someone play the Holmes role who was necessarily F-if-they-existed, and consistent with the stories that someone play the Holmes role who was necessarily G-if-they-existed.
- For some incompatible properties F and G, it is consistent with the stories that members of the species that plays the unicorn role be necessarily F-if-they-exist, and consistent with the

> stories that members of the species that plays the unicorn role be necessarily *G*-if-they-exist.

If you are a nominalist and also dislike the reference to properties, I leave to you the task of paraphrasing these formulations in your preferred style. Note that quantification over properties is not needed, because Kripke's argument would still have what force it has if we used formulations with specific properties. These formulations can thus be seen as schemas, rather than ineliminably quantificational statements. The properties in question might be that Holmes be descended from Genghis Khan or not, and that unicorns be in the order Artiodactyla, like deer, or the order Perissodactyla, like horses. (The example for unicorns is from Dummett [1993b: 346].)

Now, it is part of the stories that Holmes is a particular person, and that someone who behaved the same way would not thereby be him. We can also allow for the sake of argument that it is part of the stories that unicorns are a particular species. Just as a species superficially like tigers but with a different makeup and evolutionary history would not be tigers, fool's unicorns would not be unicorns. We have noted that different people and species could play the Holmes and unicorn roles, but which is Holmes and which the impostor? Which are the unicorns and which are the fool's unicorns? Kripke draws two conclusions. First, the epistemic conclusion that we cannot know of one particular possible person or species that it would have been Holmes or the unicorn. Second, that nothing would determine that one particular possible person or species was Holmes or the unicorn, and so nothing would be. The intuitions can be pressed further by considering worlds in which more than one thing plays the role in question, and by asking whether, had one version been actualized, the other version would still have been possible. There are ways of resisting the argument which we will examine shortly, but the straightforward conclusion which Kripke draws is that Holmes could not have existed and there could not have been unicorns.

David Kaplan [1973], another architect of the theory of direct reference and rigid designation, also saw the view as having similar consequences. Here is his statement of the argument:

> I have argued that 'Aristotle' denotes something which, at the present time, does not exist. I could now argue that 'Pegasus' denotes something which, in the actual world, does not exist. I shall not. Pegasus does not exist, and 'Pegasus' does not denote. Not here; not anywhere. What makes 'Aristotle' more perfect than 'Pegasus'?
>
> The 'Aristotle' we most commonly use originated in a dubbing of someone, our 'Pegasus' did not. Some rascal just *made up* the name 'Pegasus', and then he pretended, in what he told us, that the name really referred to something. But it did not. Maybe he even told us a story about how this so-called Pegasus was dubbed 'Pegasus'. But it was not true.
>
> Maybe he proceeded as follows. First, he made up his story in Ramsified form: as a single, existentially quantified sentence with the made up proper names ('Pegasus', 'Bellerophon', 'Chimaera', etc.) replaced by variables bound to the prefixed existential quantifiers; second, he realized that the result was possible, and that therefore it held in some possible world, and that therefore there was at least one possible individual who played the winged horse in at least one possible world; and third, he tried to dub one of those possible individuals 'Pegasus'. But he would not succeed. How would he pick out just one of the millions of such possible individuals? [Kaplan 1973: 505-6; emphasis in original.]

Kaplan is only making the argument in the case of fictional names, but you could run a similar argument for fictional kind terms. The Ramsified sentence would need to use predicate variables, but these could be within a many-sorted logic instead of a full-on second-order logic, if second-order (modal) logic was considered more problematic. The comparison with 'Aristotle', whose denotation Kaplan thinks does not exist anymore, brings out that Kaplan does not deny 'Pegasus' a denotation on the grounds of actualism, but because of underspecification. Indeed, Kaplan allows that we can name things that do not exist and never did or will provided there is not the same kind of underspecification; his example is the car which would have been made at a particular automated assembly plant if production had been halted a few seconds later [1973: 517][68]. His argument concerning 'Pegasus' is essentially the same as Kripke's and we will not treat them differently, but giving his alternative statement of it may serve to illuminate it, as well as awarding Kaplan the credit for coming up with it, insofar as he did.

- *4.222 Descriptivist Responses*

One can make various kinds of descriptivist response to Kripke's argument. The most flatfooted says that while some of the data in Kripke [1980] and Putnam [1975] seem to tell in favour of proper names and natural kind terms as being directly referential or at least rigid designators, the results about fictional terms show the view to be nonetheless absurd. We fall back on descriptivism, and say that all kinds of internally and genealogically different animals could be unicorns, and the same goes for horses and tigers and the rest. If Holmes had existed he could have had many different origins, and actual people could have had different origins too. I take that position to be implausible, and in any case part of the project of this thesis is

[68] More carefully stated: he thinks we can have terms which refer with respect to some times and worlds but not with respect to any time and the actual world.

essentially working out how to avoid being pushed into it by the problems of empty names.

A less flatfooted descriptivist response agrees that Kripke and Putnam's data make the case that 'tiger' and 'Socrates' are rigid designators unassailable, but maintains that descriptivism still wins out in the non-referring cases. It might seem *ad hoc* to go with descriptivism when but only when there is nothing to rigidly designate, but perhaps this is predicted by an independently motivated metasemantics. A fairly popular view propounded by David Lewis [1974, 1983, 1984] holds that words take the most eligible meanings in the vicinity of the conventions governing their use. If we subscribed to a view like that, we could hold that if there were animals fitting the descriptions of unicorns, 'unicorn' would rigidly designate their species, but since there are not, the most eligible meaning is a descriptive one. This does justice to the intuition that there could have been unicorns while leaving the semantics of other referring expressions alone.

One thing to dislike about this is that when competent speakers do not know whether a term is empty or not, they will not know what kind of meaning it has. Perhaps we can stomach that, but a more damaging objection is that the characters in the stories will still be using the terms rigidly, and the stories will consequently still hold that their assent to sentences like 'there could have been fool's unicorns' is correct. The consequent failures of disquotation could make discussion of such works confusing. Related to this, we will not be able to truly say things like 'there could have been both unicorns and fool's unicorns'. Note that we cannot straightforwardly solve these problems with the descriptivist device of making the descriptions rigid, with 'unicorn' meaning 'the actual occupier of the unicorn role'. While this would vindicate the characters' talk, it would falsify ours, because there is no actual occupier of the unicorn role, 'unicorn' would have a null extension at all worlds, and 'there could have been unicorns' will be false. Adopting descriptivism just for empty terms can be saved from the charge of being *ad hoc* by a suitable independently motivated metasemantics, but it generates some ugly results.

- *4.223: Dummett*

Dummett [1993b] agrees with the descriptivists that there could have been unicorns, but he still wants to do justice to the position that ‘unicorn’ is a kind term whose semantics works like those of non-empty kind terms and validates Kripke and Putnam’s data. He [1983, 1993b] seems to agree with Kripke about proper names but not general terms, although he takes it that in many cases a proper name *N* can be used to form a general term ‘such a person/thing as *N*’ which is treated as he treats ‘unicorn’:

> Consider a name which everyone in fact believes to have a reference, say “Charlotte Corday”; and suppose, for present purposes, that there actually was no such person, and that the story of Marat’s assassination is spurious. Then our use of the name is founded upon a mistaken belief; but still, that belief might have been correct, and then the name *would* have had a reference. It is the same with most empty definite descriptions or mistaken observations: there might have been something answering to the description; the observation might not have been erroneous. If a proper name had been introduced on the basis of such a mistake, we cannot say that it *could not have had* a bearer. Admittedly, in our hypothetical case, it would make no sense to say that *that person*, Charlotte Corday, might or could have existed; but we could properly say that there might have been *such a woman as* Charlotte Corday. [Dummett 1993b: 334; emphasis in original.]

He seems to decide particular cases on the basis of nuances in how the name was introduced and whether its fictionality is common knowledge. In view of this it is possible his view is not so far off Kripke's in many cases; nonetheless Dummett thinks there is a class of terms to be treated as he treats 'unicorn' and Kripke thinks there is no such class. We will examine the consequences of treating terms as Dummett thinks 'unicorn' should be treated.

He makes use of the observation that Kripke's argument relies on a problem with tie-breaking: if there were creatures fitting the description which were all or mostly of the same kind, then their actuality could break the tie, but since there are none, we have a tie between many kinds of creature. Dummett then says that if there were creatures playing the unicorn role, the tie would be broken and whatever kind of thing they were would be the unicorns.

However, since the tie could be broken in favour of many different and exclusive kinds, the worlds containing different kinds of unicorns cannot be possible relative to each other, though they are all possible relative to the actual world. This means that the proper logic for metaphysical necessity cannot be S5, because if two possible worlds are accessible from the actual world but not from each other then the accessibility relation is not Euclidean. The characteristic axiom for a Euclidean accessibility relation is $\Diamond\varphi\rightarrow\Box\Diamond\varphi$, saying that whatever is possible is necessarily possible. The two most standard ways of weakening S5 to get round this are denying transitivity, producing the logic B, or denying symmetry, producing S4. Dummett decides to deny symmetry.

He recognizes that it looks like if the world had contained one kind of unicorn then a world non-modally like the actual world would have still been possible, but holds that such a world would still have been constrained by the metaphysical necessities of the unicorn-containing world. However, it seems odd to say that the world could have been modally different without being non-modally different, and Dummett admits that perhaps we should therefore also reject transitivity. This would leave the logic T, which demands only that accessibility be reflexive, i.e. whatever is necessary is true. Now, although Dummett does not suggest this, if we are denying transitivity anyway we could make the relation non-Euclidean without rejecting symmetry, giving us the logic B. Then the unicorn worlds would be possible relative to the actual world and the actual world would be possible relative to them, but worlds with different kinds of unicorn would not be possible relative to each other. The actual world itself would vindicate the intuition that even if there had been unicorns the world could have been as the actual world is, and we would not need its modally discernible duplicates. The modal could once more determine the non-modal[69]. I think this is a more promising proposal if we want to take Dummett's side against Kripke. It appears from Dummett [1993a: xv-xvi] that he was not too wedded to the use of S4 here, and was chiefly interested in finding an argument for using any logic of metaphysical necessity weaker than S5. The B proposal still gets that. But should we want to take Dummett's side?

Marga Reimer [1997] does not think so. She says that we can accept that we would be right to call creatures playing the unicorn role

[69] It is hard to express exactly which condition the version of Dummett's proposal with just the logic T violates, because the problem worlds would not be possible relative to the actual world. This means that at any given world we could have necessary supervenience of the modal on the non-modal. There is still something strange about it though, because if the non-modal facts determine the modal facts at the actual world, we might wonder why the same non-modal facts do not determine the same modal facts at the problem worlds. However, since the problem worlds are impossible (relative to the actual world), we could probably tolerate them if the rest of the theory looked good.

‘unicorns’ if there were any, and still deny that there could have been unicorns. This is because if there were any such creatures ‘unicorn’ might have meant something different. She is right. The point is essentially the one frequently attributed to Abraham Lincoln[70]:

> “How many legs does a dog have if you call the tail a leg? Four. Calling a tail a leg doesn't make it a leg.”

If we look at things Reimer’s way we can see the tie which needs breaking as a metasemantic one: we have lots of possible species and no way of making the word ‘unicorn’ refer to one rather than the other. If there was a species playing the role it would break the tie and ‘unicorn’ would refer to it, but there isn’t and it doesn’t. This is an uncharitable interpretation of Dummett; Lincoln’s point has been acknowledged for a long time now and philosophers of language know to watch out for it. (Kaplan [1973: 505] uses the point as the basis for his Homework Problem #20.) On the other hand, people do make mistakes and maybe Dummett made one here. There is however a more charitable interpretation of him, although it does involve some substantial commitments about property ontology and essence.

Reimer sees the tie as metasemantic, but we could see the tie as more metaphysical. The idea would be that kinds get their essences in part from their instances. On this view, whatever charge electrons actually have, they necessarily have, but with uninstantiated kinds of particle things are more open. If there had been phlogiston it would have had its mass, charge and so on essentially, but since there is no phlogiston, there is nothing to give it this essence, and it could have been various ways. The situation is the same with unicorns. If unicorns had been a

[70] There is some doubt as to whether Lincoln actually said it. Rev E. J. Stearns [1853: 46] definitely did say much the same, but I will follow tradition and refer to the point as Lincoln’s.

species of artiodactylae or perissodactylae they would have been so essentially, but as things are they are not.

On two fairly plausible assumptions about essences, this picture should give us a symmetrical but non-transitive accessibility relation for metaphysical necessity. One assumption is that when a kind has a (non-disjunctive) property which that kind of kind can have essentially it does have it essentially. So unicorns are a kind of animal, and kinds of animal can have (let us suppose) their genetic makeup and evolutionary history essentially, so if unicorns have a particular genetic makeup and evolutionary history, they have them essentially. Worlds are possible relative to each other when the essences of the properties instantiated at those worlds do not exclude each other. Dummett's example of the actual world and the two different unicorn worlds is a counterexample to transitivity. Now we try to prove symmetry, at least as far as unicorn considerations go. Suppose that at w unicorns are essentially F, v is possible relative to w, and at v unicorns are essentially G. It follows that at v all unicorns are F and G, and since these are properties that unicorns can have essentially, they do. There are no non-F unicorns at w, by reflexivity. Could there be non-G unicorns at w? Well, for all we have said there could. To secure symmetry we need the additional principle that if a kind of kind can be essentially H (for non-disjunctive H), it can also be essentially non-H. This has some plausibility, at least for some properties like being descended from Genghis Khan, but perhaps it is false. With the assumption we get symmetry; otherwise we may not. But whichever way we go, denying transitivity allows us to keep what is important about Dummett's position while not allowing modal facts to vary without variation in non-modal facts.

Perhaps this kind of metaphysics of essences is not plausible. Perhaps it can be cashed out in a less metaphysically heavyweight way, and can thereby be made plausible. Perhaps it should not seem wildly implausible to someone already sympathetic to kinds having essences, since it would be good if essences came from somewhere, and it seems that the properties of a kind's instances might contribute to that. In any case, it looks like something along these lines is the way Dummett has to be interpreted to avoid Lincoln's point.

- *4.224: Reimer*

Dummett used the premise that if there had been a species uniquely filling the unicorn role then 'unicorn' would have referred to it. Kripke's tie-breaking argument would break down in that case, so we should not deny Dummett's premise without further argument. However, we have seen that Reimer's explanation of the premise in terms of Lincoln's point means that we can grant it and still keep Kripke's conclusion. Since the attempt to follow Dummett's conclusion through led to some substantial and possibly unwanted commitments about essences, perhaps we should accept Reimer's explanation and accept that unicorns are not possible, and similarly accept that Holmes is not a possible person. Reimer's way of doing this treats the terms as non-referring.

It is actually now fairly clear from Kripke's own discussion of the case that this is his response to the argument as well:

> Statements about unicorns, like statements about Sherlock Holmes, just *pretend* to express propositions. They do not really express, but merely purport to express, propositions. In the case of species, at least, this is true when the myth has not fully specified a hypothetical species, as I have mentioned in the previous paragraph. One cannot say when these sentences would have been true of a counterfactual situation, and therefore no proposition has been expressed. [Kripke 2011b: 67; his emphasis.]

Of course, just because Kripke thinks this is the proper response to his argument that unicorns and Holmes could not have existed does not mean it is. We can follow Kripke as far as his lemma but not as far as his conclusion. Nonetheless, Kripke's position is clearly that statements apparently about unicorns do not express propositions. Words like 'unicorn' and 'Sherlock Holmes', when used for flesh and blood things and not for abstract artistic creations, are coined for pretending (or lying) and not for sincerely asserting. If you try to use them for asserting, your utterances get treated like those of someone who uses the word 'Vulcan'. If this is all there is to their meanings, then the treatment of chapter one would assign utterances involving them pessimistic truth values, and presumably in a modal language they would take these pessimistic values at all worlds.

Reimer thinks it is important to take the terms as non-referring because otherwise we either have to be descriptivists about the referring terms or accept a disuniformity in our semantics. In §3.222 it was suggested that such a disuniformity might be defensible, but it is ugly. If we accept they are non-referring we get uniformity and treat the referring terms the way we want to. Something would have to be said about the pretended uses too though, to explain why fictional terms are not just semantically defective. In §4.4, especially §4.431, I will try to show how we can pin down the way they work in pretence enough that we can sometimes piggyback on the pretended use to make genuine assertions using fictional terms.

