University of Groningen

Fictional Names in Psychologistic Semantics

Maier, Emar

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Link to publication in University of Groningen/UMCG research database

Fictional names in psychologistic semantics[*]

Emar Maier
University of Groningen

Abstract Fictional names pose a difficult puzzle for semantics. How can we maintain that Frodo is a hobbit, while admitting that Frodo does not exist? To dissolve this paradox I propose a way to formalize the interpretation of fiction as 'prescriptions to imagine' (Walton 1990) within a psychologistic semantic framework in the style of Kamp (1990). In the context of an information exchange, the interpretation of an assertion triggers a dynamic update of a belief component in the interpreter's mental state, while in the context of a fictional narrative, a statement like *Frodo is a hobbit* triggers an update of an imagination component. In the computation of these updates, proper names – referential, empty, or fictional – are uniformly analyzed as presupposition triggers. The possibility of different attitude components in a single mental state sharing discourse referents and thereby referentially depending on each other ultimately allows us to account for the central paradox of fictional names and related puzzles.

Keywords: semantics; fiction; imagination; reference; (fictional) proper names; mental states; mental files; (propositional) attitudes; presupposition; Discourse Representation Theory

1 The paradox of fictional names

In a much discussed paper, Radford (1975) introduces what he calls the paradox of fiction. When someone recounts the terrible things that happened to her, I may be moved to tears, but when she admits she made the whole thing up, this sadness quickly gives way to anger (or embarrassment). With

fictional narratives this is not the case: we can feel sad for a character's tragic fate while at the same time being aware that it's all made up.

A version of this paradox comes up when analyzing the semantics of fictional names (i.e., names of fictional characters, like *Frodo* or *Sherlock Holmes*). Normally, when someone tells me what cute things their daughter said the other day, they can't consistently follow that up by saying that they don't have children. Yet, in the context of reading or discussing a work of fiction, similarly contradictory claims seem to be perfectly fine. What I will refer to as the paradox of fictional names is the intuition that we can consistently utter (and accept) both fictional statements like (1a), in which the fictional name *Frodo* seems to refer to a flesh and blood creature, and metafictional statements like (1b), which deny that the referent of *Frodo* exists.

(1)   a.   Frodo is a hobbit born in the Shire.
        b.   Frodo is a fictional character made up by Tolkien.

Although the interpretation of fiction has received little attention in formal semantics, there is a vast literature on the semantics of fiction – and fictional names in particular – in philosophy. In section 2 I briefly summarize and classify the main philosophical approaches to fictional names in order to situate my own proposal in this ongoing debate. In section 3 I further situate my proposal with respect to the traditionally distinct conceptions of meaning in truth-conditional, dynamic, and cognitive semantics. The actual proposal then is laid out in sections 4–9.

2   Background: Philosophical approaches to fictional names

In the philosophical literature on fiction and fictional names we can discern roughly three types of approaches: realism, semantic anti-realism, and pragmatic anti-realism. My own proposal will fall within the latter category, but in this section I briefly review each approach as it applies to the semantic paradox in (1).

Realists extend their ontology to include non-existent, abstract, and/or fictional objects that can serve as genuine referents of fictional names (Meinong 1904, Thomasson 1999). Realism promises a uniform semantics for the names in both (1a) and (1b), viz., as referential terms denoting a fictional object. The realist's semantics has been subjected to ridicule (Russell 1905, Quine 1948), as well as a variety of serious objections. For instance, saying that Sam carried Frodo from Mount Doom does not intuitively entail that Sam carried

a non-existent or abstract object (cf. Lewis 1978). Various strategies to circumvent this objection and others have been explored, but they always seem to involve postulating *ad hoc,* invisible distinctions in the syntax–semantics interface. Zalta (1983), for instance, puts an ambiguity in the notion of predication: (1a) involves an object "encoding" a property, while (1b) involves "exemplification". Kripke (2011), by contrast, puts an ambiguity in the name: *Frodo* in (1a) rigidly refers – within a pretense, see below – to a hobbit born in the Shire, while the similar looking name in (1b) picks out an abstract object created somewhere in the U.K. in the twentieth century. One of my desiderata for a semantics of fictional discourse is that it treats all names uniformly.

Anti-realists hold that the name *Frodo* does not refer. In a truth-conditional semantic framework, this means that the extension of the name is not defined. By compositionality, it follows that a fictional statement like (1a) cannot be true at all. Hence, according to Frege (1892), (1a) would be neither true nor false, while for Russell (1905) it would be false. Such proposals may be enhanced with the addition of a hidden, intensional "fiction operator" (Lewis 1978), so that (1a) abbreviates *in all possible worlds compatible with the given fiction, Frodo is a hobbit born in the Shire,* which could be true, even if the embedded statement itself is truth-valueless or false.[1] The motivation behind these semantic anti-realist approaches is to reconcile ontological sobriety with standard truth-conditional semantics.

One problem with such semantic anti-realist approaches is that they do not extend straightforwardly from (1a) to (1b), a metafictional assertion that contains the same name *Frodo*, but is clearly true, and cannot be viewed as prefixed with a fiction operator. In addition, if we follow Kripke (1980) in treating names as rigid designators, i.e., terms whose sole meaning is their actual referent, then (1a) and (1b) would not even express a proposition. Yet, even if we were to admit they are not literally true, these statements are surely meaningful and often pragmatically felicitous. Thus, Lewis (i) excludes metafictional statements from his investigation, and (ii) resorts to a descriptive analysis of fictional names. In conclusion, it seems impossible to reconcile a classical truth conditional semantics with a uniform semantics

---

1 Lewis's semantic approach to fiction is further explored and formalized by e.g. Bonomi & Zucchi (2003). A somewhat different incarnation of the idea of a hidden fiction operator, hinted at by an anonymous referee, would be to use Stone's (1997) analysis of modals in terms of scenarios and analyze statements like (1a) as *in the scenario dynamically constructed by the interpretation of Tolkien's text, Frodo is a hobbit born in the Shire.* Importantly, as far as I can tell, these (potential) alternative implementations are subject to the two objections raised below.

of proper names in truthful, fictional, and metafictional contexts – even with the addition of an intensional fiction operator.

Pragmatic anti-realists, finally, analyze the interpretation of fictional names and fictional discourse at the speech act level. Fictional statements like (1a) are not assertions, but pretend assertions (Searle 1975, Kripke 2011), or some wholly different speech act (Currie 1990, Bauer & Beck 2014). The most influential is Walton's (1990) analysis on which fictional texts are "prescriptions to imagine". Unlike regular assertions, fictional statements do not express information about the way the world is but rather invite the reader to imagine a certain state of affairs. Thus, (1a) asks the reader to imagine a world in which Frodo is a hobbit born in the Shire. A crucial benefit of this analysis is that it gives an account of the use of fictional statements without committing us to non-existent fictional entities in the actual world.

The main objection against Walton's analysis is that it fails to do justice to intuitions of aboutness associated with the use of fictional names, i.e., the intuition that (1a) expresses a singular proposition, about Frodo. This has been brought out especially clearly by Friend's (2011) argument from *counterfictional imagination*, which I discuss in section 7.2. In addition, like semantic anti-realism, pragmatic anti-realism does not straightforwardly extend to metafiction. After all, (1b) is clearly not an invitation to imagine that Frodo is fictional.

My aim in this paper is to propose a uniform, formal pragmatic account of the interpretation of names in fictional and metafictional statements based on Walton's suggestion that fictional statements are prescriptions to imagine. My analysis will be couched in a dynamic semantic framework – more specifically, a psychologistic version of DRT in the style of Kamp (1990, 2015), where interpretation means updating a representation of the interpreter's mental state. I will show how Kamp's DRT-based formalism for representing mental states and the way we update them with linguistic information can reconcile an imagination prescribed by (1a) with a belief conveyed by (1b), while maintaining a fully uniform analysis of proper names – referential, fictional, empty, or otherwise.

3   Psychologistic semantics

Our first job is to find a suitable formal semantic framework in which to capture the pragmatic anti-realist starting point that fictional statements are prescriptions to imagine. A traditional Montagovian conception of semantics as the compositional derivation of truth conditions seems ill-suited for the

4

job.[2] In addition to the problems for semantic approaches reviewed in the previous section, we already classified ours as a pragmatic rather than a semantic approach to fiction interpretation, i.e., we're not primarily interested in the truth conditions of fictional statements, but in the way readers interpret them.

Dynamic semantics is a more promising candidate, as it moves the focus away from truth conditions and blurs the line between semantics and pragmatics. On the dynamic conception of meaning, interpreting a fictional statement would amount to an update of a given body of information, a "context". There are different ways that this central notion of a context can be formalized, e.g. as a DRS, a set of worlds, or an information state. But, more fundamentally, there are also different views on what kind of information constitutes a context. The latter aspect is of crucial importance here, and thus requires some discussion.

Following Stalnaker (1970), a context is typically thought of as the common ground – roughly, the body of information jointly accepted by the conversational participants at a given point in the discourse. This conception of context abstracts away from the distinct individual mental states of different speech act participants. Conversation is essentially a cooperative endeavor by a group of agents trying to reduce uncertainty in the common ground. We see such a participant-neutral conception of dynamic semantics for instance in the seminal works of Heim (1982), Groenendijk & Stokhof (1991) and van der Sandt (1992).

With this abstraction in place, much progress has been made in modeling phenomena in the semantics/pragmatics interface. But there are limits to what we can achieve in this way, especially with respect to pragmatic phenomena. Recently, semanticists are increasingly interested in linguistic phenomena that require distinguishing and tracking the information states of individual speech act participants. Recent examples include Farkas & Bruce's (2009) analysis

---

2 Recently, Eckardt (2014) and Bauer & Beck (2014) have proposed analyses of fiction interpretation that stick more closely to the classic compositional semantic framework than the current proposal, though both also involve unmistakably pragmatic components (Eckardt introduces a 'story update' mechanism, and Bauer & Beck a 'fictional-assert' speech act operator). However, their aims are quite far removed from those in the current paper: Eckardt is interested in the semantics of free indirect discourse, using the story update mechanism primarily to ensure that the author is not committed to the truth of the story; while Bauer & Beck are interested in explaining how literary texts can be meaningful to the reader in the real world, given that they seem to describe distant possible worlds. Neither is concerned with giving a uniform account of fictional and regular proper names, nor with metafictional statements. I'll focus in this paper on establishing a positive proposal of my own, leaving a thorough comparison with, or perhaps integration of, these approaches for later.

of polar questions in terms of a fine grained discourse model, Sæbø's (2012) analysis of specific indefinites as referential for the speaker but existential for the hearer; Wechsler's (2010) analysis of plural pronominal paradigms in terms of speaker–hearer asymmetries in communicating *de se* attitudes; and Cohen & Krifka's (2014) analysis of superlative quantifiers using a model of complex commitment spaces.

I propose to add fiction interpretation to this list of phenomena that require us to move beyond the abstract common ground model prevalent in dynamic semantics. Intuitively, interpreting a fictional text is just not a matter of updating some abstract intersubjective common ground between speaker/writer and hearer/reader. To make sense of the philosophically compelling intuition that fictional statements are prescriptions to imagine, we need a framework that captures what happens in the mind of the reader when she interprets a text.

Discourse Representation Theory (DRT) offers just that.

> DRT is a theory of interpretation in two senses of the word. It is a theory of meaning and it is also a theory of language understanding. DRT is a cognitivist theory, which is based on the insight that a semantic theory must of necessity take into account the mental processes involved in handling language. (Geurts 1999: xi)

Despite the prevalence of abstract common ground interpretations of context in dynamic semantics, this psychologistic interpretation has always been an integral part of Kamp's conception of DRT.

> DRT has from its earliest beginnings been a theory that makes claims about the psychological relevance of the forms in which human interpreters compute and represent the semantic content of the linguistic inputs they get, rather than limiting itself to using those representations solely for the purpose of making predictions about the truth conditions of the sentences and discourses for which they have been constructed. (Kamp 2015: 266)

In sum, the idea behind DRT is to model the interpretation of a discourse in terms of the way sentences update a structured mental representation, a Discourse Representation Structure (DRS).[3]

The current application requires a more finegrained model of mental states than simply a DRS. Bringing Walton's pragmatic theory of fiction to

---

3 In recent work, Brasoveanu & Dotlacil (2015) take this one step further in the direction of actual psychology by implementing incremental DRS construction in ACT-R, a well-established computational framework for modeling human cognitive processing (Anderson & Lebiere 1998).

dynamic semantics, the idea I want to work out is that, while a plain assertion in an information exchange triggers (or is intended to trigger)[4] an update of the hearer's beliefs, a fictional statement instead triggers (or is intended to trigger) an update of her imagination. This view presupposes a model of the interpreter's mental state as a complex involving beliefs, imaginations, and, presumably, other distinct (but crucially interdependent, as we will see) attitudes. Moreover, to do justice to Kripkean intuitions of intentionality and rigidity associated with (fictional) proper names, we also need a way to represent referential intentions connecting our mental states with objects of acquaintance in the world.

Overview

The rest of the paper is structured as follows. In section 4 I develop a suitable DRT-based formalism for the representation of complex, anchored mental states, based primarily on Kamp (1990, 2003, 2015) but incorporating also some insights from the related mental files program in philosophy (Recanati 2012). In sections 5 and 6 I present my proposal for interpreting non-fiction and fiction both as updates on such complex mental states. The interpretation of names in fiction leads to a dilemma, which I discuss in sections 7 and 8. In section 9 I return to the original paradox. I demonstrate how we can analyze fictional and metafictional statements in a DRT-based psychologistic semantic framework in which fictions are prescriptions to imagine and proper names are uniformly analyzed as presupposition triggers.

4   Representing mental states

In this section I present an extension of the DRT formalism that deals with the representation of mental states as complexes of interconnected attitudes and anchors. Formal details are relegated to an appendix.

---

4 This is an oversimplification. As Searle (2001) puts it "a speaker can make an assertion quite satisfactorily without giving a damn whether the hearer assumes what he says is true. . . . He might say, 'I don't care whether you assume that it is raining, all the same it's raining'." (I owe this reference to Hans-Martin Gärtner). I will not attempt to provide here a fully worked out analysis of the speech acts involved in producing either factual/informative or fictional statements. Instead I focus merely on describing the normally intended (perlocutionary) effect on the interpreter, which I take to be a first step toward a full speech-act-theoretic analysis.

4.1 Basic DRT

A DRS is usually depicted as a box with two compartments. The top compartment, or universe, contains discourse referents $(x, y, x_1, \ldots)$, representing the entities that the discourse is about. The bottom compartment contains descriptive conditions involving these discourse referents. Conditions can be atomic (of the form $R(x_1 \ldots x_n)$) or complex (featuring logical operators like $\Box, \neg$ or $\rightarrow$ and one or more embedded subDRSs). An example DRS representation of a simple discourse will suffice to illustrate the basic DRS syntax:

(2)   John is a farmer. If he owns a donkey, he doesn't beat it.
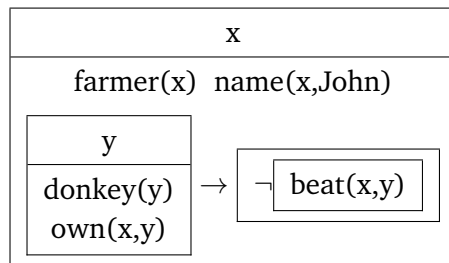
The DRS language is really just a minor variant of first-order logic in both syntax and semantics (details on both in the appendix). DRSs have a static, truth-conditional interpretation. With this semantics, the DRS in (2) is equivalent to the following first-order formula:

(3) $\exists x [\mathrm{farmer}(x) \wedge \mathrm{name}(x, \mathrm{John}) \wedge$
$$\forall y [[\mathrm{donkey}(y) \wedge \mathrm{own}(x, y)] \rightarrow \neg[\mathrm{beat}(x, y)]]]]$$

In its original formulations, the dynamic nature of DRT resides wholly in the so-called construction algorithm.[5] This is an algorithm to turn a given context DRS and syntactically parsed sentence into an updated context DRS, reflecting the information growth caused by the interpretation of that sentence. The construction algorithm for instance specifies that an indefinite noun phrase like *a donkey* adds a new discourse referent (y) to the closest DRS universe and a condition $(\mathrm{donkey}(y))$ below it. A pronoun, by contrast, does not introduce a new discourse referent but initiates a search for an already established discourse referent higher up in the DRS under construction (e.g., *it*=y). Following van der Sandt (1992), later versions of DRT typically split the construction algorithm into two stages: (i) the compositional construction of a Preliminary DRS purely on the basis of the syntactic structure of the

---

5 Later versions of DRT also provide genuinely dynamic semantic interpretations of fragments of the DRS language, i.e. mapping DRSs onto information states and context change potentials (Kamp et al. 2003).

sentence, and (ii), the resolution of presuppositions in the DRS context. I will present a version of this two-stage interpretation procedure in section 5.1 below.