- *4.225: Yagisawa*

In the absence of an account of the kind I just promised, however, it could seem unfair to treat 'unicorn' and 'Holmes' the way chapter one treats 'Vulcan'. The terms were not introduced by mistake and they seem to be meaningful, so they ought to refer to something. But if they do not refer to possible things, what might they refer to? The obvious answer is that they refer to impossible things. Unicorns are animals which could not have existed, and Holmes is a person who could not have existed. A more careful and actualistically acceptable formulation could say that the terms refer with respect to some impossible worlds but not with respect to any possible world, and behave as referring expressions only in the scope of counterpossible conditionals. It should be as unproblematic to fit such terms into an actualist semantics for a language containing modal operators and counterpossible conditionals as it is to give an actualist semantics for any modal language not validating the Barcan formula. A plausible metasemantics explaining how the terms could come to have the meanings this semantics assigns to them would be another story, but if the metasemantics is defensible then at least the semantics is coherent.

A particularly committed defender of this view is Yagisawa [2010: §10.2-10.4]. It should be noted that the view fits especially well into the rest of Yagisawa's modal metaphysics, since he already has an ontology of impossibilia and takes counterpossible conditionals seriously. The latter feature is probably essential for motivating the view; the former just makes it easier to hold, perhaps for metasemantic reasons. Daniel Nolan [1998] has however made it a lot easier for actualists to take counterpossible conditionals seriously.

For Yagisawa, fictional characters are impossible because their transworld identity conditions violate the metaphysical laws obtaining in the local possibility space. Suppose that at some Holmes worlds Darwin plays the Holmes role and at others Gladstone does. In Yagisawa's system, this means that Holmes overlaps Darwin and Gladstone. Since Holmes is a person, he obeys the transworld identity laws applying to people in his local space. If his local space obeyed our laws, that would make the Darwin stages and the Gladstone stages belong to the same person, so Darwin would be Gladstone. Since he is not, the stages must belong to a different space, and worlds in that space are impossible. Since all of Holmes' stages are at impossible worlds, Holmes is impossible. Yagisawa takes himself to have established [2010: 273-6] that all of Holmes' stages are at impossible worlds, although as far as I can tell he only establishes that his stages at possible worlds, if any, overlap at most one possible person. Nonetheless, Kripke's points that we would not know which and it would be unpalatably arbitrary which it was still stand in Yagisawa's framework.

On a slightly more committal extension of Parsons' theory, Holmes is impossible for a different reason, but still one relating to Kripke's point about the stories leaving a lot open. For Parsons, Holmes is actually a non-existent incomplete concrete object. He has the properties he has in the stories, but where the stories leave it open whether Holmes is *F* or not, he will neither have the property of being *F* or of being not-*F*. Parsons [1980: 186] is agnostic as to whether Holmes necessarily exhibits these property gaps, or only actually. If we hold that he does so necessarily, as Fine [1984: 125-6] argues that Parsons really should, then he will be necessarily non-existent, since existent objects cannot exhibit property gaps. Both of these positions take the impossibility to derive from the underspecification of the stories, and as such they are in keeping with Kripke's argument. They go different ways with this though: for the development of Parsons we leave the gaps unfilled, while for Yagisawa we fill them in too many different ways.

The two main objections to taking fictional objects to be impossible objects are that it might be hard to argue for without an implausible metaphysics and that it does not do justice to the idea that the stories are possible. After all, according to the stories Holmes is not an incomplete object, and he obeys the same transworld identity laws as the rest of us. It may also be difficult to give a plausible account of how reference to impossibilia could be secured. There are things which can be said, but I will end up rejecting this position.

It is also worth bringing up at this point a possible connection between the view that fictional characters and kinds are impossible and a superficially similar feature of Armstrong's combinatorial theory of possibility. Armstrong [1989: ch. 4] holds that there could have been individuals that do not actually exist (rejecting the Barcan formula), but there could not have been universals instantiated which cannot be constructed out of actually instantiated universals. Lewis [1986: 158-65] takes it to be an argument in favour of his modal realism that it can accommodate the possibility of such alien universals. Armstrong embraces their impossibility as a consequence of his actualism and naturalism. Nothing in the world makes them possible, so if there is nothing outside the world then they are not possible. He sees no impossibility in there having been being fewer universals than are actually instantiated though, and if *w* accesses *v* iff there are no universals in *v* which are alien to *w*, we get a transitive non-symmetric relation generating an S4 modal logic. (Note that the accessibility relation goes in the opposite direction from how Dummett's went.) Perhaps there is some mileage in connecting Kripke's argument and Armstrong's, although it is unclear to me quite how this would go. Even if no connection can sensibly be made it is still worth drawing attention to the superficial similarity and pointing out that the positions should not be lumped together.

- *4.226: Epistemicism and indeterminacy*

Kripke divided his conclusion into two, but in view of the previous response we should really divide it into these four:

S1 There is nothing that fictional terms refer to.

E1 There is nothing we can know that fictional terms refer to.

S2 There is nothing possible that fictional terms refer to.

E2 There is nothing possible that we can know fictional terms refer to.

These have been left in terms of possiblist quantification for the sake of simplicity, but we saw above that there are actualist formulations getting at the same ideas. Since knowledge entails truth but not conversely, S1 entails E1 and S2 entails E2 but not conversely. Since S2 and E2 restrict the claims to possible things, S1 entails S2 and E1 entails E2. Kripke holds all of them, but one could also hold just S2 and E2 or just E1 and E2. It is unlikely anyone would just hold E2, because the tie-breaking problem probably does not apply to impossible things, which can be incomplete (following Parsons) or have eccentric transworld identity conditions (following Yagisawa). The entailments summarized:

$$\begin{tikzcd}
& \mathbf{S1} \arrow[dl] \arrow[dr] & \\
\mathbf{S2} \arrow[dr] & & \mathbf{E1} \arrow[dl] \\
& \mathbf{E2} &
\end{tikzcd}$$

Kripke suggests that E1 gives a reason for accepting S1, although it does not logically entail it. From the quote earlier:

> One cannot say when these sentences would have been true of a counterfactual situation, and therefore no proposition has been expressed. [Kripke 2011b: 67]

This seem to appeal to a principle saying something like this: if we cannot know what something means then it cannot mean anything. If we cannot know what proposition is being expressed then no proposition is expressed, and if we cannot know what is being referred to then nothing is referred to. So if we accept E1 we should also accept S1, and if we accept E2 then we should also accept S2. Perhaps it is overinterpreting Kripke to attribute this principle to him in any strong or general form, but something in the vicinity seems to be at work here and something in the vicinity has some plausibility. Successful communication involves at least someone knowing what is being said, doesn't it?

This way of taking the argument sees the underspecification in the stories as not providing enough information about what is being said, which causes communication to fail, which causes reference to fail. Underspecification secures the epistemic thesis, which results in the semantic thesis. Another way of taking the argument sees the underspecification as meaning not enough metasemantic work is being done, which secures the semantic thesis, which entails the epistemic thesis. We can consider the difference between two ways we might think demonstrative reference could fail. One is if I point in the vague direction of several men and say 'him': perhaps not enough work has been done to secure reference, so I refer to nobody, so there is nobody anyone can know I am referring to. Another, from G. E. Moore [2004], is if I point into a dark room containing an unseen (and unhearing) man and say 'him': nobody can know who I am referring to, so perhaps I am not referring to anyone.

Suppose we grant the epistemic thesis: there are many candidate references for a fictional term and we cannot eliminate any of them from our enquiries without parity of reasoning eliminating them all. Must we accept the semantic thesis too? The entailment is not logical, and consideration of other cases of underspecification might lead us to reject the move to the semantic thesis. The cases I have in mind are those of vague terms. It seems that not enough has been done for us to know where the line between the tall people and the rest is, and maybe even that not enough has been done to guarantee that the line even is in any particular place. The tall people include everyone over two metres and no-one under one, but after a point it seems arbitrary to put a cut-off in one place rather than another. Do we conclude that reference has failed, and that no propositions are expressed by sentences containing the word 'tall'? Well, that is exactly what David Braun and Ted Sider [2007] do conclude. Theirs is a minority position though, and even they see the need to do some work to explain why we talk the way we do. They say that we can mostly ignore the vagueness for practical purposes because all the candidate meanings give the same result. This pragmatic explanation is much more conciliatory than just saying that Sherlock Holmes could not have existed: if all the (salient) Holmes candidates could have existed, can't we just ignore the underspecification for the same reason and treat 'Holmes could have existed' as if it expresses one of the true propositions in the vicinity?

Most people do not even go as far as Braun and Sider, of course. Most people say vague sentences do express propositions, because so much language is vague that the Braun-Sider position can seem unthinkable. We can follow Williamson [1994] and say the line is sharp but we can't know where it is, follow Delia Graff Fara [2000] and add that it moves around according to context, or we can say it is indeterminate where the line is. If we go down the indeterminacy route, then when a story is properly interpreted as presenting events which could have happened the candidate meanings will all be possible, and it will be determinately true that Holmes and unicorns could have existed. We also get the perhaps pleasing result that while we cannot say of any possible species that they would be the unicorns, there are also many possible species which we cannot (determinately truly) say would not be.

Earlier we stated the principle connecting the epistemic theses to the semantic ones as 'successful communication involves at least someone knowing what is being said'. Consideration of vagueness suggests that this version of the principle is too strong. We could replace it with the following: 'successful communication involves at least someone having some idea what it being said'. If we want to hold on to the strong principle, we will have to either side with Braun and Sider on vagueness or explain why not. If we reject the strong principle we do not have to take vague terms as non-denoting, and we have a choice to make about fictional terms. They can be assimilated to vague terms, and either have unknowable or indeterminate reference, or they can be assimilated to failed demonstratives. To get specific about what an analogous demonstrative would be, we need one which refers to nothing because there is more than one thing such that if the other candidates had not been there then the demonstrative would have referred to it. It is not implausible that there could be such cases, although it might follow from some theories of demonstratives that there could not be. A possible example would be where you gesture vaguely towards two men and say 'that man'.

The obvious argument for assimilating fictional terms to failed demonstratives deploys a historical explanation view about reference-transmission: you cannot refer to something unless there is a chain of communication going back to the thing being named, and the histories of fictional terms do not go back to something being named. They go back to somebody making up a story. However, we could disarm this reasonably simply by saying that in telling the story the author presents a way things could have been and refers to the characters in it, but underspecifies which way they are presenting, with the epistemicist or indeterminist consequences this has. The fictional terms thus do go back in a chain of communication to the possible world presented, and refer to the relevant objects in that world, with respect both to the world itself and to other worlds in which those objects appear.

Another case we might consider, besides vague terms and demonstratives, is that of counterfactual conditionals. The antecedents of counterfactuals do not typically specify a particular possible world. If the consequent is true at some of the candidates and false at others, we have to decide what to do. Robert Stalnaker [1981: 90-91] addresses the problem of there being multiple candidate worlds to evaluate by supervaluating across all of them. Lewis [1973] takes the counterfactual conditional as a kind of necessity-like operator, in that the conditional is false if the consequent is false at any of the candidates. We might want to bear in mind our response to this problem when deciding how to deal with the problem of multiple candidate references for fictional terms. In fact, Kaplan explicitly links his version of the argument to counterfactual conditionals. He says 'the critical invalidity is $[(\varphi \rightarrow (\psi \vee \chi)) \rightarrow ((\varphi \rightarrow \psi) \vee (\varphi \rightarrow \chi))]$ where '→' symbolizes the subjunctive conditional' [1973: 517]. The point here is that we cannot speak of *the* world where a fiction's Ramsey sentence is true, just as with a subjunctive conditional we cannot speak of *the* world where the antecedent is true. Some people like Lewis agree that this principle is invalid. Stalnaker's system upholds the principle though, and this is not obviously a mistake. It is true that we cannot speak of the unique world where the antecedent is true, but it begs the question to say we cannot speak of the unique world which would have been actualized if the antecedent had been true, and that is what matters. Note that even Lewis thinks we can speak about such a unique world in cases where the antecedent actually is true: it is the actual world[71]. Kaplan might be wrong that his argument relies on the invalidity of a principle of conditional logic, but if he is right, the argument should be controversial because the principle is controversial.

Whatever we say about counterfactual conditionals, the case of vague terms suggests that Kripke's argument only forces the epistemic

[71] The principle is closely connected to conditional excluded middle: $[(\varphi \rightarrow \psi) \vee (\varphi \rightarrow \neg\psi)]$, which is entailed by the principle Kaplan rejects and the necessity of excluded middle. However, if we modified a system along Stalnaker's lines to accommodate failures of (unconditional) excluded middle, for e.g. intuitionist or vagueness-related reasons, it would no longer validate conditional excluded middle but it would still validate the principle Kaplan rejects.

conclusion, although it gives reasons for the semantic conclusion. There could be other arguments for the semantic conclusion, but these could not rely on the underspecification issue alone. It seems that underspecification sometimes leads to referential failure, but sometimes only leads to either ignorance or semantic indeterminacy. It would be good to be able to deal with either case, and my positive account in §4.431 will have this feature.

- *4.227: Summary*

It seems to be agreed by all parties to this debate that for all the stories say there could have been various species occupying the unicorn role, and various people occupying the Holmes role. Kripke argues that this means that not enough has been done to make 'unicorn' and 'Holmes' refer with respect to any worlds, and that even if some arbitrariness did pick up the slack then we would not be able to know which referent had been selected. He concludes that unicorns and Holmes could not have existed, but this is puzzling because it seems that the stories are possible and according to the stories they do exist. We have seen several ways of responding to Kripke's argument.

The metaphysical option takes fictional terms to denote impossible things. The epistemic option takes them to denote possible things, but within the bounds left open by the stories we cannot know which things. The semantic option takes the terms either not to refer at all, or to refer indeterminately.

If we disagree with Kripke, we could abandon rigid designation altogether in favour of descriptivism, but I am trying to avoid going down that road. We could however embrace a descriptivist semantics just for fictional terms. This can be defended in the context of a suitable metasemantics, but it generates some ugly results. Alternatively, we could keep the references of fictional terms rigid but allow the modal profiles of the references themselves to vary from world to world. That is one way of interpreting Dummett, but involves some substantial commitments about the metaphysics of properties.

The situation is complicated, and all the available solutions have some *prima facie* unlovable features, which must either be tolerated or explained. Probably some tolerance will be necessary, but hopefully our positive account will be able to explain as much as possible, and these explanations should improve our understanding of how fictional terms work, and what we mean when we use them.

- **4.23 Double Lives**

- *4.231: Caulfield's fame and other problems*

In the last section we alluded to there being at least two uses of fictional terms: one to refer to artistic creations, which exist and are abstract, and one to either pretend or try (unsuccessfully) to refer to concrete things which (in fact) do not. That is Kripke's way of doing things, but not everyone follows him in this. Zalta seems to hold that in all the cases the fictional terms are used to refer to the fictional character. However, even if we do not make the distinction Kripke's way, we have to make it somehow, to deal with what I will call the *double lives* problem. That problem is the topic of this section.

As we saw in §4.1, different theories of fictional characters have different resources to deal with this sort of problem. Zalta's theory makes use of the distinction between two ways objects can instantiate properties: encoding and exemplifying. Parsons talks about two different classes of properties, nuclear and extranuclear, and gives extranuclear properties watered-down nuclear equivalents. Yagisawa can distinguish between the properties fictional characters have in different worlds. Van Inwagen can distinguish between the properties fictional characters have and the properties ascribed to them by the stories they come from. Kripke distinguishes assertions about abstract things from corresponding pretences ostensibly about concrete things.

If it was just a problem about whether to say the characters existed or not, we could probably just distinguish between two senses of 'exist' and have done with it. We cannot do that though, because the double lives problem is more wide-ranging and systematic than that. It arises in a few forms.

First, there is a generalized version of the existence problem: there are several kinds of property besides existence which we might want to attribute to the fictional characters in their roles as artistic creations but not necessarily in their roles as concreta. The artistic creations, since they are abstract, will automatically have lots of negative properties: not having been kicked, not being alive, not being extended in spacetime and so on. These are, at least for the most part, the negations of what Priest [2005: ch. 3] calls 'existence-entailing properties'. (Priest holds that abstracta are non-existents.) The creations also get attributed some properties they do not have automatically by being abstract, though. These are the kinds of property Parsons [1980: 23] classes as extranuclear. To recap, he divides them into ontological properties like existence, modal properties like possible existence, intentional properties like being famous, worshipped or thought about by Parsons, and technical properties springing out of whatever theory of fictional characters we have adopted. Note that these properties are frequently the kind of properties that the characters can have in the stories, too. The main category of properties which characters can interestingly have or lack in either of their roles is the intentional properties. Holden Caulfield, the protagonist of *The Catcher in the Rye*, is a famous creation but not famous in the stories. Poirot is famous both as a creation and in the stories. Examples of fictional characters who are not famous in reality are necessarily obscure, but it is clear that they are possible, and that their fame in the stories is independent of their fame out of them.