4.2 Attitude Description Sets

We've seen that Kamp originally intended DRT not just as a logical description of truth-conditions and abstract common ground updates, but rather as a model of what goes on in the mind of the individual interpreter. To work this out, Kamp has been developing a formalism for the representation and interpretation of complex mental states. I will use the term "Attitude Description Theory (ADT)" to refer to this theory of attitudes and mental state representations in terms of "Attitude Description Sets (ADS)".

Formally, an ADS is a set of labeled DRSs representing the content of the various interrelated attitudes that make up an agent's mental state. Each DRS is paired with a label indicating the mode of the attitude it represents, e.g. BEL for belief and DES for desire. In addition to attitudes proper, Kamp assumes something like mental files (Perry 1980, Recanati 2012) – DRSs that serve as descriptive internal representations of objects the agent is acquainted with. These so-called *internal anchors* are labeled with the mode indicator ANCH.

Consider first an example involving direct perception. The ADS in (4) represents a fragment of the mental state of an agent who sees a glass in front of her, thinks it contains water, hopes it's cold, imagines it's wodka and finally intends to pick it up and drink from it.

$$(4) \quad \left\{ \left\langle \text{ANCH}, \boxed{\begin{array}{c} x \\ \hline \text{glass(x)} \\ \text{see(i,x)} \end{array}} \right\rangle , \left\langle \text{BEL}, \boxed{\text{water(x)}} \right\rangle , \left\langle \text{DES}, \boxed{\text{cold(x)}} \right\rangle , \\ \left\langle \text{IMG}, \boxed{\text{wodka(x)}} \right\rangle , \left\langle \text{INT}, \boxed{\begin{array}{c} \text{pick.up(i,x)} \\ \text{drink(i,x)} \end{array}} \right\rangle \right\}$$

The first component of (4) is an internal anchor. It tells us how the agent is acquainted with a given object, and introduces a discourse referent to stand for that object. The other components represent the various attitudes the agent has toward this object, viz., the belief that it contains water, the intention to drink from it, etc. The special discourse referent i, finally, represents the

9

agent's *self-file*, a non-descriptive, irreducibly indexical representation of herself.[6]

Below I elaborate on two crucial aspects of the ADT formalism that I'll rely on in my analysis of fiction interpretation: the notion of an anchor as a mental file, i.e., a way to reconcile singular attitudes with descriptive modes of presentation (4.3), and the sharing of discourse referents across attitudes, which allows us to model *parasitic attitudes* (4.4).

4.3   Referential intentions and external anchors

Internal anchors are meant to capture how the agent is acquainted with objects in the external world. Hence, we could say that an internal anchor in an ADS *refers* to an extra-mental entity. In this respect, Kamp's anchors are very similar to Recanati's (2012) mental files: on the one hand they are descriptive bodies of information, functioning like cognitive modes of presentation, but on the other hand they are directly referential. As Recanati puts it:

> mental files are 'about objects': like singular terms in the language, they refer, or are supposed to refer. They are, indeed, the mental counterparts of singular terms. What they refer to is not determined by properties which the subject takes the referent to have (i.e. by information – or misinformation – in the file), but through the relations on which the files are based. The reference is the entity we are acquainted with (in the appropriate way), not the entity which best 'fits' information in the file. (Recanati 2012: 35)

Applied to ADT, formalizing such a non-descriptive, relational interpretation of anchors requires that we introduce a formal device to specify the actual object of acquaintance in addition to the (merely descriptive) internal anchor. For this purpose, Kamp introduces *external anchors*. Formally, an external anchor is just a partial assignment function, mapping internally anchored discourse referents to the objects that are the actual source of the information described in the corresponding internal anchors. A contextually "situated" version of the ADS in (4) should thus include an external anchor mapping the discourse referent x to some actual glass of water.[7]

Note that the agent herself does not have access to the external anchor. Her behavior and practical reasoning are guided by the information in the internal anchors and attitudes. In philosophical terminology, external anchors are

---

6 Formally, i should be treated as a separate internal anchor that always picks out the center of any doxastic alternative.

7 I discuss cases of perceptual error and faulty anchoring in section 7.3.

needed to capture the *wide content* of an ADS, but irrelevant for determining its *narrow content* (and *vice versa*). What the presence of an internal anchor in an ADS does is signal a referential intention on the part of the agent. In other words, the agent takes all her internal anchors to have corresponding external anchors that link them to the objects she believes she's acquainted with. As (Kamp 2011: 5) puts it, internal anchors "presuppose" external anchors, and therefore internal anchors without external ones are faulty.

For our current purposes we'll henceforth restrict attention to the narrow, (i.e., psychologically relevant) content of mental states, and refrain from adorning example ADSs with external anchors. We return to the relation between internal and external anchoring in section 7.3, where I argue against an analysis of fictional names in terms of faulty anchors. Furthermore, in section 5.2 I'll introduce vicarious anchoring, a type of anchoring not grounded in perceptual acquaintance but in causal–historical reference chains.[8]

4.4   Parasitic attitudes

In ADT, anchors are used to represent *de re* attitudes. In the glass-of-water example we represented the agent's *de re* belief about the glass, that it contains water, by using in the belief the discourse referent x that was introduced by an internal anchor. This configuration captures the intentionality of *de re* attitudes in that, if all is well, the internally anchored discourse referent leads us via its external anchor to an actual *res* – independently of whether or not that *res* satisfies the descriptive content in the internal anchor.

This analysis of *de re* thought requires that we allow DRSs with free variables in our ADSs. We'll say that an ADS is globally well-formed if all free variables of each component are grounded in the universes of other components. Thus, our simple example in (4) is well-formed. We'll say that the open attitude DRSs in (4) are *referentially dependent* on the internal anchor.

In ADT, referential dependence is not restricted to the dependence of attitudes on internal anchors. Let me illustrate this with an application to a linguistic puzzle. Consider the attitude ascription in (5) (cf. Heim 1992, Elbourne 2010).

---

8 I should point out here that my use and interpretation of anchors differs from Kamp's, especially considering the multiply anchored entity representations of Kamp (2015). In fact, the combination of an internal and external anchor in the current setup corresponds more closely to Recanati's notion of a mental file than to Kamp's notion of an entity representation. In section 7.3 we'll encounter another crucial difference between Kamp's anchors and mine.

(5)   Hans wants the ghost in his attic to be quiet tonight.

This report arguably has a reading that is neither *de re* (there's a unique ghost and Hans wants it to be quiet) nor *de dicto* (Hans wants it to be the case that there is a unique ghost that is quiet). On this reading Hans may have a *de dicto* belief that there is a ghost with a desire that it be quiet. In a classic Hintikka-style theory of belief and desire as propositional attitudes we cannot describe such a belief–desire complex. The closest we can get is a logical form like (6), but there the final x is actually a free variable, which gives the wrong interpretation.

(6) $\text{BEL}_h[\exists x[\text{ghost}(x) \wedge \text{in.attic}(x)]] \wedge \text{DES}_h[\text{quiet}(x)]$

In ADT, by contrast, we can straightforwardly capture the mental state ascribed to Hans as involving a desire referentially dependent on a (*de dicto*) belief:

(7)
$$\left\{ \left\langle \text{BEL}, \begin{array}{|c|} \hline x \\ \hline \text{ghost}(x) \\ \text{in.attic}(x) \\ \hline \end{array} \right\rangle, \left\langle \text{DES}, \boxed{\text{quiet}(x)} \right\rangle \right\}$$

I should stress that (7) is just a representation of Hans's mental state, not of the truth conditions of the English sentence in (5). However, having an adequate DRT-based syntax and semantics for representing parasitic mental states like this is an important first step towards a compositional semantics of linguistic ascriptions like (5).[9]

    Dependence of non-doxastic attitudes on anchors and beliefs is a common theme in philosophical and linguistic puzzles about attitudes. We can extend the ADT analysis of the parasitic reading of (5) to solve, for instance, one of Karttunen's (1973) puzzles about presupposition projection:

(8)   Bill believed that Fred had been beating his wife and he hoped that Fred would stop beating her.   [Karttunen 1973]

On the classical approach, where *believe* and *hope* are intensional operators, the introduction of the event of Fred beating his wife in the first conjunct cannot bind the pronoun *her*, nor satisfy the presupposition triggered by *stop beating her*. But, intuitively, the presupposition does get satisfied somehow, and the pronoun bound, since the sentence as a whole is perfectly felicitous in a context in which Bill's belief is mistaken and Fred never beat his wife, or

---

9 For the second step, see Maier (2015a).

is not even married. In ADT we could account for this intuition by analyzing the second conjunct as introducing a hope referentially dependent on the earlier belief inside a single, complex description of Bill's mental state.

On the one hand, the expressive power of Kamp's system goes well beyond representing these types of belief parasitism. It freely allows chains of referential dependencies between any modes of attitudes, and even between multiple attitudes simultaneously. Indeed, we occasionally encounter evidence of non-belief-dependencies in puzzles involving natural language ascriptions:[10]

(9)   Alice fears that there is a squirrel in her kitchen cabinets. She hopes to trap it alive.   [Schoubye 2013]

I will exploit the full expressivity of ADT in my analysis of fictional names, for instance, by having (counterfictional) imaginations and beliefs depend on (fiction-induced) imaginations.

On the other hand, it is also worth emphasizing the limits of the ADT approach to parasitic attitudes. In particular it does not extend to seemingly closely related puzzles involving intentional identity, (10a), and modal subordination, (10b).

(10)   a.   Hob thinks a witch has blighted Bob's mare and Nob wonders whether she killed Cob's sow.   [Geach 1967]
           b.   A wolf might come in. It would eat you first.   [Roberts 1987]

We see in both of these a pronoun bound by an embedded antecedent that is predicted to be inaccessible on a classical possible worlds analysis of the embedding operator. However, despite superficial similarities, there are important differences between the phenomena in (10) and the parasitic attitudinal dependencies discussed above.

Regarding intentional identity,[11] note that the attitude of wondering ascribed to Nob in the second conjunct of (10a) is not itself parasitic or

---

10 There appear to be some restrictions on parasitic ascriptions involving dependence on other attitudes than belief. I return to this matter in section 8.2. See in particular footnote 27 for a comparison with a simpler approach in which parasitism is restricted to non-doxastic attitudes depending on a doxastic base.

11 In (Maier 2015a) I also note that at the level of linguistic reports, anaphora across attitude ascriptions in (multi-agent) intentional identity cases is subject to some additional constraints as compared to (single agent) parasitic ascriptions, as witness the following minimal pair:

(i)   Mary knows that Sue won't come. John is more optimistic. John believes that Sue will come. He/*Mary hopes that Sue's sister will come too.

referentially dependent on Hob's beliefs in the way that the hope ascribed to Bill in (8) is parasitic on his belief. Despite the anaphoric dependency in (10a), the underlying attitudes ascribed to Hob and Nob are independent in the sense that (10a) can be true even if Nob knows nothing about Hob's thoughts. In (8), by contrast, Bill's hope is dependent on his belief in that it cannot even be described or paraphrased without recourse to the belief.

Regarding modal subordination, note that the "subordinated" second sentence is usually assumed to get a conditional interpretation, i.e. *If a wolf came in, it would eat you first* (Roberts 1989, Frank & Kamp 1997, Geurts 1999). For our parasitic ascriptions, by contrast, such a conditional paraphrase seems off: (5) does not mean *if there was a ghost in the attic, Hans wants {it/the ghost in his attic} to be quiet*.[12] I tentatively conclude that parasitic attitudes, intentional identity, and modal subordination are truly distinct phenomena. I leave it for future research to determine whether some existing account of modal subordination (e.g. Stone 1997) and/or of intentional identity (e.g. Edelberg 1992, van Rooy 2000) may be extended to account for parasitic attitude ascriptions, or even fictional names and (meta-)fictional statements, without resorting to something like ADT.

4.5   On the semantics of ADT

We can exploit the DRT foundations of ADT to give not only a precise syntax but a model-theoretic semantics for ADSs. In this respect our model has a crucial advantage over similar but more or less informal mental representation frameworks, such as Fauconnier's (1994) Mental Spaces in linguistics, or Recanati's (2012) Mental Files in philosophy. Unfortunately, the flexibility of referential dependence requires a rather complex formalism, so I will here attempt only a rough outline of the form of the semantics laid out by Kamp et al. (2003) and Maier (2016), moving the details to the appendix.

A semantics for ADSs should tell us under what conditions a given ADS correctly represents part of an agent's mental state. The first decision to make here is how we want mental states to be given set-theoretically in a model. On a classical possible worlds approach we'd have sets of possible worlds for each attitudinal mode, i.e., a set of doxastic alternatives (*Dox*), a set of buletic alternatives (*Bul*), etc.[13] On a sententialist approach, by contrast, we'd have

---

12 The root of the problem seems to be that Kratzer's (1981) bipartite analysis of modals in terms of modal base and ordering source, and the subsequent unification of conditionals and modals, does not straightforwardly extend to attitude ascriptions.

13 The use of the essentially indexical self-representation i actually requires the use of centered propositions, but I will not go into such details here.

fully structured syntactic objects, isomorphic to ADSs themselves perhaps (Asher 1986). Kamp explores an interesting middle route between these two extremes, but to better grasp the motivation behind his ADS interpretation formalism, let's first consider what the classical possible worlds approach would look like (and why it fails).

Consider a simple example ADS that contains an anchor DRS (K_{ANCH}), a belief DRS (K_{BEL}), and a desire DRS (K_{DES}). A classical semantics for such an ADS has three components. First, the model specifies the relevant agent's beliefs and desires as sets of possible worlds, i.e., $Dox, Bul \subseteq W$. Second, it appeals to standard DRT semantics to determine the possible worlds propositions expressed by the belief ($[\![K_{\text{BEL}}]\!]$), the desire ($[\![K_{\text{DES}}]\!]$), and the anchor ($[\![K_{\text{ANCH}}]\!]$). Third, it checks whether each of these propositions is entailed by the corresponding attitude from the model: $Dox \subseteq [\![K_{\text{BEL}}]\!]$, $Bul \subseteq [\![K_{\text{DES}}]\!]$, and, following Kamp et al.'s (2003) interpretation of the narrow content of internal anchors as beliefs,[14] $Dox \subseteq [\![K_{\text{ANCH}}]\!]$. If these relations hold, the ADS accurately – if partially – captures the (narrow, propositional) mental state of the agent. This third component is essentially the psychologistic version of a Tarskian definition of truth: instead of specifying when a formula is true (in a world and/or time provided by the model) we now specify when a formula correctly captures a mental state (as given by the model).