This part of the problem is where distinctions like those between encoding and exemplifying and between neat and watered-down properties are most comfortable. The theories give us two kinds of relation which an object can stand in to a property, and we say that whether one relation obtains between a fictional character and a property is independent of whether the other does. Matters are more complicated than this though, and the complicated cases can put pressure on the theories. Two pressure points are where objects are in danger of appearing in a story both as abstract and as concrete, and where fictional characters depend on or stand in relations to other fictional characters. These two cases are discussed in §4.232 and §4.233. In §4.234 I examine a potential paradox in Zalta's system arising from these issues, and in §4.235 I explain what kind of distinction a theory needs to be able to make to solve the double lives problem effectively. Then in §4.3 I will be able to sketch a positive theory of fictional characters which deals with the problems we have discussed so far.

- *4.232: Kripke's story about Sherlock Holmes*

Kripke [1980: 157-8] tells a story about Doyle writing his stories and their coincidentally matching up perfectly with actual people and events. Kripke tells this story in connection with the issue about unicorns discussed in the last section, but it also raises an important aspect of the double lives problem. Some of the things we say about this problem will relate to what we say about the unicorns problem, but they are distinct. When we say 'Holmes could have existed', we are talking about Holmes in his concrete role: as a part of how the concrete things could have been. Now, we might think that if there was someone who by coincidence actually filled this role, then this use of 'Holmes' would refer to him. Kripke does not want to say this, but Dummett says this at least about unicorns, and maybe it is what we want to say about some cases. Nonetheless, even if we say this, we can still hold that the other use of ‘Holmes’ refers to an abstract artistic creation. This is necessary to vindicate our saying ‘Doyle created/selected Holmes’, and indeed our knowing anything much about how Holmes actually is.

Now, supposing we do want to maintain that in Kripke’s story there is still an artistic creation as well as a concrete detective, we have the makings of a problem. Everything determinately true in the Holmes stories is determinately true in Kripke’s story, but there is the additional information that an author – let’s call him ‘Conan’ – wrote a story coincidentally matching the events recounted in the Holmes stories. Kripke’s story leaves open strictly less than Doyle’s stories leave open. Perhaps we want to say that this makes Kripke’s detective different from Doyle’s detective, and since Conan’s stories are the same as Doyle’s we can say that the two roles are filled by the two detectives. To close this loophole, let’s have Conan add Kripke’s postscript to his stories, saying that someone just like him coincidentally wrote some stories in which all the events recounted played out. To be clear, here is the story, which we will still call ‘Kripke’s story’:

> The events of Doyle’s stories occurred, and someone called Conan unknowingly wrote stories like Doyle’s as fiction, but with a postscript adding that an author unknowingly wrote stories just like Conan’s (including the postscript) as fiction.

Now we can state the problem. In this story, we have three objects. They are a novelist, Conan, a detective we can call Sherlock, and an artistic creation we can call Sherlock*. Conan’s story is the same as Kripke’s, but Sherlock* must not be the same as Sherlock, because Conan created/selected one but not the other, one but not the other is concrete and so on. But we might think that Kripke (in reality) and

Conan (in the story) are confronted with the same abstract objects as each other, and so Sherlock* and Sherlock must be the same. This makes sense with numbers: Poirot has the same number of heads as I do, viz. one. If, like Zalta, we want to fit fictional characters into a system of necessarily existing abstract objects, these abstract objects should appear in the stories just as they appear in reality. When Kripke writes a story that selects an abstract object, and when Conan writes the same story it selects the same object. Or so we might like to say.

Now, this example is a little Byzantine, and there are several fixes one could try in response to it. When producing a fully worked out theory of fictional characters we might need some sort of fix, and an *ad hoc* fix for this one problem is probably not difficult to find. We want a principled fix though, and if we do not understand the problem properly then any fix can be expected to run into trouble later. As such, I will try to state the problem in a more intuitive way.

The worlds of fictions will contain abstract objects, just like the real world does. We might well want to say that these are the same objects, imported into the fiction: Holmes lives in London and Holmes has one head, so London and the number one are real objects imported into the fiction. If fictional characters are abstract objects too, then it seems they will be able to be imported into stories too. But if that happens, there will be a danger of them running into themselves. In that case it will not be enough to say that they have one set of properties in reality and one in the stories: they will have one set of properties in reality and two in the stories. That might be more than a distinction like Zalta's or Parsons' can handle.

We should note also that problems might arise even for stories which do not involve people writing coincidentally accurate fictions. According to Zalta's theory and any other selectionist theory, Holmes would have been out there even if Doyle had not written his stories.

Suppose that we precisify Doyle's stories to say no such coincidental stories were written about Holmes. In that story, Holmes the detective will be famous and Holmes the abstract object will be unheard of. This will not be an issue for creationists, because for them Holmes the abstract object exists only if the stories are written. Kripke's story (or our modification of it) might still be a problem for them though, and as we saw in §4.21, it is as well to have a coherent plenitudinous theory of fictional characters to cope with the available types, even if ultimately we want to identify characters with the tokens and be creationist about those. Fine [1982: 120] also makes the point that if we can't have a coherent plenitudinous theory, even a creationist theory risks having its coherence be contingent on the plenitude of objects, or the incoherent parts of it, not getting created.

- *4.233: Interdependent fictional characters*

Garlic is bad for vampires and kryptonite is bad for Superman. Holmes is from England and the Daleks are from Skaro. If we have a theory which, like Parsons' or Zalta's, holds that fictional characters are somehow defined by or dependent on the properties they have in the stories they come from, then these properties will probably include properties involving other objects. Sometimes these objects will be real, like garlic and England, and sometimes they will be other objects from the stories, like kryptonite and Skaro. Where the properties involve real objects this does not create any obvious problems: we already have the objects and on a suitably abundant conception of properties we can have the properties too. However, where the properties involve objects from the same stories as the characters we are using them to define, we might have a problem. Parsons discusses this issue in his [1980: 194-7].

One of Superman's properties is expressed by 'kryptonite is bad for x' and one of kryptonite's properties is expressed by 'x is bad for Superman'. If we were thinking of the fictional characters as constructed in stages, as with the hierarchy of sets, we would not be able to construct Superman until we had kryptonite and we would not be able to construct kryptonite until we had Superman. Perhaps that is not the way we want to think about fictional characters. We could just say that since there are those objects, there are those properties, and the objects make the relevant instances of the comprehension schema for fictional objects true. There is a worry of arbitrariness here though: if neither of two co-dependent objects existed, neither of them would have to exist. In fact, since all the objects native to a story will have the properties of co-existing with the others, there is a case to be made that they would all be co-dependent, and so we could not have any guarantee that the arbitrary process supplied them rather than not. Our comprehension schema by itself would not supply the objects for any story with more than one character native to it.

Perhaps we could get round this with some kind of principle saying that the default position was having more objects rather than fewer, although this would run into trouble if there were multiple different extensions of the theory each of which could not be extended further. Even if the technical problem was soluble though, a metaphysical problem would remain about dependence. If fictional objects depend on their nuclear properties, and object-involving properties depend on the objects they involve, then we would have circular chains of ontological dependence. We might not want those. Kit Fine [1982: 129; 1984: 118] offers to solve this problem for Parsons by defining groups of objects together, so that they all depend on the same properties and not on each other. The system sketched in §4.3 will deal with the problem along similar lines.

We have a further problem about fictional characters being defined by properties involving other fictional characters though, and this relates more closely to the double lives problem. I think about Sherlock Holmes, and Watson thinks about Sherlock Holmes, but what I am doing is different from what Watson does. I am thinking about an abstract object in a story, and Watson (according to the story) is thinking about a detective. Christopher Boone, the protagonist of *The Curious Incident of the Dog in the Night-time*, also thinks about Holmes, but what he does is like what I do and not like what Watson does. It would be good if our theory brought this distinction out.

If Holmes and Watson are defined or generated together, we get to make this distinction. The set of properties Holmes, Watson and the other characters in their stories depend on does not include Holmes-involving properties or Watson-involving properties. The set of properties generating Boone can include Holmes-involving properties like thinking about Holmes. However, we have a further complication, which is that fictional characters appear in other works not just as fictional characters, but as real people[72]. Suppose Boone met Poirot and thought about him. Then Boone would have to be generated from Poirot-involving properties too, but to do justice to the way he thinks about Poirot being different from the way he thinks about Holmes, we

[72] The status of immigrant characters in stories is delicate. It is certainly possible to have a fictional character based on an actual person but distinct from them; for example Charles Strickland in *The Moon and Sixpence* is based on but distinct from Paul Gauguin. Likewise, fictional characters can be based on but distinct from other fictional characters; for example Simba in Disney's *The Lion King* is based on Hamlet. Giving the derivative character the same name as the source should presumably not automatically make the characters imported, and it is probably natural to say that Marvel's Thor is distinct from Norse mythology's Thor, and maybe Philip Roth's Philip Roth is a fictional character distinct from the real Philip Roth. We can argue over cases, but the conceptual space is there for fictional characters based on but distinct from either real characters or other fictional characters. These create no new problems. However, the conceptual space is also there for both real characters and characters from other fictions appearing in fiction, and the positive account in §3.3 will attempt to accommodate them.

should have the two properties playing a different role in the way they generate Boone. The theory in §4.3 will try to do that.

Another complication is that we might have two fictional characters reading each other's stories, if two stories were written with sufficient co-operation. There the holistic generation of characters seems less appropriate. What should we do? Well, let's examine the scenario a bit more closely. Suppose that in *Othello* Desdemona went to see *Hamlet*, and in *Hamlet* Ophelia went to see *Othello.* Then Desdemona would depend on Ophelia and Ophelia would depend on Desdemona, or so it seems. But this is a very odd case. Desdemona goes to see a play in which one of the characters goes to see a play in which she, Desdemona, is one of the characters. This would be like me going to a play in which the characters see a play in which I am one of the characters. That would be bizarre, of course. I could not actually be one of the characters in a play, so I cannot be native to the play they are watching, if it really features me. Thus we could say that Desdemona watches a play like *Hamlet* but distinct from it, and Ophelia watches a play like *Othello* but distinct from it. In §4.3433 we will see that the system I propose may be able to accommodate the particular case a little more smoothly than that, but it is unlikely that all cases involving this sort of circularity can be fully accommodated.

- *4.234: The modesty paradox*

Is there a property of non-self-instantiation? You might think there was: some things, like the property of being a property or the property of being such that 2+2=4, do not have non-self-instantiation, but most things seem to. Nonetheless, if there is such a property, it seems to lead to a paradox. If it does not instantiate itself then it instantiates itself, and if it does then it does not. Perhaps we should try to construct our ontology of properties without it; in any case our response to this paradox will be informed by our responses to the liar paradox, Russell's paradox and Grelling’s (heterological) paradox, since it is similar to all of those.

Zalta’s theory of fictional characters fits into his general theory of abstracta, and that theory includes a theory of properties. He has to be careful to avoid the problem of the non-self-instantiation paradox. Ignoring the restriction Zalta imposes to avoid the paradox for now, let’s try to construct the paradox within his system. Abstract objects can encode properties as well as instantiating them. We write *a* encodes *F* as *aF*. Properties are typically expressed by predicates, but we can have stand-ins denoted by terms. When an object encodes exactly one property, we can think of that object as the Platonic form of that property. This is just what Zalta [1983: 42] does. He defines property identity, written F=G, as holding between two properties whenever they are encoded by just the same objects [1983: 13]. We can express that *x* is the form of the property *F* with this open formula:

$$x = \iota y[A!y \mathbin{\&} \forall G[yG \leftrightarrow G=F]]$$

Informally, this says ‘x is the thing which is abstract and encodes all and only properties identical to *F*’, or ‘x is the abstract object encoding just *F*’. We can abbreviate ‘$[A!y \mathbin{\&} \forall G[yG \leftrightarrow G=F]]$ as $[\Phi yF]$, so ‘the form of F’ is rendered as $\iota y[\Phi yF]$. We can say that a property F is not instantiated by its form: $\neg F\iota x[\Phi xF]$. We can also say that an object x is not a self-instantiating form: $\neg\exists F[\Phi xF \mathbin{\&} Fx]$. Using a lambda formula we can turn this into a predicate expressing the property of not being a self-instantiating form: $[\lambda x.\neg\exists F[\Phi xF \mathbin{\&} Fx]]$. Now we can define the form of non-self-instantiation, *n*:

$$n = \iota x[A!x \mathbin{\&} \forall F(xF \leftrightarrow F=[\lambda y.\neg\exists F[\Phi yF \mathbin{\&} Fy]])]$$

In full, this will be:

$n = \iota x[A!x \,\&\, \forall F(xF \leftrightarrow F=[\lambda y.\neg\exists F[[A!y \,\&\, \forall G[yG \leftrightarrow G=F]] \,\&\, Fy]])]$

Does *n* instantiate the property it is the form of? That property is instantiated by all and only properties that do not instantiate a property they are the form of, so *n* instantiates its property iff it does not. We have a paradox. Zalta understandably imposes a restriction on his system in order to avoid this problem. Here is his comprehension schema [1983: 34-5]. Where x is not free in φ:

$\exists!x(A!x \,\&\, \forall F[xF \leftrightarrow \varphi])$

Ordinarily φ will have only F free, and so express a condition on properties. To avoid paradoxes, Zalta imposes a restriction on his language which ends up imposing a restriction on φ: lambda predicates cannot be constructed from formulas containing either encoding subformulas or quantifiers binding predicate variables [1983: 18]. The lambda predicate in the definition of *n* breaks both rules, so no paradox arises.

Can we construct a different paradox? Let's try. Of all the objects there are, some will encode liking themselves, and some will not. I do not mean encoding self-liking, but rather encoding the object-involving property which happens to be liking yourself. In Zalta's system Holmes encodes liking Holmes while Obama does not encode liking Obama. Let's call objects which do not encode liking themselves *modest*, symbolized as 'M', and symbolize 'likes' as 'L'. Now we define an object, Larry, which encodes just liking all the modest objects:

$\text{Larry} = \iota x(A!x \,\&\, \forall F[xF \leftrightarrow \exists y(My \,\&\, F=[\lambda z.Lzy])])$

This says that whenever an object does not encode the property of liking itself then Larry will encode the property of liking that object, and Larry will not encode any other properties. Does Larry encode liking

Larry? If not then it is modest, so it will encode liking Larry, but if it encodes liking Larry then it is not modest, so it will not encode liking Larry. We have a contradiction, so there can be no object defined as Larry was defined. But there were no encoding subformulas or predicate quantifiers in the lambda formula, so the formula defining Larry was an instance of the comprehension schema.

What can we say? Well, ‘M’ was in the formula, and ‘M’ was defined as not encoding liking yourself, which would be this predicate:

$$[\lambda x.\neg x[\lambda y.Lyx]]$$

That is banned because it contains an encoding subformula. Does this resolve the issue? Not really, for two reasons. First, ‘M’ is just a simple predicate. Perhaps we can say that since it is synonymous with a banned expression, it is itself banned. I am not sure how we would go about hunting down all the banned expressions, since many simple predicates were defined long ago, but let us suppose that we can do this. We get a problem now though, because ‘Holmes’ is the object encoding all and only the properties in the Holmes role, and as such is equivalent to an expression with encoding subformulas. We want to say that Watson encodes liking Holmes, but now we cannot, because that expression is this:

$$w[\lambda x.Lxh]$$

If ‘h’ cannot appear in lambda terms, we cannot say this. If it can, on what grounds was ‘M’ banned from the language? But the cases are not quite the same: ‘M’ had a banned definition, whereas ‘h’ had an allowable definition such that if we replaced ‘h’ with its definition would bar the resulting expression from appearing in a lambda formula. This is not pretty, but it gives Zalta a place to stand.

However, the paradox can return even if ‘modest’ is not synonymous with ‘not encoding liking yourself’. Can we deny that a predicate in the language could apply, contingently, to all and only the things which do not encode liking themselves? The things are out there, and what if they all just happened to be God’s favourite things? (Or mine?) The comprehension schema has no magical force; it is just something which Zalta thinks is true and entails the existence of all the abstract objects there are. Since there is no object encoding liking all the objects which encode liking themselves, if they are God’s favourite things, there is no object encoding liking all and only God’s favourite things.

Perhaps we think it very unlikely that our language contains a predicate which could combine with Zalta’s comprehension schema to generate a false instance. If we are falling back on that though, we see that it only entails the existence of the objects it does because our predicates have the extensions they contingently have. It is therefore unlikely that the schema entails the existence of all the abstract objects there are, although that was supposed to be a point in favour of believing it (see the quote in note 46, above).