The problem with this simple semantics is that it can't handle referential dependencies, the hallmark of ADT, because, taken in isolation, a dependent attitude DRS by definition contains free variables and hence doesn't express a proposition. Kamp's solution is to switch from a possible worlds semantics for the attitude DRSs, to a genuinely dynamic semantics in terms of context change potentials (CCP, Heim 1983). The advantage is that DRSs with or without free variables always express such a dynamic content. It does mean that we have to give up the classical model of propositional attitudes as sets of possible worlds (Dox, Bul, etc.). Kamp represents mental states model-theoretically as Information State Based Attitudinal State representations (ISBAS) – sets of CCPs paired with attitudinal mode labels. Concretely, a simple Kampian ISBAS suitable for the interpretation of the ghost-in-the-attic example in (7) would be $\{\langle \text{BEL}, J_1 \rangle, \langle \text{DES}, J_2 \rangle\}$, where the J's are CCPs.[15] The

---

14 We briefly return to the idea of reducing internal anchors to beliefs below in section 7.3. Note that this reduction is initially plausible as long as we restrict our attention to narrow mental content. The agent holding the glass of water in (4) believes that he is holding a glass. To determine the wide content of an ADS we'd have to ignore the descriptive content of the internal anchor completely and instead just fix the reference of the anchored discourse referents with the external anchor.

15 A suitable ISBAS may also contain additional components beyond the ones corresponding to the ADS components.

central definition of Kamp's semantics is then that an ADS **K** captures an ISBAS **A** iff for each labeled attitude DRS $\langle l, K' \rangle \in \mathbf{K}$ there is a corresponding labeled attitude $\langle l, J \rangle \in \mathbf{A}$ such that the CCP J entails the CCP expressed by the corresponding attitude DRS K'.

The main challenge in this setup is to define a sensible notion of entailment between (dependent) CCPs within the context of their surrounding ADS/ISBAS. In fact, I've argued elsewhere that Kamp's answer to this challenge is insufficient for dealing with counterfactual attitudes like imagination (Maier 2016). Since imagination is crucial to our current enterprise I present in 8.3 (and the appendix) my alternative implementation, which shares the general ideas behind Kamp's proposal, but uses so-called 'two-dimensional information states' rather than CCPs to model the contents of individual attitudes (relative to other, background attitudes). For now, I hope the informal sketch of the Kampian semantics above suffices to give an impression of what it means to give a semantics for ADT, and what such a semantics looks like. The curious reader may want to skip ahead to 8.3 and/or the appendix for full details.

Summing up, ADT is a logical framework for representing mental states as complexes of attitudes and anchors. An important feature of the framework is the sharing of discourse referents across distinct attitude DRS within a single ADS, modeling referential dependencies between different attitudes. This feature can be used to solve a variety of puzzles involving non-doxastic attitudes parasitic on belief. I will exploit it below to analyze our semantic paradox of fictional names.

5   Interpreting proper names

With the ADT formalism for representing mental states in place, we turn to linguistic interpretation. In this section I sketch an interpretation algorithm based on the theory of presuppositions-as-anaphora developed by van der Sandt (1992) and Geurts (1999). The focus is on the interpretation of proper names, for which I combine elements from two recent DRT analyses, by Maier (2015b) and Kamp (2015).

5.1   Interpretation as presupposition resolution

According to van der Sandt (1992), interpretation proceeds in two stages. First the construction algorithm turns a sentence into a preliminary DRS (PrelDRS), which simply represent definites and other presupposition triggers *in situ*, merely marking them as "to be resolved". In the second stage we add

this PrelDRS representation of the sentence to the context DRS and resolve all presuppositions. The resolution algorithm then looks for suitable antecedents for the presuppositions to bind to, or else accommodates such antecedents.

This theory offers a straightforward analysis of proper name interpretation: proper names are presupposition triggers (Geurts 1997). By way of illustration, consider the interpretation of an utterance of *Mary is a spy*. The construction algorithm analyzes *is a spy* as a unary predicate, while *Mary* triggers the presupposition that there is someone named 'Mary'. The preliminary DRS representation of the sentence thus looks like this, with the dashed DRS representing the unresolved presupposition:

(11)   a.   Mary is a spy

b.

$$\begin{array}{|c|}\hline \mathrm{spy}(x) \\ \hline \begin{array}{|c|} \hline x \\ \hline \mathrm{name}(x,\mathrm{Mary}) \\ \hline \end{array} \\ \hline \end{array}$$

In van der Sandt's original formulation we proceed to stage two by adding the preliminary DRS to a DRS representing the common ground. In our psychologistic framework we instead add it to an ADS representation of the hearer's mental state. Let's assume that the interpreter, Sue, already has a perception-based anchor for the person currently speaking to her, say John. For the interpretation of the name *Mary*, we have to distinguish two possibilities: either Sue already knows the Mary that John is talking about, or this is the first time she hears this name.

In the first case, Sue's mental state looks something like this, prior to interpreting John's utterance:

(12)

$$\left\{ \left\langle \mathrm{ANCH}, \begin{array}{|c|} \hline y \\ \hline \mathrm{name}(y,\mathrm{Mary}) \\ \mathrm{colleague}(y,z) \\ \mathrm{friend}(i,y) \\ \hline \end{array} \right\rangle, \left\langle \mathrm{ANCH}, \begin{array}{|c|} \hline z \\ \hline \mathrm{name}(z,\mathrm{John}) \\ \mathrm{talk.to}(z,i) \\ \hline \end{array} \right\rangle \right\}$$

For an assertion that's part of a straightforward information exchange, where the hearer trusts the speaker to know what they are talking about, the hearer adds the preliminary DRS to a (new) belief box in her ADS, i.e., $(12) \cup \{\langle \mathrm{BEL}, (11\mathrm{b})\rangle\}$.[16] At this point we apply the presupposition resolution algorithm to identify potential antecedents for the unresolved presuppposi-

---

16 In other types of communicative exchanges, other attitude compartments and/or modalized contents (e.g., that the speaker herself believes, or wants me to believe, the proposition

tions. In this example the presupposition that there is an x named 'Mary' is satisfied by the already established anchor y representing the interpreter's friend Mary. We bind x to y by unifying these discourse referents and removing the (now satisfied) presuppositional condition, which leads to the following final output mental state:

$$
(13) \quad \left\{ \left\langle \mathrm{ANCH}, \begin{array}{|c|} \hline y \\ \hline \mathrm{name(y,Mary)} \\ \mathrm{colleague(y,z)} \\ \mathrm{friend(i,y)} \\ \hline \end{array} \right\rangle, \left\langle \mathrm{ANCH}, \begin{array}{|c|} \hline z \\ \hline \mathrm{name(z,John)} \\ \mathrm{talk.to(z,i)} \\ \hline \end{array} \right\rangle, \right. \\ \left. \left\langle \mathrm{BEL}, \begin{array}{|c|} \hline \mathrm{spy(y)} \\ \hline \end{array} \right\rangle \right\}
$$

This output captures the basic intuition that the name is interpreted referentially (as opposed to, say, descriptively or anaphorically/bound), not because names are analyzed as directly referential singular terms, but because the lexically triggered presupposition is bound by an internal anchor, which in turn is the mental correlate of the agent's acquaintance with an external *res*.

In the second case, where Sue does not have an independently grounded anchor for Mary, we also want to derive a referential reading. To achieve this we need to extend our notion of anchoring.

5.2 Vicarious anchors and causal chains

When Sue's mental state provides no suitable antecedent to bind the name to, she either has to reject John's utterance ("Huh, who are you talking about?") or somehow "accommodate" a suitable antecedent. Let's explore the latter option.

Names want to be bound by internal anchors, but what does it mean to accommodate an anchor? The anchors we've encountered so far have been based on perceptual acquaintance with an external object. Merely hearing someone use a proper name doesn't put us in such a direct relation of acquaintance with the bearer of that name, so we cannot just accommodate a perceptual anchor. Still, following Kripke (1980), hearing someone use a name does, typically, put us in a position to refer rigidly to that name's bearer and form singular attitudes about her. The acquaintance relation that affords this type of reference is an indirect one, it connects the current use of the name to its actual bearer via a causal chain of communication.
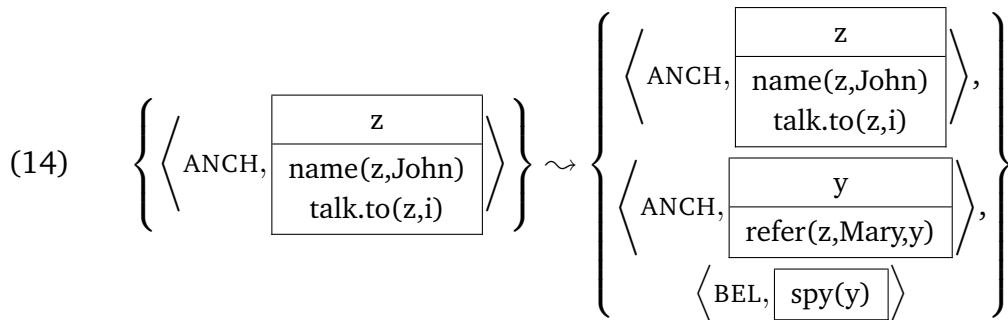
---

expressed) may be appropriate. Here I focus on the highly idealized situation where the hearer simply accepts whatever the speaker asserts.

Kamp (2015) brings Kripke's analysis of names to his cognitive framework by means of *vicarious anchors*.

> A vicarious anchor is established by some agent H who is witness to an act of reference by another agent S, and who, on the basis of this, establishes an entity representation R for the referent of that act. The vicarious internal anchor of that representation is the mark of this referential intention on the part of H and it is what makes [the vicarious anchor] into a representation of that referent. (Kamp 2015: 283–284)

A vicarious anchor is like a regular perceptual anchor in that it signals a referential intention. It differs from a perceptual anchor in that it doesn't directly refer to its source, but rather it defers the interpretation to another agent. Vicarious anchoring thus allows the agent to have singular attitudes about some individual she has no direct perceptual acquaintance with, other than hearing someone use their name.

Formally, we can capture vicarious anchoring in ADT by adding a condition of the form 'refer(x,y,z)' (speaker x used expression y to refer to z) to an internal anchor.[17] To simplify notation I'll assume that 'refer(x,y,z)' entails 'name(z,y)'. Applied to our example: Sue has no relevant anchor for anyone named 'Mary' when she hears John say "Mary is a spy", so she accommodates a vicarious anchor to refer to Mary via John. The input and output mental states of this interpretation are as follows:

$$(14) \quad \left\{ \left\langle \text{ANCH,} \begin{array}{|c|} \hline z \\ \hline \text{name(z,John)} \\ \text{talk.to(z,i)} \\ \hline \end{array} \right\rangle \right\} \rightsquigarrow \left\{ \begin{array}{l} \left\langle \text{ANCH,} \begin{array}{|c|} \hline z \\ \hline \text{name(z,John)} \\ \text{talk.to(z,i)} \\ \hline \end{array} \right\rangle, \\ \left\langle \text{ANCH,} \begin{array}{|c|} \hline y \\ \hline \text{refer(z,Mary,y)} \\ \hline \end{array} \right\rangle, \\ \left\langle \text{BEL,} \begin{array}{|c|} \hline \text{spy(y)} \\ \hline \end{array} \right\rangle \end{array} \right\}$$

The recursive application of vicarious anchoring allows a reconstruction of Kripke's analysis of reference via causal–historical reference chains. I use the name *Aristotle* referentially through a vicarious anchor that defers my referential intention to some high school teacher's usage, which itself refers

---

17 For details on the metalinguistic 'refer' predicate I refer to Maier (2014), where the very same predicate plays a key role in the analysis of mixed quotation (showing the close connection between naming and quoting, taken to its extreme by Shan (2007), who reduces names to mixed quotations).

via a vicarious anchor deferring to someone else's earlier usage and so on all the way back to someone who has a direct acquaintance link to Aristotle.[18] We'll return to vicarious anchoring when we consider how to accommodate fictional names in section 8.

Summing up, our psychologistic semantics includes a uniform presuppositional analysis of definites. Among other things this provides us with a powerful semantic analysis of proper names. Referential interpretations of a proper name arise in two different ways: either (i) the hearer binds the name presupposition to an already established anchored discourse referent, thereby connecting the incoming occurrence of the name to one of her independently referential mental files, or (ii) she accommodates a new vicarious anchor, thereby effectively connecting her referential intention to that of the speaker who uttered the name.[19]

6 Names in fiction

In this section I first implement Walton's (1990) account of fiction as prescriptions to imagine within our psychologistic semantic framework. Our presuppositional analysis of proper names brings out a fundamental dilemma for the interpretation of unfamiliar fictional proper names: do we accommodate them globally, as vicarious but faulty anchors, or locally, inside the fiction-induced imagination, leading to a descriptive rather than strictly referential interpretation?

6.1 Interpreting fiction as imagination updates

We've seen above that in the context of a cooperative information exchange between speaker and hearer, interpreting an assertion means adding its preliminary DRS representation to a belief compartment in the ADS representing the hearer's mental state. After resolving all presuppositions from that preliminary DRS we then end up with an updated mental state representation in which the hearer has acquired the belief that what the speaker said is true. Of course, in the real world the hearer may have reason to doubt that what the

---

18 For this reason, something like vicarious anchoring is a central ingredient of all current mental file frameworks in philosophy. Thus, Perry (2001) talks of notion networks, and Recanati (2012) of indexed mental files.

19 Following Geurts (1997), various attested non-referential interpretations will arise if the name presupposition is bound or accommodated at some local or intermediate DRS embedding level, as in *If a child is christened 'Bambi' they will sue Bambi's parents* or *If presidents were elected alphabetically, Aaron Aardvark could be president*, respectively (cf. (25) below).

speaker says is true and for that reason ignore it, mark it as possibly true, or as something that the speaker believes, or something else entirely. Nonetheless, the norm for an assertion (in a cooperative information exchange) is to convey true information and hence the pragmatic success conditions for an assertion include that the hearer update her belief with the asserted content.

We can extend this normative picture to other speech acts, e.g. the successful interpretation of a command like "Go home!" involves the hearer adding the preliminary DRS representation of its propositional content ($\approx$ the addressee goes home) to an intention compartment of her ADS. In this vein we can now naturally understand Walton's analysis of fictional statements as prescriptions to imagine: In the context of a fictional narrative, successfully interpreting a simple indicative statement means adding its preliminary DRS to an imagination compartment.

Consider first a simple example. Say, I open a book of fairy tales and come across the following opening sentence.

(15)   Once upon a time there was a princess named Isabella.

The first step in the interpretation process is always to parse the sentence and construct a preliminary representation of its semantic content.[20]

(16)

y	t
$t<n$ live(y,t)	
princess(y)	
name(y,Isabella)	

Next, I update my mental state with (16). I started with a mental state containing at least an anchor for the book I'm holding, and, say, the *de re* belief that it contains fairy tales. Since, I thus take the book to contain fiction rather than factual statements, interpreting (15) consists not in updating my beliefs with its semantic content, (16), but in engaging in an act of imagination based on that semantic content. Concretely, my updated mental state will look like this:[21]

---

20 n is the temporal counterpart of i, a special indexical discourse referent denoting the present.
21 For the sake of simplicity, I do not explicitly represent the dependency between the book and the imagination it induces. We could easily add an extra parameter with the label IMG to keep track of this dependency. Concretely, an ADS component $\langle$IMG, φ,$x\rangle$ would indicate that the subject imagines that φ, based on reading book x.

(17)
$$\left\{ \left\langle \text{ANCH}, \begin{array}{|c|} \hline x \\ \hline \text{book(x)} \quad \text{read(i,x)} \\ \text{fairytales(x)} \\ \hline \end{array} \right\rangle, \left\langle \text{IMG}, \begin{array}{|cc|} \hline y & t \\ \hline t<n & \text{live(y,t)} \\ \multicolumn{2}{c}{\text{princess(y)}} \\ \multicolumn{2}{c}{\text{name(y,Isabella)}} \\ \hline \end{array} \right\rangle \right\}$$

Since in this case there are no presuppositions to resolve this concludes the interpretation of the opening sentence. The output ADS describes the reader as someone who is reading a book of fairy tales and imagines that there is a princess named 'Isabella', who lived some time in the past.[22]

So far so good, but what we're really interested in is fictional names, and that first sentence only introduced a name by mentioning it. Say, the next sentence uses that same name:

(18)   Isabella lived in a castle.