It seems that Zalta’s theory is in some trouble, and there are two issues to think about as a result of that. One is how to fix Zalta’s general theory of abstract objects in general in response to the modesty paradox, and the other is how to stop our systematic theory of fictional characters falling foul of something similar. We will deal with the first issue first.

The modesty paradox is not just a technical thing, and indeed we saw that it might contain some technical loopholes. Escaping through these loopholes will not do much good though, because it raises a systemic worry about theories like Zalta’s. Some of our predicates have contingent extensions, and the comprehension schema uses those contingent extensions to determine which collections of properties are necessarily encoded by abstract objects. If we relied on the contingent extensions we would most probably not get all the collections we needed, and if we used all the possible extensions we would run into paradoxes. Probably something can be done, but we will not explore exactly what works here.

Before leaving the topic though, it is worth drawing attention to some similar problems, and in particular Kripke's [2011c] paradox about time and thought. Kripke argues that if we want to say that the times form a set then a paradox threatens. For every set and expressible condition on objects, there is meant to be a subset of the set containing just those members satisfying the condition. (This is just a version of the separation axiom for set theory.) Suppose that the set is the set of times, and the condition is that *t* be a time at which he (Kripke) is thinking of a set of times of which *t* is not a member. Since the 1960s, Kripke has sometimes thought about the set of times satisfying that condition. Are those times in the set or not? They are iff they are not, which leads to a contradiction[73]. In his discussion, Kripke notes that his paradox is similar to Kaplan's [1995] paradox for possible worlds semantics. The normal response to Kaplan's paradox is not to give up on using possible worlds semantics, and time will tell how people respond to Kripke's paradox. In any case, the way defenders of Zalta's theory respond to the modesty paradox and the general problem it raises should be informed by their responses to Kaplan's and Kripke's problems.

For present purposes we can scale our ambitions back from a systematic theory of all abstract objects to a theory of fictional characters. We need to make sure our theory does not fall foul of the modesty paradox or something like it. This is where the double lives

[73] This technique can be used to generate a predicate to use for 'M' in the modesty paradox. At *t* I think – contingently of course – about the set of objects not encoding liking themselves, and then we use the predicate 'x is a member of the set I am thinking about at *t*'

problem comes in. The problem arose from properties involving abstract objects. If we think there are such objects at all, then these objects will in some sense be already out there and things can have properties involving them. The same goes for properties involving fictional characters: I have the property of thinking about Holmes. We must also allow that fictional characters can have these properties in the stories, because of cases like Christopher Boone thinking about Sherlock Holmes. However, when Boone has this property it still involves Holmes as a fictional character. When we go into Holmes' world and look at the properties of characters in that story, including Holmes himself, Holmes is no longer present as a fictional character. In that world Holmes is a detective instead, and not among that world's stock of fictional characters. They can have a fictional character a lot like Holmes, but Holmes himself is busy being a detective. Larry's problem was like Conan's problem. Conan thinks about Sherlock* but not about Sherlock. Larry is like a character in a story who likes all the objects which do not encode liking themselves. There is no contradiction there: as a concrete character he does not encode liking himself, which places him among the things he exemplifies liking. But if he was also in that world encoding *qua* abstract object all the properties he exemplifies *qua* concrete object, we would have a problem. We need an account which takes into account the way different stories' stocks of concrete objects generate different collections of fictional characters, and which does not give anything both roles in the same story. We examine this desideratum further in the next section.

- *4.235: Abstract and concrete natives and immigrants*

We need to have a concept of a character being *native* to a story as opposed to being an *immigrant*, the terminology of Parsons [1980: 51]. The idea has been broadly assumed so far, but we should examine it more closely. When you tell a story you can import characters, either from the real world or from other stories, and you can also make up new characters. The natures of the immigrant characters will not be changed by the story, except insofar as having a story about you is itself a change. Their intrinsic natures are determined by what they are like in reality, if they are real, or what they are like in the stories they were native to, if they are imported from other fictions. We can distinguish between immigrants and new characters based on old, as with Marvel's Thor, and if we want to do justice to the continuity of Thor's incarnations we can research intertextual genealogies; we don't have to say the characters are strictly identical. The choices to be made in this kind of case should not generate problems relevant to the present project. (For more on Thor see note 59, above.)

Often conflicts between the properties of characters in different stories can be resolved by saying the character is native to one story and an immigrant in the other. It may not always be appropriate to resolve conflicts in this way, though. Consider Kripke's story from §3.232. The fictional detective determined by Kripke's story has different properties in the story from that determined by Conan's story, but since they seem to be the same story they seem to be the same character. But this seems impossible: Conan created one and not the other, one solves crimes and one is fictional and so on. We could try to resolve this conflict by saying the character is native to one story and an immigrant in the other. But should we do this? It seems that Kripke's story is the origin of Sherlock, a fictional detective, and Sherlock*, a fictional fictional detective. There is a sense in which both are native to Kripke's story, but there is also an apparent sense in which Sherlock* is native to Conan's fictional story too. How do we resolve that?

I propose we distinguish between two ways an object can be native to a story. It can be a *concrete native* if the story says it is one of the concrete objects, and it can be an *abstract native* if it would be one of the abstract objects generated by the world of the story. The plenitude of available fictional characters is, according to Zalta and Parsons and perhaps everyone else with a systematic ontology of fictional characters, determined by what the world is like and could be like and what things it does and might contain. But if Kripke is right about unicorns and (especially) Sherlock Holmes, fictional worlds contain objects which do not exist in the real world, and could not. These new objects can generate new possibilities, which in turn generate new fictional characters. The concrete natives are the new concrete objects introduced into a story, and the abstract natives are the new fictional characters which are generated as a result. Sherlock is a concrete native of Kripke's story, and Sherlock* is an abstract native of the same story.

As well as making a distinction between the concrete and abstract natives of a story, we can also make a similar distinction among immigrants. Suppose again that Christopher Boone thinks about Sherlock Holmes and meets Poirot. Holmes is an abstract immigrant to this story and Poirot is a concrete immigrant. Holmes will appear in the story by Boone having a property, thinking about Holmes, which is native to our world. Poirot will appear by having his own properties in the story, although only as an immigrant. The ontology sketched in §4.3 will be constructed so as to make the distinction fall out of it naturally.

4.3 A Sketch of a Fictional Ontology

- **4.31: Introduction**

We have seen some reasons to think there are fictional characters, some of the systems on offer, and some issues and choices those systems have to face. Now we are ready to put these things together and sketch out a system which should be able to solve those problems. It is only a sketch, and as such will have some limitations which will be laid out in §4.34. A fuller development of it will be left for further work. The sketch will however try to solve the problems we have discussed in a principled way, which gives reason to hope that a fuller development along the same principles would inherit its solutions to these problems.

Even a full development of the position sketched in this section would however have some limits to its ambitions. As already noted, there is a distinction between fictional characters as abstract creations (or selections) – as theoretical entities of literary criticism – and fictional characters as concrete things which we only pretend exist. Van Inwagen's argument that there are fictional characters applies only to the former. Once we have them in our ontology, of course, we could press them into service in an analysis of the latter kind of discourse, and this is, at least to an extent, what Parsons and Zalta do. It is not what I will do. I will follow Parsons and Zalta in setting up a systematic ontology of fictional characters, but I will follow Kripke in using them only as theoretical entities of literary criticism. For the other kind of discourse I will offer a pretence account, which I will set out in §4.4. The ontology sketched below will include elements from Zalta [1983], Parsons [1980] and Fine [1982, 1984], and some elements are intended as new.

- **4.32: Informal presentation of the theory**

The most standard way of avoiding paradoxes in set theory is to have the axioms for generating sets apply to the members of a set rather than to everything at once. For example, we cannot (consistently) say "for every collection of things, there is a set of those things", since this would generate the Russell set of non-self-members, which is a member of itself iff it isn't. But we can say "for every set, for every collection of members of that set, there is a set of those members". This doesn't generate any paradoxical sets, or so we hope: we saw that Kripke [2011c] has a puzzle for us even when we adopt this restriction. We will ignore that puzzle here, though.

Zalta and Parsons tried to avoid paradoxes in a different way: they have a single comprehension schema which applies to everything at once, including the objects being generated, to generate their universe of objects, but there are restrictions on the schema which rule out paradoxical objects like the thing that encodes loving everything that doesn't encode loving itself. There is probably room for debate over just how deep the difference is between the two strategies, but my proposal will be at least superficially more similar to the standard strategy in set theory. We have a way of making new fictional characters from the things we have at one level of the hierarchy, and then these new characters are included in the things generating the objects at higher levels of the hierarchy.

The outline of the theory is as follows. We have a notion of our *possibility space*, which is the set of possible worlds and the set of things that can exist at those worlds. These worlds and things give us a set of intensions, which are functions from the worlds to sets of n-tuples of the things. These intensions stand in as properties and relations. We can then generate a *story*, which is a distribution of these properties and relations over some things. These things can be new things *native* to the story, or they can be *immigrants* to the story. Immigrants can either be non-fictional things, from our possibility space, or fictional things from other stories generated at an earlier level than the story they are immigrants to. The natives to a story are new characters, which can appear as immigrants in new stories further up the hierarchy.

We can follow Zalta in having characters in a story encode the properties and relations they have in the story, but encoding can be thought of as much more specific to this theory of fiction. Zalta takes encoding to be a kind of instantiation, but there is less pressure on us to say that, because we have a separate analysis of statements like 'Poirot is a detective' given in §4.4, where 'is' is the 'is' of ordinary instantiation. We can extend encoding to cover relations as well as properties, and have it relative to a story. For example, 'admires' denotes the intension of admiring over our home possibility space: the function from worlds *w* to ordered pairs <x, y> such that x admires y at *w*. Watson admires Holmes in Doyle's stories, so we can say *Doyle's stories*[(*Watson*, *Holmes*)*Admires*]. If I write a story about them in which Watson doesn't admire Holmes, we can say ¬*My story*[(*Watson*, *Holmes*)*Admires*]. If we like, we could say that an object or some objects encode a property or relation *simpliciter* iff they encode it relative to the story to which they are native.

Now, we must also have a way of letting characters like Christopher Boone admire Holmes, in a different way from how Watson does. This is because Boone admires Holmes qua fictional character, while Watson admires him qua concrete character. Admiring Holmes as Boone does is the sort of property that non-fictional characters can have, so now we treat it as a monadic property, which will be a function from worlds to sets of things that admire Holmes at those worlds. We don't have to take a stand now on whether admiring something concrete (like Watson does) or something fictional (like Boone does) are the same kind of admiring, although we will need to address something similar shortly when we talk about attitudes towards fictional fictional characters like Gonzago. Now we are just trying to generate enough fictional characters without running into paradoxes.

A complication arises from fictional properties. The problem is that it isn't just fictional characters that can migrate into other stories; fictional properties can as well. Having the characters was relatively easy, but having the properties is a bit harder. There are three kinds of fictional properties we need to deal with.

First, properties such as *being an ewok*. Ewoks are a fictional species in the *Star Wars* universe, and while there might be a case for saying that unicorns are part of some kind of public domain myth such that they are native to a composite of all the stories about them (and so not immigrant to any story), you definitely can't make this case about ewoks. Ewoks are native to *Star Wars* but could appear as immigrants. Ewokhood has no non-null intension over our possibility space, assuming Kripke is right about unicorns, and ideally we would be able to generate it from a possibility space associated with the *Star Wars* universe.

Second, properties such as *living with Holmes*. This also has no non-null intension, but the theory as previously sketched can deal with it. Since this is a property which things can only have if Holmes appears in their story, we can always give things this property by putting Holmes in the story, and having things encode *living with* to him relative to that story. *Living with* has an intension generated by our possibility space.

Third, properties such as *pitying Gonzago*. Gonzago is (let us assume) a fictional character in the play within a play in *Hamlet*. Maybe I can pity Gonzago, but this is arguably not the same as what Hamlet would do if he pitied Gonzago, because when I do it I am interacting with *Hamlet*, whereas when Hamlet does it he is not. Pitying a fictional fictional man is thus arguably not the same as pitying a fictional man. So it seems that while pitying Gonzago qua fictional fictional man corresponds to an intension over our possibility space, pitying him qua fictional man corresponds to an intension over a space associated with *Hamlet*.

The natural solution here is to let stories have possibility spaces associated with them, rather than just worlds. Perhaps we should worry that stories are not specific enough about what is possible according to them, but this is just an instance of a feature already present in the basic case. Stories aren't maximally specific, but just as plenty is implicit about what is actual, e.g. Hamlet has ten toes, plenty is implicit about what is possible, e.g. Hamlet could have had only nine toes. Now, having only nine toes is a non-fictional property which appears in *Hamlet* as an immigrant, but pitying Gonzago qua fictional character appears in *Hamlet* as a native, and corresponds to an intension over the possibility space of the *Hamlet* story. Likewise, being an ewok corresponds to an intension over the possibility space of the *Star Wars* universe.

We ought to say that a story's possibility space does not have just one designated actual world, but a set of actual worlds, and say that what is true according to a story is what is true at all the actual worlds. Multiple actual worlds are used in a different context by Williams [2008], to deal with metaphysical indeterminacy. Here they allow us to reconcile the fact that while the possibilities according to a story (unless it is some kind of logical fantasy) will presumably be complete, what is true according to a story will tend to be incomplete. The simplest thing would be to say that what is possible according to a story is whatever is true at some world in the space, and what is true according to a story is whatever is true at all the actual worlds. We should not say quite this though, because fictions can sometimes be underspecified with respect to necessary truths. Returning to Dummett's example about unicorns mentioned in §4.221, the unicorn myth might be non-specific about whether they are in the order Artiodactyla, like deer, or the order Perissodactyla, like horses. But whichever order they are in, they are necessarily in that order, so just as it is not true according to the myth that they are Artiodactylae, it is not true according to the myth that they could be Artiodactylae. We solve this by having a reflexive accessibility relation on worlds in a story's possibility space, and say (as is standard) that something is possible at a world iff true at some accessible world, and so it is possible according to the story if every actual world accesses at least one world where it is true. All the Artiodactyla worlds could access each other, but they would not access the Perissodactyla worlds. These relations could still be equivalence relations if possibility obeyed S5 according to the story, but if the actual worlds were from more than one equivalence class then the story would not fix all the necessary truths.

Now that we have clarified how the multiple actualities work, we can show that there is no violation of actuality entailing possibility here; whenever something is true according to a story it will be possible according to it. This can easily be seen since actuality entailing possibility is equivalent to necessity entailing actuality, and whatever is true at all the worlds accessed by an actual world of a story will clearly be true at all the actual worlds of the story, since accessibility is reflexive. We have to distinguish between what is [not] [true according to *S*] and what is [not true] [according to *S*], but this distinction is standard.

That is basically the whole thing. We distinguish between objects appearing or being involved in properties qua concrete, fictional, fictional fictional and so on. Each character and property is native to a possibility space, and we can generate new stories and their possibility spaces from objects and properties native to a space earlier in the hierarchy. In §4.34 we will explain how the theory meets the desiderata we laid out, and in §4.35 we will mention some limitations and areas for further work. First, however, we will clarify the theory by presenting it in a more formal way.

4.33 Formal presentation of the theory

Since we are simplifying, we can follow Zalta and Williamson in using a constant domain modal logic, and worlds and possibilia of the intended model will serve as the base for the rest of the system. So our home possibility space contains a set W of worlds and a set D possibilia, and the expressions of our actual language pick out members of D (for names) and functions from W to n-tuples of D (for n-place predicates). Now we can give the general procedure for generating objects. Objects aren't in general generated directly from a possibility space. Objects come from *stories*, and stories are generated by taking some immigrant objects and properties and adding in some native objects. At the first level all the immigrant objects and properties will be native to our home possibility space, but at other levels they can come from different places. So we need a way of generating properties and relations from a possibility space, and a way of generating stories and their native objects from a set of objects and a set of properties and relations.

We begin with our home possibility space P, which is the pair $<W, D>$ of possible worlds and possible objects. Let R_{P} be the set of functions F_{n} from W to sets of finite n-tuples of D. In general, R_{S} is the set of relations generated by possibility space S, including properties as 1-place relations.

Now we can generate a set of stories from the home possibility space P. A set of stories is generated from some objects and relations in S. A *migration M* from P is any pair $<d_P, r_P>$, where $d_P \subseteq D_P$ and $r_P \subseteq R_{\mathrm{P}}$. A story σ generated from M will be a quintuple $<W_\sigma, A_\sigma, D_\sigma, V_\sigma, @_\sigma>$. These are defined as follows:

- W_σ is the set of possible worlds for σ. We impose a limit on the size of any W_σ; following Lewis [1986: 118] this could be Beth-2 (the power of the continuum).
- A_σ is a reflexive accessibility relation on these worlds.
- D_σ is $d_M \cup \delta_\sigma$, where δ_σ is the set of objects native to σ. We impose a limit on the size of δ_σ.
- V_σ assigns each member F_n of r_M a function from W_σ to n-tuples of D_σ. This says what worlds in σ's possibility space are like.
- $@_\sigma$ is a subset of W_σ which are the actual worlds of σ.

The stories' worlds and natives and are the non-set-theoretic commitments of the theory. (If you have set theory and the natives and worlds, you get the stories themselves for free.) The selectionist and creationist versions of the theories commit to different stories, and I remain neutral between them. For a creationist, each story must correspond to an actual authorship event, and indiscernible authorship events can give rise to more than one indiscernible story, with the same immigrants, differing only in their set of natives. For the selectionist, each migration generates a plenitude of stories, with one of each kind. Indiscernible authorship events would select the same story, instead of creating similar ones. Since there is a limit on the size of members of δ_σ and W_σ, there is a limit on the number of possible equivalence classes of indiscernible quintuples of this kind that could be constructed out of a single migration. The selectionist will have once story from each available class.

A story σ brings new objects, δ_σ. It also brings new relations. R_σ is the set of relations native to σ, which is the set of functions from W_σ to n-tuples from D_σ, excluding those in the range of V_σ, which were immigrant properties. Now we need a way of generating new stories from migrations including objects and properties not native to the home possibility space P. For this we introduce the concept of a *universe*, which is a pair <O, R> of objects and relations. Universes are characterized by these clauses:

- $<D, R_P>$ is a universe.
- If <O, R> is a universe, then any pair <*d, r*> is a migration from <O, R>, where $d \subseteq O$ and $r \subseteq R$.
- If M is a migration from <O, R>, Ω_M is a set of objects native to stories generated from M, and R_M is a set of relations native to stories generated from M, then $<O \cup \Omega_M, R \cup R_M>$ is a universe.
- If M_0 is <O, R>, Ω_{Mi} and R_{Mi} are sets of objects and relations native to stories generated from the M_i as above, and each M_{i+1} for $i \geq 1$ is a migration from $<\bigcup \Omega_{M0\ldots}\ \Omega_{Mi}, \bigcup R_{M0\ldots}\ R_{Mi}>$, then $<\bigcup \Omega_{Mi}, \bigcup R_{Mi}>$ is a universe.

These universes generate more migrations, which in turn can generate more stories. Note that a migration will in general be a migration from many universes. For the selectionist each migration generates a plenitude of stories, while for the creationist they generate stories corresponding to all and only the actual authorship events. That is about as far as we'll take the formal presentation of the theory. We will see in §4.34 how the theory solves the problems laid out in §4.2, and in §4.35 we will see some of its limitations and areas for further development.

4.34 How the theory solves the problems

§4.2 raised some issues which any ontology of fictional objects needs to be able to deal with. It either needs to have a line on these issues, or leave room for a more developed version of the theory to have one. The theory just sketched was expressly designed to address the issues, and in some places it may be reasonably obvious how it would deal with them. Nonetheless, rather than leave them as exercises for the reader, I'll go over the issues and say how the theory deals with them.

- *4.341 Creations or discoveries?*

I noted that some people think fictional objects are created by their authors, and some people think they are out there waiting to be discovered. I don't have a strong line on this, but the theory sketched out fits nicely with either view. If fictional objects are out there irrespective of our practices, all the objects in the hierarchy defined by the theory will exist. If not, only some will exist.

A nice feature of the hierarchical system is that a certain amount of independence between the objects in different stories is guaranteed. Recall that fictional objects are generated from a migration consisting of objects and properties, which can either be native to our possibility space or to other stories. Everything native to our possibility space is guaranteed to exist, as are the immigrants from other stories, since you can't import Holmes into a story unless the Holmes stories have been written. So whatever a given fictional character depends on, if that character existed then everything they depend on would have also had to exist. Looked at another way, even if some parts of the hierarchy are missing, what remains will still be a properly grounded hierarchy. This can be compared with an ontology of impure sets: even though Socrates and {Socrates} exist contingently, we don't have to worry about {{Socrates}, Plato} existing without {Socrates} existing.

It is probably also worth pointing out that the present theory allows for rather a lot of fictional objects, and if we take them to be created rather than discovered then our ontology won't be so quantitatively bloated. If we want to allow for distinct indiscernible fictional objects, we can do just that: whereas the selectionist asserts that each migration generates one of each possible type of story and its natives, the creationist can assert that each migration generates as many of each type as there are corresponding literary practices, and each story has its own natives.

- *4.342 Kripke and unicorns*

The straightforward response this theory gives to the unicorns problem is the same as that given by Reimer and Kripke: Holmes and unicorns couldn't have existed, because in the relevant context 'Holmes' and 'unicorn' don't pick out an object or a species. Holmes and the property *being a unicorn* do exist qua fictional object and property (which is just to say they exist and are a fictional object and a fictional property), but when we say 'Holmes could/couldn't have existed', we are using 'Holmes' qua name of a person, not a fictional object. Used this way it picks out nobody, and so on the negative free logic proposed in chapter one, 'Holmes could have existed' is false, its negation is true, and likewise *mutatis mutandis* for unicorns.

Since we take 'Holmes' and 'unicorn' not to pick out elements of the fictional ontology, we need a different account of their semantics. We could just say they don't refer, but we won't quite say that. I will treat them similarly to the proposal of Lewis [1978], which I will elaborate in §4.43. This will give the fictional names clearly different meanings, which lays the ground for the role in attitude contexts I will give them in §4.432. In 'Holmes could have existed', however, they will still end up not denoting anything.

- *4.343 Double lives*

The present proposal probably distinguishes itself most sharply from those of Parsons and Zalta in its treatment of the double lives problem. Zalta has both Watson and Christopher Boone encode admiring Holmes. Parsons has both of them literally admiring Holmes. My theory treats the cases quite differently: Watson and Holmes are generated together in Doyle's stories, and are picked out by their roles, part of which involves one encoding admiring the other relative to the story at the actual worlds of the story. Boone admires Holmes the way people in our world do: relative to his story and its actual worlds, he encodes the function from our possible worlds to things that admire Holmes (in whatever sense real people do admire the fictional Holmes) at those worlds. That was the problem of abstract and concrete natives and immigrants. Now we can look briefly at the other parts of the problem raised in §4.23

- *4.3431 Caulfield's fame*

This was just the problem that objects can have incompatible properties in reality and in their fictions. Every theory has a way of dealing with it, whether by distinguishing encoding from instantiation, nuclear properties from extranuclear, or something else. The present proposal uses the encoding/instantiation distinction. Caulfield instantiates being famous, but relative to his story he does not encode being famous at its actual worlds.

- *4.3432 Kripke's story about Sherlock Holmes*

The problem here was meant to be that if fictional characters are abstract objects like numbers, then stories should have the same universes of fictional objects as we do, and so an object might appear in a story both as fictional and as concrete. In Kripke's thought experiment someone writes about someone just like Holmes and there is someone just like Holmes, so Holmes seems to appear both ways in the story. The way we have set things up, this should not be able to happen. The fictional objects generated from the possibility space of a story are different from those generated by our possibility space, so the fictions within Kripke's story would contain different characters from the one Kripke's story contains. If you wanted Holmes to be imported into a story as a fictional character created either by a fictional author or by an (also imported) Doyle, you would do this by importing Holmes-involving properties from our possibility space, as with Christopher Boone's admiring Holmes.

○ *4.3433 Interdependent fictional characters*

The treatment of fictional characters' interdependence is very explicit in the theory, because a story generates all its characters together. We do not need to worry about what explains why either Holmes or Watson exist, because the migration generating the story explains why they both exist. Perhaps we can deal with the problem of Ophelia and Desdemona watching plays with each other in relatively smoothly: properties like *admiring Ophelia* and *admiring Desdemona* (qua fictional) are native to our world and so nothing stops Ophelia admiring Desdemona qua fictional and vice versa. We would not do so well if natives of two stories appeared as concrete immigrants in each other's stories. We can treat at least one as a similar but distinct character, or combine the two stories into a single one (as is usual in interpreting comic book universes). It isn't ideal, but it has at least one companion in guilt: the same piece of ugliness arises in the view of properties as (well-founded) set of their instances, when two properties appear to instantiate each other.

○ *4.3434 Paradoxes*

This ontology of fictional objects ought to be able to avoid paradoxes at least to the extent that the set theoretic universes it is modelled on can. It does this while still allowing fictional characters to depend on other fictions, and to migrate into other fictions either as concrete or as fictional. It avoids the modesty paradox, the paradoxes Fine raises for Parsons, and the standard set-theoretic paradoxes. Whether it avoids all paradoxes without curtailing its expressive capacity too much is an open question, as it is with set theory, but avoiding existing paradoxes gives reason for optimism.

4.35 Limitations of the theory

I already mentioned one limitation, in dealing with characters migrating to each other's stories. This is the price of a properly grounded hierarchy, but an unavoidable price is still a price, and it shouldn't be swept under the carpet. While in fact such cases seem to be mostly dealt with by consumers of literature (including comic books) through the 'one big story' approach, our literary practices could have been different and made the problem bigger. I won't go into it further; I'll just flag it up as a limitation.

Another limitation is that it bases the possibility spaces on a constant domain modal logic. The orthodoxy, *pace* Williamson [2010], is that it is contingent what exists, and in particular, there could have been things that do not actually exist. Formally, to make the scope clear: $\Diamond\exists x\neg @\exists y[y=x]$. This probably means that the intensions we took over our possibility space will also not exist, since they are functions from worlds (presumably ersatz worlds, which many people are happy to accept) to sets of possibilia (which are considerably less popular). Zalta's theory is in the same boat, since he takes himself to be committed to the possibilia he quantifies over even if he takes them as Meinongian non-existents. I won't try to solve the problem now.

Another problem arising from the use of intensions as properties is that we cannot individuate properties hyperintensionally. Any two necessarily co-extensive properties will be treated as the same property when they appear in fictions. We can mitigate this to some extent by widening the class of worlds in our possibility space to include some impossible worlds. Once we have done that, properties may be individuated finely enough that it is plausible that they do not come apart in fiction. If this solution is not deemed satisfactory, maybe we will have to resort to some kind of *sui generis* property ontology and use those instead of intensions. It is possible that a solution to this problem would also provide a solution to the previous one, since it gives intensions over non-actual possibilia less work to do.

The last limitation to mention is that the theory probably does not do very well with what Fine [1982] calls logical and philosophical fantasies, particularly revenge fantasies cast in the terms of the theory itself. These are worlds where logic is different, or something like the metaphysics of properties or fictional characters is different. Maybe we want a fiction where Bob loves Joe but Joe isn't loved by Bob. It is possible that any theory could run into examples like this which it can't really deal with, and you can either provide an unsatisfactory solution or add an epicycle to theory. The epicycles in any presentation will have to stop somewhere though, and I won't go any further with mine. Hopefully enough has been done to show that the proposal deals well with the normal range of cases, and leaves room for epicycles to expand the range if desired.

4.4 Pretence

So far, this chapter has been about the place of fictional characters in our ontology. I argued that there is a role for them, at least as theoretical entities of literary criticism. Fictional characters are artworks, like novels or musical works. I surveyed different kinds of fictional ontology people have put forward, looked at some issues any ontology of fictional characters must address, and put forward a systematic ontology of my own. The creationist version of the theory is less ontologically profligate than the selectionist version, but both commit to a lot of things.

There is another point of view which is much less ontologically committing. That is pretence theory. The *locus classicus* of pretence theory is Walton [1990], which develops and sets out ideas going back to his [1978a] and [1978b]. The central idea of pretence theory is that fiction is pretence. Writers of fiction pretend to recount facts, and readers join in with this pretence, but it is all just so many make-believe games, and none of these things exist in any sense.

- ### 4.41 Leaving your sports car in the garage

Having argued for committing to fictional characters, you might think that pretence theory has nothing to offer us. I still think it has. The arguments of §4.1 said that we should understand sentences like these as being about fictional characters:

> There are characters in some nineteenth century novels who are presented with a greater wealth of physical detail than is any character in any eighteenth century novel.
>
> Every female character in any eighteenth century novel is such that there is some character in some nineteenth century novel

> who is presented with a greater wealth of physical detail than she is. [van Inwagen 1977: 302-3]

These sentences are from discourse about literature. They sound like straightforward assertions, and wouldn't do well as paradigm cases for the pretence theorist. But sentences like this sound much more like pretence:

> It was a dark and stormy night; the rain fell in torrents — except at occasional intervals, when it was checked by a violent gust of wind which swept up the streets (for it is in London that our scene lies), rattling along the housetops, and fiercely agitating the scanty flame of the lamps that struggled against the darkness. [Opening of Bulwer-Lytton 1830]

When Edward Bulwer-Lytton writes that and we as readers go along with it in whatever way readers of fiction do, we don't have to commit ontologically to anything, because we don't think it's true. If all discourse related to fiction was like that, van Inwagen's argument would not work, because we only commit to the commitments of the things we believe, not the things we pretend to believe. Maybe we in fact need to commit to things to understand the things we pretend to believe, but that would involve a separate argument.

Kripke [2013] has been influential in supporting fictional ontologies for sentences like van Inwagen's, but he has also been influential in supporting pretence theory for sentences like Bulwer-Lytton's, while others like Nathan Salmon have wanted to make fuller use of the ontology they worked so hard to persuade us of. Kripke says this:

> In various writings [Salmon 1987, 1998, 2000], he has argued that I ought to have made greater and more effective use of the ontology of fictional characters I propose. Instead of saying that

> Conan Doyle only *pretends* to name any one entity, why not say that he *does* name one entity – the fictional character? [Kripke 2013: xii; his emphasis]

And here is an example of Salmon making the point:

> Once fictional characters have been countenanced as real entities, why hold onto an alleged use of their names that fails to refer to them? It is like buying a luxurious Italian sports car only to keep it garaged. [Salmon 1998: 298]

Salmon goes on to say how he thinks we should view the matter:

> The matter should be viewed instead as follows: Conan Doyle one fine day set about to tell a story. In the process he created a fictional character as the protagonist, and other fictional characters as well, each playing a certain role in the story. These characters, like the story itself, are man-made abstract artifacts, born of Conan Doyle's fertile imagination. The name 'Sherlock Holmes' was originally coined by Conan Doyle in writing the story (and subsequently understood by readers reading the Holmes stories) as the fictional name for the protagonist. That thing – in fact merely an abstract artefact – is according to the story, a man by the name of 'Sherlock Holmes'. In telling the story, Conan Doyle pretends to use the name to refer to its fictional referent (and to use 'Scotland Yard' to refer to Scotland Yard) – or rather, he pretends to be Dr. Watson using 'Sherlock Holmes', much like an actor portraying Dr. Watson on stage. But he does not really so use the name; 'Sherlock Holmes' so far does not really have any such use, or even any related use (ignoring unrelated uses it coincidentally might have had). At a later stage, use of the name is imported from the fiction into reality, to name the very same thing that it is the name of according to the story.

> That thing – now the real as well as the fictional bearer of the name – is according to the story a human being who is a brilliant detective, and in reality an artifactual abstract entity created by Conan Doyle. [Salmon 1998: 300]

It is possible that some of this disagreement is verbal: a disagreement over whether uttering a name as part of a pretended utterance is a use of the name or just a pretended use. But Salmon does say that the very same object is the thing literary critics talk about and the thing which is, according to the stories, a detective. It isn't far from this to say that Doyle is pretending that this thing is a detective, rather than just pretending to use the name to refer to a detective. If that is Salmon's view, then I don't agree with it. I don't have any catastrophic objections to it, but I will present some minor ones, and it's worth showing how you can dissociate yourself from it without being a full-on pretence theorist, and how a version of pretence theory can sit with the ontology of fictional characters I put forward in §4.3.

4.42 Anchored and unanchored pretence

We can distinguish between two kinds of pretence. An *anchored* pretence is when you take something real and pretend that it is a certain way. Walking through the woods, I might pretend I was a bear. According to my pretence, I am a bear. There's a thing, me, such that according to my pretence it is a bear. An *unanchored* pretence is when the things in your pretence are not real things. Walking through the woods, I might pretend there was a bear following me. According to my pretence, there is a bear. But there isn't a thing such that according to my pretence it is a bear. That's a natural way to see things, anyway, and if we can't make this distinction somehow then we're doing things wrong.