Let's assume that the past tense in *lived*, not bound by a temporal quantifier like *once upon a time* in (15), triggers a temporal presupposition, looking for a salient time before n. The proper name also triggers a presupposition, looking for a salient individual named 'Isabella'. Adding the preliminary DRS for (18) to the imagination component of (17), and binding the presuppositions gives the following resolution:

(19)

$$\left\{ \left\langle \text{IMG}, \begin{array}{|ccc|} \hline \multicolumn{3}{|c|}{\vdots} \\ \hline y & t & y' \\ \hline \multicolumn{3}{|c|}{t<n \quad \text{live(y,t)}} \\ \multicolumn{3}{|c|}{\text{princess(y)}} \\ \multicolumn{3}{|c|}{\text{name(y,Isabella)}} \\ \multicolumn{3}{|c|}{\text{live.in(x',y',t')} \quad \text{castle(y')}} \\ \hline \end{array} \right\rangle \right\} \rightsquigarrow \left\{ \left\langle \text{IMG}, \begin{array}{|ccc|} \hline \multicolumn{3}{|c|}{\vdots} \\ \hline y & t & y' \\ \hline \multicolumn{3}{|c|}{t<n \quad \text{live(y,t)}} \\ \multicolumn{3}{|c|}{\text{princess(y)}} \\ \multicolumn{3}{|c|}{\text{name(y,Isabella)}} \\ \multicolumn{3}{|c|}{\text{live.in(y,y',t)} \quad \text{castle(y')}} \\ \hline \end{array} \right\rangle \right\}$$

with the presupposed material $\begin{array}{cc} x' & t' \\ \text{name(x',Isabella)} & t'<n \end{array}$

In the reader's imagination there now exist a princess and a castle and a time before now, such that the princess lived, at that time, in the castle.

As I read on, every sentence which I consider part of the same narrative is interpreted as an update on this IMG-labeled DRS, K_{IMG}. We've seen how

---

22 It may be more accurate to analyze n here as the time of narration, which need not coincide with either the time of the production of the narrative, or of its interpretation by the reader.

indefinites like *a castle* add new discourse referents to the universe of K_{IMG}, predicates like *lived in* add conditions to K_{IMG}, and presuppositions are bound by discourse referents previously introduced into K_{IMG}.

But what if a presupposition does not find a suitable antecedent within K_{IMG}? The general resolution algorithm predicts that proper names (and other definites) can be bound by discourse referents in other boxes, including anchors, as demonstrated in example (12)–(13) in section 5.1 above. In fact, when I introduced the presuppositional analysis of names there I suggested that binding to anchors should be the default behavior for names, as it is only that option that leads to a referential reading. But should this also apply to names in fiction?

6.2 Non-fictional names in fiction

To explore the projection behavior of fictional names consider first the case of a fictional narrative referring to some familiar historical figure or place. For instance, take a sentence like (20) from *War and Peace*.

(20)   "Fine men!" remarked Napoleon, looking at a dead Russian grenadier, who, with his face buried in the ground and a blackened nape, lay on his stomach with an already stiffened arm flung wide.

Reading this sentence I imagine Napoleon looking at a dead soldier. Given that I know *War and Peace* to be historical fiction rather than a factual description of the Napoleonic war, I am not committed to believing that Napoleon really did say such words while looking at a dead grenadier. Nonetheless, I do take Tolstoy's use of the name *Napoleon* to denote the actual Napoleon, i.e., my imagination is *de re* about Napoleon.

Our current framework captures this *de re* interpretation by letting the proper name presupposition project out of the imagination DRS and bind to my pre-existing internal anchor representing Napoleon, leading to an ADS output like the following:

$$(21) \quad \left\{ \left\langle \text{ANCH}, \boxed{\begin{array}{c} x \\ \hline \text{name(x,Napoleon)} \\ \text{emperor(x)} \end{array}} \right\rangle, \left\langle \text{IMG}, \boxed{\begin{array}{c} y \\ \hline \text{say(x,'fine men!')} \\ \text{grenadier(y)} \quad \text{look.at(x,y)} \\ \dots \end{array}} \right\rangle \right\}$$

Following the general resolution preferences for names laid out in section 5.1, such a global resolution should be the preferred outcome whenever we

encounter a name for something we are already acquainted with outside the fiction.

6.3   The dilemma: fictional anchoring or existential interpretation

In the case discussed above, the name 'Napoleon', though occurring in a fictional text is not a fictional name. It's just a regular, referentially used proper name picking out someone with whom the reader was probably already acquainted. The question is now, what happens when a genuinely fictional name, i.e., a name of a fictional or imagined character, is not bound by an explicit existential quantifier, as in the formulaic fairy tale illustrated in (15)–(18). For instance, consider a novel that starts *in medias res*, using an unfamiliar proper name in the opening sentence.

(22)   Barry Fairbrother did not want to go out to dinner.[23]

As this is the first line of the novel, there is no antecedent in the dedicated imagination DRS assigned to the interpretation of this narrative, nor is there one in any other attitude or anchor, let's assume. So, where do we accommodate the proper name presupposition? Do we globally accommodate an internal – but non-referential – anchor, or do we locally accommodate the existence of an individual named 'Barry Fairbrother' inside the imagination DRS, effectively leading to a descriptive interpretation of the name? Both horns of this dilemma have apparent advantages and disadvantages, and, as I will demonstrate below, the choice between them has important consequences for solving some philosophical puzzles about fictional names, including the semantic paradox of fictional names that we started out with in section 1. In the following two sections we carefully explore each option, before eventually settling on the second.

7   Horn I: fictional anchors

Kamp and others working in closely related mental files frameworks have pursued the first option: accommodating a global anchor for the fictional character. In 7.2 I present a strong argument for this position, but in 7.3 I reject it on the grounds that it requires not just faulty but intentionally faulty anchors.

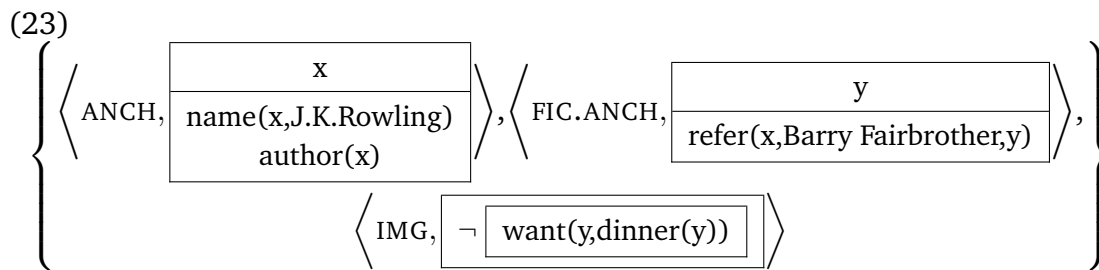23 *The Casual Vacancy*, by J.K. Rowling (Little, Brown Book Club, 2012).

7.1 Networking with vicarious fictional anchors

Kamp's (2015) starting point is to treat reference to fictional entities exactly like reference to real entities, viz., as mediated by entity representations ($\approx$ internal anchors).

> entity libraries do not just consist of representations that stand or purport to stand for real entities, but also entity representations 'of fictional entities'. We know of such entity representations in our libraries that they do not stand for real entities – that they are 'make-believe' so to speak. But they nevertheless function largely like the entity representations that we do take to stand for entities that have an independent identity. (Kamp 2015: 307–308)

Since the difference between fictional internal anchors and non-fictional ones is cognitively significant – the reader is aware that, unlike Napoleon, Frodo and Fairbrother do not really exist – Kamp must add some kind of formal marking of the distinction between regular internal anchors and fictional ones. We'll just introduce a new mode indicator FIC.ANCH, alongside ANCH. Applying Kamp's suggestion to the example at hand means that the reader of (22) globally accommodates a fictional anchor representing Barry Fairbrother.