The distinction fits nicely into the ontology of fictional characters laid out in §4.3, and corresponds to the distinction between natives and immigrants to a story. When a pretence is anchored, the real objects are imported into the story, and when it is unanchored, the unanchored roles generate new objects native to the story. Having admitted that the unanchored roles generate new objects, however, we could then say that the pretence was anchored on these objects. We can still make the distinction between two kinds of pretence, but instead all pretences are anchored, while only some are *creative*. That's the position Salmon may have been pushing, and it's the position I want to resist.

As I've said, I don't have a catastrophic objection. The most important reason to provide an alternative is to show that you can accept a fictional ontology and still take creative pretences to be unanchored. Even if you're sure that Doyle and his readers aren't pretending that an abstract object is a detective, you can still accept van Inwagen's argument that Doyle created Holmes, and so there's a thing called Holmes that Doyle created.

So, one minor objection to the view is just that it's unintuitive. Some people will be quite sure that creative pretences are unanchored and demand an alternative. But the other minor objections come from when the names occur outside of pretences, but with the same meanings as they have within pretences. The classic example is 'Holmes doesn't exist'. If Holmes is an abstract artefact, then he (it?) does exist. I mentioned this earlier in §4.11 as a problem for van Inwagen and put it to one side, but now we can try for a solution: 'Holmes' can refer to an abstract artefact (or whatever fictional characters are), but it also has another use, coined by Doyle for use in his stories and associated pretences, which also appears in the true 'Holmes does not exist'. I'll give a semantics verifying this in §4.431.

Negative existentials are not the only times the pretend use of fictional names can occur outside of pretences. There are other statements of the same tone as negative existentials: ‘Batman doesn’t live in our city’, ‘Holmes is not real’, ‘Poirot isn’t a famous Belgian’. There are false positive existentials: ‘Holmes exists’, ‘Achilles existed’, and non-existential statements in the same category. There are also attitude ascriptions, where the attitude is itself part of the pretence: ‘Smith admires Holmes’ or ‘Charles fears the slime’ [Charles is from Walton 1978a]. At a pinch we could probably manage to treat the attitude ascriptions as referring to the abstracta, but once we have the pretend use it treats the attitude ascriptions in a more satisfying way. We will give an account of this in §4.432 The problem with treating the other uses as referring to the abstract artefacts is that you either get the wrong truth values as with ‘Holmes doesn’t exist’, or the right ones for the wrong reasons, as with ‘Poirot isn’t a famous Belgian’.

If the pretend use only occurred in pretences and the non-pretend use only occurred in serious assertions, then we could presumably gather together what we needed from the two kinds of use into one word, giving us a theory that worked and a slight feeling of unease. But in fact the pretend use does occur outside of pretences, so putting the two kinds of use into the same word would mean either understanding those non-literally, or in a way that severs the link between truth and assertability. This seldom makes for a satisfying final theory. It would be better to take unanchored pretence at face value as unanchored if we can, and consequently take the pretend use as not referring to anything that exists, including when it occurs outside of pretence. This means we can’t just use the simple semantics for the names which has them refer to the abstracta. We need another one.

4.43 Lewis on truth in fiction

David Lewis [1978] gives an analysis of fictional discourse which does not have the names refer to anything at the actual world. I am going to modify Lewis's account to give a semantics for the pretend use of fictional names which also explains their behaviour outside of pretences. Lewis is doing a few things with his account though, and it is partnered with his notoriously implausible genuine modal realism, so we should break it down a bit so as not to make commitments we don't have to.

One thing he wants to do is give an explanation of why we can accept the premises of the following argument and reject its conclusion:

> Holmes lives at 221B Baker Street.
> 221B Baker Street is a bank.
> __________________________________
> Holmes lives in a bank.

We assert the first premise as part of the pretence, and assert the second when talking about the real world, but we wouldn't ever assert the conclusion. Lewis explains this by saying that the first premise is implicitly prefixed by a fictionality operator: 'In the Holmes stories…'. The second premise is not prefixed by the operator, so the argument is invalid, and the conclusion can be false, prefixed and unprefixed.

Another thing Lewis wants to do is give an account of what determines what is true in a story. Some things are true in a story although not explicitly stated, and some things are left open by stories. Lewis deals with this by having what the story explicitly says pick out some but not all of the worlds in which the explicit statements are true. I won't be arguing with Lewis over the particular class of worlds a story picks out;

we can just follow him in saying that some class is picked out. For Lewis, the natures of his concrete possible worlds can do some work here, by fixing what similar possibilities are like independently of what we think or know about them. If we are not genuine modal realists, we will need some other way of fixing the class of worlds, perhaps using whatever we do think fixes facts about possibility and counterfactuals.

A third thing he wants to do, related to the other two, is give a semantics for the fictionality operators. He takes it that a statement is true in the story iff it is true at all the worlds picked out by the story. However, since the Holmes role is played by different people at different worlds, and even by people who are not counterparts of each other on an ordinary counterpart relation (perhaps they have different origins), 'Holmes' has to refer to different people at different worlds for 'Holmes is clever' to be true at all the worlds. 'Holmes' refers at *w* to whoever plays the Holmes role at *w*, which makes it a *non-rigid designator*.

This account suffers from a technical problem, however, dealing with sentences like *In the Iliad, Hector could have survived the war.* This is true: Hector died but he could have survived. The problem is finding a world to make this true. We want a world where the person playing the Hector role survives, but this is problematic because the Hector role includes dying in the war. Mark Sainsbury [2010: 90] takes this problem to be unsolved. Let's try to solve it.

- *4.431 Semi-rigid designation*

First I should point out that we know full well what we want a model satisfying *In the Iliad, Hector could have survived the war* to look like. The problem is getting a way of interpreting sentences systematically so that those models satisfy that sentence. The model should be one in which for all Iliad worlds *w*, there is a world *v* possible relative to *w* such that the person playing the Hector role at *w* survives the war at *v*. The difficulty is that since at *v* the person survives the war, they don't play the Hector role at *v*, and so *Hector* doesn't refer to them at *v*, and *Hector survived the war* isn't true at *v.*

The solution is to evaluate sentences relative to pairs of worlds, one for fixing the referents of the names relative to the pair, and one for seeing if objects satisfy predicates relative to the pair. Evaluating sentences relative to pairs of worlds is not a new idea; it goes back at least to Davies and Humberstone [1980] as a device for accommodating an actuality operator in modal logic, and has developed into what is now known as *two-dimensional semantics*. For an overview of this area, see Schroeter [2012] and Chalmers [forthcoming]. For present purposes, a fictionality operator like *In the Iliad* should shift both worlds, while a possibility operator like *could have* shifts only the world determining which objects satisfy which predicates. So we have these clauses:

- V(*Hector*) at <u, v> is the thing playing the Hector role at u.
- V(*Survived the war*) at all worlds is the function f from worlds to sets such that f(w) is the set of things that survive the war at w.
- Where F is a predicate and n is a name, V(Fn) is true at <u, v> iff V(n) at <u, v> is a member of V(F)(v).
- V(*In the Iliad,* φ) is true at <u, v> iff for all Iliad worlds w (relative to v), V(φ) is true at <w, w>.
- V(◊φ) is true at <u, v> iff for all worlds w accessible from v, φ is true at <u, w>
- V(φ) is true in a model M iff V(φ) is true at <@, @> where @ is the actual world of M.

It may be helpful to think of the first world of a pair as determining what the names refer to, and the second world determining what goes on at the pair. The work gets done by having fictional names like *Hector* behave rigidly with respect to possibility operators and non-rigidly with respect to fictionality operators. Working through the example, we paraphrase *In the Iliad, Hector could have survived the war* as *In the Iliad, ◊Survived the war*(*Hector*).

- V[*In the Iliad, ◊Survived the war*(*Hector*)] in M is
- V[*In the Iliad, ◊Survived the war*(*Hector*)] at <@, @>, which is true iff
- V[◊*Survived the war*(*Hector*)] is true at <w, w> for all Iliad worlds w (relative to @), which is true iff
- For all Iliad worlds w (relative to @), there is a world v accessible from w, such that V[*Survived the war*(*Hector*)] is true at <w, v>, which is true iff
- For all Iliad worlds w (relative to @), there is a world v accessible from w, such that V(*Hector*) at <w, v> is a member of V(*Survived the war*) at <w, v>, which is true iff
- For all Iliad worlds w (relative to @), there is a world v accessible from w, such that the person playing the Hector role at w survived the war at v.

That is the result we wanted. Note that the framework above can still accommodate rigid designators, which refer to the same thing at all worlds, and also non-rigidly designating definite descriptions which refer relative to <u, v> to the thing satisfying the description at v (not u). These will have their references shifted by possibility operators as well as fictionality operators, as they should.

We should also note that outside the context of a fictionality operator, fictional names will tend not to refer. That looks like a good result, and allows *Hector* to be univocal in *Hector does not exist* and *In the Iliad, Achilles killed Hector*. But we can run into some difficulties where a story is coincidentally played out. We wanted to side with Kripke and philosophical orthodoxy in saying that *Hector does not exist* is true as long as the Iliad was made up, even if coincidentally events like it all happened. We can get round this by saying that non-actuality is part of the Hector role. We don't have to worry about sentences like *Hector could have been Obama's brother* being true either. Even if Obama could have had a brother who played the Hector role, *Hector could have been Obama's brother* is only true if *Hector is Obama's brother* is true at <@, w> for some world w accessible from @. *Hector* still doesn't refer to anyone relative to <@, w>, because nobody plays the Hector role at @. The only way a fictional name could start referring is if we introduce the appropriate fictionality operator, to shift the first world of the pair away from @. Outside of fictionality operators, fictional names will always be empty, because at the actual world their history of use goes back to a block, in the sense of Donnellan [1974]. This will allow such uses to be treated according to the negative free logic argued for in chapter one, just as if they were straightforwardly non-referring.

- *4.432 Fictional names in attitude ascriptions*

Now we have a view according to which there are two uses of fictional names. In the pretence use they are semi-rigid designators as described above, and in the literary critical use they are ordinary rigid designators which refer to abstract objects. This should be thought of more as polysemy than as straight-up ambiguity; it is of course no coincidence that the two uses are homophonic. Now we need to think about how they will be incorporated into the account of propositional and objectual attitude ascriptions given in chapter three. The literary critical use will just treat them as names in the ordinary way, but we need to explain how semi-rigid designators work in attitude ascriptions, and we need a way of deciding how to classify particular occurrences of names in attitude ascriptions.

How exactly we integrate semi-rigid designators into the treatment of attitude ascriptions will depend on the details of the proposal for ordinary cases, and the metaphysics of the senses involved was not fully worked out. To see how it might go, however, we can take the treatment of senses as Chalmers' two-dimensional intensions as a case in point. Two dimensional intensions can be seen as functions from epistemically possible scenarios to intensions. Empty names from mistakes and lies, like 'Vulcan', all have the null intension with respect to the actual scenario, and as such will all have the same intension with respect to the actual scenario. This means all their difference in meaning is meta-representational.

Semi-rigid designators have non-meta-representational differences in meaning, but they still have the null intension with respect to the actual scenario in the sense that 'FN could not have existed' is true for any fictional name FN. To deal with this, we can instead think of senses as 3D intensions, or functions from epistemically possible scenarios to the 2D intensions used in the semantics for semi-rigid designators. Mistaken names do not refer with respect to any world pair, since they do not refer in the context of fictionality operators. This means that with respect to the actual scenario they will still have the null function as on world pairs as their 2D intension, and so any differences in meaning will still be meta-representational. However, fictional names can have different 2D intensions with respect to the actual scenario. This gives a way of incorporating senses for semi-rigid designators into an extension of a framework designed to deal with the non-fictional case. Once we have these senses, we can treat them just like any other names. Propositions constructed out of the new senses will encode enough information to assess their truth values with respect to different fictions, so if we want to say something like 'Fred believes that Poirot is a detective, but *that* is only true in Christie's stories', we should be able to manage it. We should also note that since there are sometimes elliptical fictionality operators in attitude ascriptions, so really we a say that we believe that *in the Holmes stories* Holmes is a detective, the propositions will often involve the senses of fictionality operators. This doesn't generate any obvious problems though; the fictionality operators pick out a set of worlds and can be ascribed an intension accordingly.

A full development of the view would demand a fuller development of the ordinary case than we have, but this will suffice for present purposes. Now we need to decide how to classify different uses of fictional names in attitude ascriptions. The two main factors guiding us here were identified in chapter three. They are that the connections between propositional and objectual attitudes should be preserved, and the proposition referred to in describing the attitude should be the content of an utterance expressing that attitude.

Lots of cases will be easy to classify: in 'I believe that Poirot is a detective', 'Poirot' is the pretence use. In 'I believe that Christie created Poirot', it is the literary critical use. The most interesting cases are those ascribing objectual attitudes, and ascribing the beliefs corresponding to those attitudes. Walton [1978a] talks about someone called Charles fearing some slime while watching a horror movie. Suppose he also fears Dracula. Which use is 'Dracula' in 'Charles fears Dracula'? I think the most satisfying thing is to say that it is the pretend use, and follow Walton in saying that the fear is not genuine fear, but part of the pretence. (This explains, as Walton notes, why his fear does not have the motivational consequences of fear, although it does have some of the physiological consequences.) The pretend propositional attitudes which are the reasons for and consequences of his fear would be expressed by Charles using the pretend use, and so we should ascribe them using the pretend use. It is most satisfying to be able to ascribe the objectual attitude that way too. The pretend use could also be used in truly ascribing genuine fear, if the subject thought that Dracula was real, and really feared him, keeping a supply of garlic, stakes and so on. When they expressed the propositional attitudes corresponding to the fear, they would not do so using a tacit fictionality operator. Charles would, because he is just pretending. It will be possible for the literary critical use to feature truly in attitude ascriptions, but these will typically be the kinds of attitudes which people bear to any artistic creations, like musical works or novels, because that is the sort of thing that fictional characters are.

At this stage, you might think that there is a perhaps unwelcome convergence between my treatment of the pretence use of fictional names and Meinongianism. We have names which have different meanings, not just at the meta-representational level, but which don't refer to anything existent. Where 'n' is one of these names, 'n does not exist is true', and now, to analyse intentional attitude ascriptions, I have ontologically committed to the names' senses. You might worry that the superficial similarities point to a deep similarity, and the view is really a notational variant of Meinongianism, with the senses playing the role of non-existent objects.

You could object to this at the deep level, saying that the superficial similarities are just that, and deeper down there are real differences. It is part of the explicit view that the senses exist and denote nothing with respect to <@,@>, which makes them different from the Meinongian's objects in at least that way. If that doesn't satisfy, we can add that if we accept that there are meaningful fictional names which refer to nothing existent, then we are already committed to the senses, Committing to the names being meaningful shouldn't make you a Meinongian by itself. The names do refer to non-actual objects with respect to world-pairs other than <@,@>, but since these worlds are just the ones we already use in our model theory for ordinary modal discourse, my treatment of fiction need be no more Meinongian in its commitments than that. This was one of the advantages of Lewis's original proposal, and is part of why we have followed and developed it.

Finally, we can point to an irreconcilable superficial difference between my proposal and the Meinongian's. We saw in §1.12 that we can identify a scope ambiguity in negative existentials, and the difference between the two readings can be brought out using lambda predicates. If 'n' is an empty name, then 'n does not exist' is true when the negation has wide scope, and false when it has narrow scope. The same goes for the pretence use of fictional names. ¬Exists(Holmes) is true, but [λx.¬Exists(x)](Holmes) is false. The Meinongian can also acknowledge both readings, but they will say that both readings are true, because Holmes is non-existent (narrow scope) and non-existent things do not exist (wide scope). This superficial difference corresponds to the deep difference that the Meinongian thinks that there are non-existent things and I don't, so they hold that atomic predications of non-existence are sometimes true, and I don't. There are still superficial similarities, but there are bound to be since both views are trying to account for the same data. The fundamental difference between Meinongianism and the orthodoxy's robust sense of reality remains, both in the theory and on the surface.

- ### *4.433 Note on 'according to'*

An objection could be raised here that the treatments of 'according to Christie's stories, Poirot is Belgian' and 'according to Leverrier, Vulcan is a planet' are too different[74]. 'Vulcan' is treated as a non-referring name which only contributes meta-representational content, whereas 'Poirot' is treated as a semi-rigid designator. The truth conditions of the 'Poirot' sentence involves whatever metaphysics grounds truths about possibility (model-theoretically cashed out as possible worlds), while the truth conditions of the 'Vulcan' sentence just involve Leverrier

[74] Thanks to my examiners for raising this objection.

standing in a relation to a Fregean GP. They seem similar, so why treat them so differently?

My response to this has two parts. First, the treatments are not quite as different as they might appear, and the treatment of the Vulcan sentence could without contradiction be developed into something very like the treatment of the 'Poirot' sentence. Second, there are some disanalogies between the two kinds of sentence which could justify treating them differently, and would at least demand some more work on the part of someone who wanted to treat them alike.

What we need from the semantics of an operator like 'S believes that', 'according to S' or 'according to Christie's stories' is this: a recipe for getting from a point in a model and the semantic value of a sentence to a set of propositions (or a sense which determines a set of propositions). With 'according to Christie's stories', we do this using the machinery of world-pairs and semi-rigid designation. With 'S believes that' we more or less just look at the propositions which S, or an idealization of S, would be disposed to sincerely assent to. Now, we could try to make these treatments more similar to each other. We could use the THAT operator in the fiction case as well as the belief case, and instead of letting it be a black box, we have a structured entity encoding all the information of the 2D or 3D intensions of the sentence's parts, in a manner similar to Chalmers [2011]. Then the entity in a way parallel to the semi-rigid designation machinery. We could also use similar machinery with 'S believes that', where the operator's semantic value determines a set of pairs or triples of worlds corresponding to S's global belief state, and 'Vulcan' determines a Vulcan-role, presumably corresponding to its A-intension.

This has the makings of a unified treatment of fictionality and belief operators, which does not contradict what I have already said in any very serious way. So why not do it? The main reason is that what is true according to Leverrier should track what (idealized) Leverrier will and won't assent to, whereas there is no analogous set of dispositions for truth in fiction to be beholden to. The opposite is true, Lewis's proposal is designed to fill a gap in our theory of truth in fiction, by letting what is true according to a story depend on objective facts about possibility and counterfactuals. This difference manifests itself in two ways. First, we need something to determine truth in fiction, whereas we already have people's idealized dispositions to assent to determine, or at least indicate, what they believe. Second, Christie's stories determine a Poirot role which matches the set of world-pairs determined by the semantic value of 'according to Christie's stories'. It isn't clear that the A-intension of 'Vulcan' could determine a Vulcan role which matches Leverrier's belief-world-pairs or triples. Different people believe different things about Vulcan, and these beliefs all have a bearing on the public-currency A-intension of 'Vulcan'. Where Leverrier's beliefs about Vulcan diverge from the composite Vulcan role, his idealized dispositions to assent will diverge from what is true according to him, and that's bad. The difference is that what is true according to a person is individualistic, whereas what is true according to a story is public. The A-intension, if words are to mean the same thing in different people's mouths, is presumably public, which makes it difficult to mesh with a person's belief state in the way the semi-rigid designation machinery demands. Perhaps these problems can be solved, but until they are it makes sense to stick with the separate treatments I have offered.

4.5 Conclusion

This chapter was about how to understand fictional names, by which we mean names introduced in the context of fictional works, like ‘Sherlock Holmes’ and ‘Poirot’. We saw that uses of these names divide into two main categories: talking within fiction and talking about fiction. Some uses are harder to classify than others, and that could motivate a unified semantics which applied to all the uses. Nonetheless, I offered two analyses, one for each category, and I suggest we do the best we can to classify difficult cases. The use within fiction is not confined to fiction, since people can be confused and think things are really true when they are actually only true in the stories, or they can just remark on how reality and the stories are different. Then the semantics for the use within fiction will still apply, and what they say will tend to be evaluated pessimistically, at least in non-intentional contexts. The use for talking about fiction will always result in people talking about fiction in a sense, because it refers to fictional characters, and talking about fictional characters is (in a sense) talking about fiction. There may however be scope for people talking about fictional characters without realizing it, although when people are very wrong about what they are talking about this may result in referential failure instead.

§4.1 looked at van Inwagen’s argument that there are fictional characters, and defended it against some objections. His argument does not say anything much about what fictional characters are like though, so we looked at different kinds of fictional ontology. These can be categorized in several different ways, and we looked at some of the more popular options: fictional characters as abstract artefacts, as Platonic abstracta, and as non-existent concreta.

§4.2 presented and explored several issues that fictional ontologies should be able to deal with, such as whether the characters are created or discovered, Kripke's problem about unicorns, and problems arising from characters from one fiction appearing in another, either as real characters or as fictional characters. We saw how different kinds of fictional ontology will have to approach these problems in different ways. This helps us decide which ontology we should adopt. Ultimately I went with a systematic ontology of abstract objects, which can be adopted in either a creationist or a selectionist version, and this positive view was presented in the next section.

§4.3 set out a simple version of this systematic ontology, drawing on elements from work by Zalta, Parsons and Fine, and showed how it should have the means to address the issues discussed in §4.2. We also saw some of this system's limitations and some areas which could be further developed.

This system is best suited for the use of fictional names for talking about fiction, and §4.4 looked for an alternative analysis of the use which primarily occurs within pretence. I proposed a semantics for this use based on David Lewis's system. It had to be modified, however, to deal with Mark Sainsbury's puzzle relating to sentences like 'according to Doyle's stories, Holmes could have been a vicar'. I suggested we treat fictional names as semi-rigid designators, which behave non-rigidly with respect to fictionality operators, and rigidly with respect to possibility operators. This analysis also allowed us to deal with this use of names occurring outside of pretence, where the sentences get pessimistic truth values in non-intentional contexts, and in particular negative existential statements like 'Holmes does not exist' could be taken at face value and evaluated as true. In §4.432 I showed how fictional names would behave in attitude contexts, which was quite similar to how ordinary names behaved in chapter three. Finally in §4.433 I looked at the possibilities for use the treatment of fictional names in fictionality operators to enrich the treatment of mistaken names in belief contexts, but concluded that this was probably unnecessary and might not work very well anyway. The treatment from chapter three would suffice.

Chapter 5 – Concluding Remarks

5.1 What we have learned about empty names

Apparently empty names are used in many different ways and give rise to many different problems. It would probably not have been possible to solve all the problems here, so that the reader would immediately know how to analyse any sentence containing an apparently empty name. (Or so that if they couldn't, it wouldn't be the name's fault.) My aim has been to get our understanding of apparently empty names into a position where we have the means to solve all the problems. This has meant treating apparently empty names in one of three ways.

Names introduced in the contexts of mistakes and lies will tend to be straightforwardly non-referring, which gives rise to pessimistic truth values in non-intentional contexts. In intentional contexts the truth values need not be pessimistic, because the names can have metarepresentational content even if they do not succeed in otherwise representing. Names introduced in the context of fiction are used in one of two ways. They can refer to a fictional character, which exists and is the same sort of thing as a play or a musical work. Alternatively, they might be semi-rigid designators, which don't refer to anything with the respect to the actual world and so give rise to pessimistic truth values in non-intentional contexts, but do refer with respect to other possible worlds in the context of a fictionality operator. The pessimistic truth values in non-intentional contexts allow us to take true negative existential statements like 'Sherlock Holmes does not exist' at face value. The non-trivial 2D intensions also give us more resources for assigning the truth values we want in contexts involving attitude verbs and fictionality operators.

5.2 What we have learned about other things

Hard cases can make bad law, but sometimes getting your theory to account for hard cases can lead to a theory which deals with the easier cases better. Hopefully this thesis has involved a bit of that. Much work has been done by others on problems arising from co-referring names, and they have much in common with the problems of empty names. Both involve a mismatch between how things are in the world and our attempts to represent them.

Chapter two tried to give an account of how a person's beliefs can be consistent or inconsistent, and how their deductions can be valid or invalid, without appealing directly to the logical relations between the contents of their beliefs. This was helpful because of the danger that some beliefs don't have propositional contents, or at least not the kind of contents which can be true. Our account instead appealed to the co-ordination relations between belief tokens, and the consequences of these for what the beliefs could potentially be about, and what their contents could be. We motivated this machinery independently though, because it is just as useful for dealing with Kripke's puzzle about belief. That puzzle suggested that the rational relations between belief tokens don't always track the logical relations between the contents of utterances expressing them, where contents are individuated so as to express useful interpersonal generalizations. Perhaps if we didn't have to deal with empty names we could have tried to get by with just publicly accessible and expressible Fregean contents, but the solution which can handle empty names best is also the one that does best with Kripke's puzzle. Consideration of empty names helps us draw the line between interpersonal and intrapersonal content in the most satisfying place.

Chapter three also had some bearing on puzzles of co-reference. This time the issue was about co-referring words appearing in different attitude ascriptions. We could probably just about make Millianism plausible even where attitude ascriptions were concerned, if we didn't have to deal with empty names, but allowing all empty names to be substitutable *salva veritate* in attitude contexts may push these intuitions to breaking point. Maybe everyone who worships Phosphorus worships Hesperus, but not everyone who worships Zeus worships Thor, Vulcan and Santa. Once empty names have motivated a Fregean semantics for attitude ascriptions, however, we can use it for non-empty names too, in line with the intuitions the Millians had to resist.

The Fregean semantics motivates some metaphysical work as well, because we need to see what kinds of thing could play the role of Fregean senses and Fregean gappy propositions, assuming something could. There is promising work in this area already, which looks at a metarepresentational level of content, whether this is in terms of Cumming's discourse referents, Chalmers' primary intensions, or Gillian Russell's reference determiners.

5.3 What we have not learned about empty names

I said in §5.1 that I didn't have an analysis for every possible sentence containing an apparently empty name, although I hope to have put us on the right track in the search for such analyses. There is still some work to be done. I just mentioned the metaphysical project of giving a full account of what metarepresentational content is like, and until we have one of those we won't know exactly what empty names (or indeed any names) are up to in attitude contexts.

Another thing I haven't solved is how to deal with weird mixed cases, where empty names are used in ways which are hard to classify according to the various analyses I have proposed. There is no limit to the zeugmatic convolutions that could be made by mixing the pretence and literary critical uses of fictional names. Meinongians do well on this, but they do worse elsewhere. In particular, anything we do to unify the pretence and literary critical uses to deal with mixed cases will also end up conflating them, leading to other kinds of ambiguities or demands for awkward paraphrases. Nonetheless, disentangling the odd cases remains a challenge.

Two final areas for further investigation are mythical names and the names used in false scientific theories. These seem to fall somewhere between fictional names and the names used in the contexts of mistakes and lies. It seems plausible that a name could start out as part of a mistake or a lie, but the misconception could live on as fiction after being exposed, with the name being used as a fictional name. That doesn't create any immediate problems for my account, as it can be seen just as a change in meaning. Perhaps there are problems lurking when we deal with intermediate cases, however. Perhaps the machinery we already have can accommodate it, or perhaps we need some more. Along with metarepresentational content and odd mixed cases, it is an area for further work.

Bibliography

- Adams, F. and R. Stecker 1994: ‘Vacuous singular terms’, *Mind and Language* 9(4): 387-401
- Armstrong, D. M. 1989: *A Combinatorial Theory of Possibility* (Cambridge: Cambridge University Press)
- Bacon, A. 2013: ‘Quantificational logic and empty names’, *Philosophers’ Imprint* 13(24): 1-21
- Barcan, R. 1946: ‘A functional calculus of first order based on strict implication’, *Journal of Symbolic Logic* 11(1): 1-16
- Barnes, E. and Williams, J. R. G. 2011: ‘A theory of metaphysical indeterminacy’, in Karen Bennett and Dean Zimmerman (eds.) *Oxford Studies in Metaphysics* 6
- Bench-Capon, M. MSa: ‘Dialetheist disjunctive syllogism’
- Bench-Capon, M. MSb: ‘Tolerance, transparency, transitivity, and tonk’: https://docs.google.com/file/d/0B2zCttQcG8k2anZ1b1pKSjJmVmM/
- Bench-Capon, M. MSc: ‘Belief ascriptions and sential quantification’: https://docs.google.com/file/d/0B2zCttQcG8k2b3FSWVFRd2J6bkE/edit?usp=sharing
- Bennett, K. forthcoming: *Making Things Up* (Oxford University Press)
- Boolos, G. 1984: ‘To be is to be the value of a variable (or the values of some variables)’, *Journal of Philosophy* 81(8): 430-49
- Boolos, G. 1985: ‘Nominalist Platonism’, *Philosophical Review* 94(3): 327-44
- Braun, D. 1993: ‘Empty names’, *Noûs* 27(4): 449-69
- Braun, D. 2005: ‘Empty names, fictional names, mythical names’, *Noûs* 39: 596-631

- Braun, D. and T. Sider 2007: ‘Vague, so untrue’, *Noûs* 41(2): 133-56
- Brontë, E. 1847: *Wuthering Heights* (Thomas Cautley Newby) http://www.gutenberg.org/ebooks/768
- Bulwer-Lytton, E. 1830: *Paul Clifford* (Colburn and Bentley) http://www.gutenberg.org/ebooks/7735
- Cameron, R. 2010: ‘How to have a radically minimal ontology’, *Philosophical Studies* 151(2): 249-64
- Cameron, R.: ‘There are no things that are musical works’, *British Journal of Aesthetics* 48(3): 295-314
- Caplan, B. 2012: ‘Never been kicked’, in T. E. Wartenberg (ed.) *Philosophers on Film: Fight Club* (Routledge): 132-161
- Carnap, R. 1947: *Meaning and Necessity* (University of Chicago Press)
- Carroll, L. 1876: *The Hunting of the Snark* (MacMillan) http://www.gutenberg.org/ebooks/13
- Chalmers, D. 2003: ‘The content and epistemology of phenomenal belief’, in Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Perspectives* (Oxford University Press)
- Chalmers, D. 2011: ‘Propositions and attitude ascriptions: a Fregean account’ *Noûs* 43(4): 595-639
- Chalmers, D. forthcoming: ‘Two-dimensional semantics’, in E. Lepore and B. Smith (eds.) *Oxford Handbook of the Philosophy of Language* (Oxford University Press); http://consc.net/papers/twodim.pdf
- Church, A. 1950: ’On Carnap’s analysis of statements of assertion and belief’, *Analysis* 10(5): 97-9
- Church, A. 1954: ‘Intensional isomorphism and identity of belief’, *Philosophical Studies* 5(5): 65-73
- Correia, F. 2008: ‘Ontological dependence’, *Philosophy Compass* 3(5): 1013-32

- Crimmins, M. and J. Perry 1989: 'The prince and the phone booth: reporting puzzling beliefs', *Journal of Philosophy* 86(12): 685-711
- Cumming, S. 2007: *Proper Nouns* (PhD thesis, Rutgers): http://rucore.libraries.rutgers.edu/rutgers-lib/22845/PDF/1/
- Cumming, S. 2008, 'Variabilism' *Philosophical Review* 117(4): 525-54
- Cumming, S. forthcoming: 'Discourse content', in A. Burgess and B. Sherman (eds.) *Metasemantics* (Oxford University Press)
- Davidson, D. 1965: 'Theories of meaning and learnable languages', Y. Bar-Hillel (ed.) *Logic, Methodology and Philosophy of Science* (Amsterdam: North-Holland): 383-94, reprinted in Davidson [1984: 3-15]
- Davidson, D. 1984: *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press)
- Davidson, D. and G. Harman 1972 (eds.): *Semantics of Natural Language* (Dordrecht: D. Reidel, 1972)
- Davies, M. and I. L. Humberstone 1980: 'Two notions of necessity', *Philosophical Studies* 38(1): 1-30
- Davies, M. 1981: 'Meaning, structure and understanding', *Synthese* 48(1):135-61
- Dickie, I. 2011: 'How proper names refer', *Proceedings of the Aristotelian Society* 111(1pt1): 43-78
- Donnellan, K. S. 1970: 'Proper names and identifying descriptions', *Synthese* 21(3-4): 335-358, reprinted in Davidson and Harman [1972]
- Donnellan, K. S. 1974: 'Speaking of nothing', *The Philosophical Review* 83(1): 3-31
- Dorr, C. 2005: 'What we disagree about when we disagree about ontology', in M. E. Kalderon (ed.) *Fictionalism in Metaphysics* (Oxford University Press): 234-286

- Dummett, M. 1983: 'Existence', in D. P. Chattopadhyaya (ed.) *Humans, Meanings and Existences*, *Jadavpur Studies in Philosophy* 5 (Delhi); reprinted in Dummett [1993a: 277-307]; page references to reprint.
- Dummett, M. 1993a: *The Seas of Language* (New York: Oxford University Press)
- Dummett, M. 1993b: 'Could there be unicorns?' pp.328-48 in Dummett 1993a; revised and translated version of 'Konnte es Einhorner geben?', *Conceptus* 17(40-41): 5-10
- Dummett, M. et al 1974a: 'First general discussion session', *Synthèse* 1974 27(3): 471-508
- Dummett, M. et al 1974b: 'Second general discussion session', *Synthèse* 1974 27(3): 509-21
- Evans, G. 1973: 'The causal theory of names', *Proceedings of the Aristotelian Society*, Supp. 47: 187-208, reprinted in Moore [1993: 208-227]
- Evans, G. 1981: 'Semantic theory and tacit knowledge', in S. Holtman and C. Leich (eds.) *Wittgenstein: to Follow a Rule* (Routledge and Kegan Paul)
- Evans, G. 1982: *The Varieties of Reference* (ed. John McDowell) (Oxford: Clarendon Press)
- Everett, A. 2003: 'Empty names and "gappy" propositions', *Philosophical Studies* 116: 1-36
- Fara, D. G. 2000: 'Shifting sands: an interest-relative theory of vagueness', *Philosophical Topics* 28(1): 45-81. Originally published under the name 'Delia Graff'.
- Fine, K. 1982: 'The problem of non-existents', *Topoi* 1: 97-140
- Fine, K. 1984: 'Critical review of Parsons' *Non Existent Objects*', *Philosophical Studies* 45(1): 95-142
- Fine, K. 2003: 'The role of variables', *Journal of Philosophy* 100(12): 605-631
- Fine, K. 2007: *Semantic Relationism* (Wiley-Blackwell)

- Fodor, J. A. 1975: *The Language of Thought* (New York: Thomas Y. Crowell)
- Fodor, J. A. 2008: *LOT2: The Language of Thought Revisited* (Oxford University Press)
- Forbes, G. 2013: 'Intensional Transitive Verbs', in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), URL = <http://plato.stanford.edu/archives/fall2013/entries/intensional-trans-verbs/>.
- Frege, G. 1952: 'On sense and reference' (tr. Max Black), *Translations from the Philosophical Writings of Gottlob Frege*, eds. Peter Geach and Max Black (Oxford: Blackwell 1952): 56-78. Reprinted in Moore [1993: 23-42]; page references to reprint.
- Frege, G. 1956: 'The thought: A logical inquiry' (tr. A. M. and Marcelle Quinton), *Mind* 65(259): 289-311
- Frege, G. 1980: 'Letter to Jourdain', in his *Philosophical and Mathematical Correspondence*, ed. B. McLaughlin, trans. H. Kaal. Reprinted in Moore [1993]
- Goldstein, L. 2009: 'Pierre and circumspection in belief-formation', *Analysis* 69(4): 653-5
- Grice, H. P. 1969: 'Vacuous names', *Words and Objections* eds. D. Davidson and J. Hintikka (Dordrecht: Reidel): 118-45
- Haddon, M. 2003: *The Curious Incident of the Dog in the Night-time* (Jonathan Cape)
- Hare, C. 2009: *On Myself, and Other, Less Important Subjects* (Princeton University Press)
- Huddleston, R. and G Pullum (eds.) 2002: *The Cambridge Grammar of the English Language* (Cambridge University Press)
- Kaplan, D. 1968: 'Quantifying in', *Synthese* 19(1-2): 178-214
- Kaplan, D. 1973: 'Bob and Carol and Ted and Alice', in Hintikka et al (eds.) *Approaches to Natural Language* (Dordrecht/Boston: D. Reidel): 490-518

- Kaplan, D. 1989: ‘Demonstratives’, in *Themes from Kaplan* eds. J. Almog, J. Perry and H. Wettstein (New York: Oxford University Press)
- Kaplan, D. 1990: ‘Words’, *Aristotelian Society,* Supp. 64: 93-119
- Kaplan, D. 1995: ‘A problem in possible-world semantics’, in W. Sinnott Armstong, D. Raffman and N. Asher (eds.) *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus* (Cambridge University Press): 41-52
- Kleene, S. C. 1952: *An Introduction to Metamathematics* (Amsterdam: North-Holland)
- Kripke, S. A. 1963: ‘Semantical considerations on modal logic’, *Acta Philosophica Fennica* 16: 83-94, reprinted in L. Linsky (ed.) *Reference and Modality* (OUP 1971): 63-72; page references to reprint.
- Kripke, S. A. 1979: ‘A puzzle about belief’, *Meaning and Use* ed. A. Margalit (Dordrecht: Reidel 1979): 239-83
- Kripke, S. A. 1980: *Naming and Necessity* (Blackwell); originally in Davidson and Harman [1972]
- Kripke, S. A. 2008: ‘Frege’s theory of sense and reference: some exegetical notes’, *Theoria* 74(3): 181-218
- Kripke, S. A. 2011a: *Philosophical Troubles: Collected Papers, Volume 1* (New York: Oxford University Press)
- Kripke, S. A. 2011b: ‘Vacuous names and fictional entities’, pp. 52-74 in Kripke 2011a
- Kripke, S. A. 2011c: ‘A paradox about time and thought’, pp. 373-9 in Kripke 2011a
- Kripke, S. A. 2011d: ‘Unrestricted exportation and some morals for the philosophy of language’, pp. 322-350 in Kripke 2011a
- Kripke, S. A. 2013: *Reference and Existence* (Oxford University Press)
- Kroon, F. 1994: ‘Make-believe and fictional reference’, *The Journal of Aesthetics and Art Criticism* 52(2): 207-14

- Larson, R. and G. Segal 1995: *Knowledge of Meaning* (Cambridge, Massachusetts: The MIT Press)
- Larson, R., M. den Dikken and P. Ludlow 1997: 'Intensional transitive verbs and abstract clausal complementation', manuscript at http://semlab5.sbs.sunysb.edu/~rlarson/itv.pdf
- Levinson, J. 1980: 'What a musical work is', *Journal of Philosophy* 77(1): 5-28
- Lewis, D. and S. Lewis 1970: 'Holes', *Australasian Journal of Philosophy* 48(2): 206-12
- Lewis, D. 1973: *Counterfactuals* (Blackwell)
- Lewis, D. 1974: 'Radical interpretation', *Synthese* 23: 331-44
- Lewis, D. 1978: 'Truth in fiction', *American Philosophical Quarterly* 15(1): 37-46
- Lewis, D. 1979a: 'Scorekeeping in a language game' *Journal of Philosophical Logic* 8: 339-59
- Lewis, D. 1979b: 'Attitudes *de dicto* and *de se*', *Philosophical Review* 88: 513-43
- Lewis, D. 1983: 'New work for a theory of universals', *Australasian Journal of Philosophy* 61: 343-77
- Lewis, D. 1984: 'Putnam's paradox', *Australasian Journal of Philosophy* 62(3): 221-36
- Lewis, D. 1986: *On the Plurality of Worlds* (Blackwell)
- Lewis, D. 1990: 'Noneism or allism?', *Mind* 99(393): 23-31
- Lewis, D. 1991: *Parts of Classes* (Oxford: Basil Blackwell)
- Linnebo, Ø. 2012: 'Reference by abstraction', *Proceedings of the Aristotelian Society* 112(1pt1): 45-71
- Linsky, B. and E. Zalta 1995: 'Naturalized Platonism versus Platonized naturalism', *The Journal of Philosophy* 92(10): 525-555
- Linsky, B. and E. Zalta 1996: 'In defense of the contingently non-concrete' *Philosophical Studies* 84(2-3): 283-94

- Linsky, B. and E. Zalta 1994: 'In defense of the simplest quantified modal logic', *Philosophical Perspectives* 8: 431-58
- Lycan, W. 1979: 'The trouble with possible worlds', in Michael J. Loux (ed.) *The Possible and the Actual* (Ithaca and London: Cornell University Press): 274-316
- McDowell, J. 1984: '*De re* senses', *Philosophical Quarterly* 34(136): 283-94
- Martinich, A. P. 1996: *The Philosophy of Language* (New York: Oxford University Press)
- Maugham, W. S. 1919: *The Moon and Sixpence* (William Heinemann): http://www.gutenberg.org/ebooks/222
- Meinong A.1960: 'On the theory of objects', in R. Chisholm (ed.) *Realism and the Background of Phenomenology* (Free Press)
- Melia, J. 1995: 'On what there's not', *Analysis* 55(4): 223-9
- Melia, J. 2008: 'A world of concrete particulars', in D. W. Zimmerman (ed.) *Oxford Studies in Metaphysics* 4 (New York: Oxford University Press): 99-124
- Montague, R. 1973: 'The proper treatment of quantification in English', in P. Suppes, J. Moravcsik and J. Hintikka (eds.) *Approaches to Natural Language* (Dordrecht): 221-42
- Moore, A. W. 1993: *Meaning and Reference* (Oxford: Oxford University Press)
- Moore, G. E. 2004: *Commonplace Book 1919-1953* (Routledge)
- Nolan, D. 1998: 'Impossible worlds: a modest approach', *Notre Dame Journal of Formal Logic* 38(4): 535-73
- Nolan, D. 2004: 'Classes, worlds and hypergunk', *The Monist* 87(3): 303-21
- Parsons, T. 1980: *Non-existent Objects* (New Haven and London: Yale University Press)
- Pitcher, G. (ed.) 1964: *Truth* (Englewood Cliffs, N. J.: Prentice-Hall)

- Priest, G. 1997: ‘Sylvan’s box: a short story and ten morals’, *Notre Dame Journal of Formal Logic* 38(4): 573-82
- Priest, G. 2005: *Towards Non-Being: The Logic and Metaphysics of Intentionality* (Oxford University Press)
- Priest, G. 2006: *In Contradiction*, 2nd ed. (Oxford University Press)
- Putnam, H. 1975: ‘The meaning of “meaning”’ in his *Mind, Language and Reality* (Cambridge: Cambridge University Press): 215-71
- Quine, W. V. O. 1948: ‘On what there is’, *The Review of Metaphysics* 2(5) (Sep 1948) pp21-38. Reprinted in Quine [1953: 1-19], page references to reprint.
- Quine, W. V. O. 1951/1953: ‘Two dogmas of empiricism’, originally in *The Philosophical Review* 60(1):20-43, revised version in Quine [1953: 20-46], page references to latter.
- Quine, W. V. O. 1953: *From a Logical Point of View* (Cambridge, Massachusetts: Havard University Press)
- Quine, W. V. O. 1961: ‘Reply to Professor Marcus’ 13(4): *Synthese* 323-30
- Rayo, A. 2007: ‘Ontological commitment’, *Philosophy Compass* 2(3): 428-44
- Rayo, A. 2008: ‘On specifying truth conditions’, *Philosophical Review* 117(3): 385-443
- Reimer, M. 1997: ‘Could there have been unicorns?’, *International Journal of Philosophical Studies* 5(1): 35-51
- Reimer, M. 2001: ‘A “Meinongian” solution to a Millian problem’, *American Philosophical Quarterly* 38(3): 233-48
- Richard, M. 1990: *Propositional Attitudes* (Cambridge University Press)
- Richard, M. 2001: ‘Seeking a centaur, adoring Adonis: Intensional transitives and empty terms’, *Midwest Studies in Philosophy* 25(1): 103-107

- Routley, R., *Exploring Meinong's Jungle and Beyond*, Paper 1: http://digitalcommons.mcmaster.ca/meinong/1
- Russell, B. 1905: 'On denoting', *Mind*, New Series 14(56): 479-493
- Russell, G. 2008: *Truth in Virtue of Meaning* (Oxford University Press)
- Sainsbury, R. M. 2005: 'Names in free logical truth theory', in J. L. Bermudez (ed.) *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans* (Clarendon Press)
- Sainsbury, R. M. 2010: *Fiction and Fictionalism* (Routledge)
- Salmon, N. 1986: *Frege's Puzzle* (Cambridge, Massachusetts: MIT Press)
- Salmon, N. 1987: 'Existence', *Philosophical Perspectives* 1: 49-108
- Salmon, N. 1998: 'Nonexistence', *Noûs* 32(3): 277-319
- Saul, J. M. 1998: 'The pragmatics of attitude ascription', *Philosophical Studies* 92(3): 363-89
- Schaffer, J. 2009: 'On what grounds what', in D. Manley, D. Chalmers and R Wasserman (eds.) *Metametaphysics: New Essays on the Foundations of Ontology* (Oxford University Press): 347-83
- Schellenberg, S. 2011: 'Ontological minimalism about phenomenology', *Philosophy and Phenomenological Research* 83(1): 1-40
- Schnieder, B. 2007: 'Mere possibilities: A Bolzanian approach to non-actual objects', *Journal of the History of Philosophy* 45(4): 525-50
- Schroeter, L. 2007: 'Illusion of transparency', *Australasian Journal of Philosophy* 85(4): 597-618, DOI: 10.1080/00048400701654820
- Schroeter, L. 2012: 'Two-Dimensional Semantics', *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), E. N. Zalta (ed.), URL =

<http://plato.stanford.edu/archives/win2012/entries/two-dimensional-semantics/>.

- Searle, J. R. 1995: *The Construction of Social Reality* (Penguin Books)
- Sider, T. 1995: 'Three problems for Richard's Theory of Belief Ascription', *Canadian Journal of Philosophy* 25(4): 487-514
- Sider, T. and D. Braun 2006: 'Kripke's revenge', *Philosophical Studies* 128(3): 669-82
- Sinhababu, 2008: 'Possible girls', *Pacific Philosophical Quarterly* 89(2): 254-60
- Stalnaker, R. 1981: A defense of conditional excluded middle', in W. L. Harper, R. Stalnaker and G. Pearce (eds.) *Ifs* (Dordrecht: Reidel): 87-104
- Stearns, E. J. 1853: *Notes on Uncle Tom's Cabin* (Philadelphia: Lippincott, Grambo & Co.)
- Thomason, R. 1980: 'A model theory for propositional attitudes', *Linguistics and Philosophy* 4(1): 47-70
- Thomasson, A. L. 1999: *Fiction and Metaphysics* (Cambridge University Press)
- Van Inwagen, P. 1977: 'Creatures of fiction', *American Philosophical Quarterly* 14(4): 299-308
- Van Inwagen, P. 1990: *Material Beings* (Ithaca: Cornell University Press)
- Walton, K. L. 1978a: 'Fearing fictions', *Journal of Philosophy* 75(1): 5-27
- Walton, K. L. 1978b: 'How remote are fictional worlds from the real world?', *Journal of Aesthetics and Art Criticism* 37(1): 11-23
- Walton, K. L. 1990: *Mimiesis as Make-Believe* (Cambridge, Mass.: Harvard University Press)
- Williams, J. R. G. 2008: 'Multiple actualities and ontically vague identity' *Philosophical Quarterly* 58(230): 134-54

- Williams, J. R. G. 2012: ‘Requirements on reality’, in F. Correia and B. Schnieder (eds.) *Grounding and Explanation* (Cambridge University Press): 165-185
- Williamson, T. 1994: *Vagueness* (Routledge)
- Williamson, T. 1998: ‘Bare possibilia’ *Erkenntnis* 48(2): 257-73
- Williamson, T. 2000: *Knowledge and its Limits* (Oxford University Press)
- Williamson, T. 2002: ‘Necessary existents’, in A. O’Hear (ed.) *Logic, Thought and Language* (Cambridge: Cambridge University Press): 233-51
- Williamson, T. 2008: ‘Why epistemology can’t be operationalized’, in Q. Smith (ed.) *Epistemology: New Essays* (Oxford University Press): 277-300
- Williamson, T. 2010: ‘Necessitism, contingentism and plural quantification’, *Mind* 119(475): 657-748
- Wodehouse, P. G. 1976: *Something Fresh*, reprinted in *The World of Blandings* (Arrow Books): 11-262.
- Wolterstorff, N. 1975: ‘Toward an ontology of art works’, *Noûs* 9(2): 115-42
- Yagisawa, T. 1993: ‘A semantic solution to Frege’s puzzle’ *Philosophical Persepctives* 7: 135-154
- Yagisawa, T. 1995: "Reference *ex machina*", *Karlovy Vary Studies in Reference and Meaning*, J. Hill and P. Kotátko (eds.), (Prague: FILOSOPHIA Publications): 215-42
- Yagisawa, T. 2001: ‘Against creationism in fiction’, *Noûs* 35(s15): 153-72
- Yagisawa, T. 2010: *Worlds and Individuals, Possible and Otherwise* (New York: Oxford University Press)
- Zalta, E. 1983: *Abstract Objects: An Introduction to Axiomatic Metaphysics* (Dordrecht: D. Reidel)
- Zalta, E. 2000: ‘The road between pretense theory and abstract object theory’, in T. Hofweber & A. Everett (eds.), *Empty Names,*

Fiction, and the Puzzles of Non-Existence (Stanford: CSLI Publications): 117-47