As argued in section 5.2 above, accommodation of a name presupposition generally leads to a vicarious anchor. If fictional names indeed function like regular names we'd expect accommodation of fictional names to involve the introduction of a fictional anchor vicariously linked to the producer of the name, i.e. the author or storyteller. Applied to our example, when I read the Fairbrother sentence at the start of the Rowling book I accommodate a vicarious fictional anchor representing the character named 'Barry Fairbrother', linking it vicariously to J.K. Rowling.[24]

(23)

$$
\left\{
\left\langle \text{ANCH}, \begin{array}{|c|} \hline x \\ \hline \text{name(x,J.K.Rowling)} \\ \text{author(x)} \\ \hline \end{array} \right\rangle,
\left\langle \text{FIC.ANCH}, \begin{array}{|c|} \hline y \\ \hline \text{refer(x,Barry Fairbrother,y)} \\ \hline \end{array} \right\rangle,
\left\langle \text{IMG}, \neg\, \boxed{\text{want(y,dinner(y))}} \right\rangle
\right\}
$$

24 We could add a fourth argument to the 'refer' relation to restrict it to uses of the name in the relevant work of fiction, i.e., refer(x,y,z,w) $\approx$ author x uses name y to refer to character z in book w.

7.2 Counterfictional imagination

Kamp's position, as reconstructed above, resonates with much of the recent philosophical literature on fictional names. For instance, Recanati (2012) associates empty and fictional names with 'unloaded indexed files', i.e., vicarious mental files which are not internally equated with ('linked to') regular acquaintance-based mental files within the agent's own mental state. Second, Friend (2011) analyzes fictional name reference as participation in an intersubjective 'notion network' (following the terminology of Perry 2001). And finally, Salis (2013) defends a variation based on Sainsbury's (2005) 'empty name using practices'.[25]

In all these approaches, fictional names are treated as directly referring expressions, whose interpretation leads to *de re* attitudes. Friend (2011) provides a compelling argument for this position, by extending Kripke's argument concerning the reference of names in counterfactuals to the fictional domain. Reading Kafka's *Metamorphosis* I imagine that Gregor Samsa turned into a beetle, but, at the same time, . . .

> . . . I might imagine what the Samsa family's life would have been like had Gregor never changed into a vermin. Even though I imagine contrary to what Kafka's story prescribes – thinking of Gregor in ways contrary to the fictional descriptions – I continue to imagine about the same character. (Friend 2011: 188)

Following Kripke's arguments further, I can even imagine that Gregor's parents decided against the name 'Gregor' and instead named him 'Josef'. Intuitively, such counterfictional imaginations are nonetheless about the Gregor Samsa that Kafka wrote about. It follows that the fictional name indeed behaves like a rigid designator, and the imaginations behave like *de re* attitudes about the referent.

The fictional anchoring approach is ideally suited to account for these intuitions of rigidity in counterfictional contexts. Concretely, with a fictional anchor we can represent any number of different *de re* imaginations about Samsa in addition to the imagination directly prescribed by the book:

---

25 One desirable feature of these analyses is that they can cash out the apparent intersubjectivity of fictional characters in terms of the intersubjective networks of vicariously linked mental files. The difficulty, and the point where these views start to diverge, is what happens at the 'root' of such a network, and, consequently, what the fictional name/file actually refers to. I refrain from discussing the various proposals in more detail here, because, ultimately, they all fall prey to the fundamental objection raised in 7.3 below.

(24)
$$\left\{ \begin{array}{l} \left\langle \text{ANCH,} \begin{array}{|c|} \hline x \\ \hline \text{name(x,Kafka)} \\ \text{author(x)} \\ \hline \end{array} \right\rangle, \left\langle \text{FIC.ANCH,} \begin{array}{|c|} \hline y \\ \hline \text{refer(x,Gregor Samsa,y)} \\ \hline \end{array} \right\rangle, \\ \\ \left\langle \text{IMG,} \begin{array}{|c|} \hline \text{turn.into.beetle(y)} \\ \hline \end{array} \right\rangle, \left\langle \text{IMG,} \begin{array}{|c|} \hline \text{name(y,Josef)} \\ \text{turn.into.horse(y)} \\ \hline \end{array} \right\rangle \end{array} \right\}$$

7.3 Against fictional anchors

My objection to the fictional anchoring approach is its reliance on intentionally non-referential anchors, i.e. anchors that the subject herself knows do not have referents. How do we model-theoretically interpret a fictional internal anchor in such a way that it doesn't entail that the agent believes that its descriptive content is satisfied?

Just to be clear, the problem is not just that of internal anchors lacking external anchors. Human perception is faulty: I can think I'm seeing John but actually it turns out it was his twin brother Mark, or a hallucination. Faulty perception leads to faulty anchors, i.e., internal anchors without corresponding external ones. With vicarious anchoring I'm perhaps even more prone to faulty anchoring. Reading a 19th century astronomy article describing sightings of Vulcan, the hypothetical planet causing peculiarities observed in the orbit of Mercury, I may form a vicarious anchor intending to refer to the planet the author referred to. But since that planet doesn't actually exist, both the author's anchor and mine are then faulty. Crucially, with faulty anchors, both perceptual and vicarious, the subject herself takes the anchor to be grounded in reality. This referential commitment makes thought involving such anchors 'formally *de re*', meaning that, for all the subject knows, her anchored thought is a singular thought about the individual described in the internal anchor.

In line with the idea of referential commitment, Kamp et al. (2003) semantically interpret the (narrow, psychological) content of anchors as beliefs. For the interpretation of the ADS in (24) that means the agent is committed to believing there exists a book she's reading, and an author named 'Kafka' who wrote that book. If it turns out she's actually reading some Kafka-inspired fan fiction (or that she hallucinated the whole book reading episode) the book anchor would be faulty. Consequently, her thoughts about the book would express no wide semantic content (or at least a different content than she thinks). But as far as her narrow, internal psychological state

is concerned that is irrelevant: all that we need for (24) to be psychologically correct is that she believes the content of the anchors to be satisfied.

What distinguishes fictional anchors from faulty anchors is that in addition to not referring they are not even intended or assumed to refer. As discussed in 1, a Tolkien reader is well aware that Frodo doesn't exist, so what does it mean to have an internal anchor describing its referent as a hobbit named 'Frodo'? Kamp doesn't explicitly address how fictional anchors should be interpreted semantically, but it is clear that we can't interpret them as beliefs. So we'll have to find another attitude to interpret fictional anchors.

Perhaps acceptance, in the sense of van Fraassen (1980) or Stalnaker (1984), comes close.[26] I can choose to accept something in order to make scientific progress, or just for the sake of argument, without actually believing it to be true. Nonetheless, the acceptability of negative existentials (*Frodo doesn't exist*) and other metafictional statements (*Frodo is fictional*) seems incompatible with a fictional anchor indicating that the agent accepts that Frodo really does exist.

Looking for a neutral mode of attitude that doesn't entail any referential commitment, we quickly end up with imagination or pretense as a plausible semantic interpretation for fictional anchors. Based on reading the fiction, I imagine that a hobbit named 'Frodo' exists. But then the *de re* account collapses. We'd effectively interpret *Frodo is a hobbit* as an imagination referentially dependent on another imagination. What then is the point of accommodating the name outside of the original imagination box in the first place? We might as well leave everything in the original fiction-induced imagination DRS. As it happens, that is precisely the option identified as the second horn of the dilemma in section 6.3 above. I will defend this approach in the following section.

Summing up: interpreting fictional names via internal anchors, following suggestions from Kamp, Friend, Recanati, and other contemporary philosophers, is attractive because it unifies the interpretation procedures for fictional and regular names. Moreover, it seems to offer a way to capture the apparent intentionality of fictional name reference, as brought out by counterfictional imagination scenarios. However, readers do not commit to the existence of fictional characters, as indicated by the acceptability of metafictional statements like *Frodo is fictional* or *Frodo doesn't exist*. This means that fictional anchors would differ essentially from regular anchors, including faulty ones, in that they carry no existential commitment, and when we get to the model-theoretic interpretation this becomes quite problematic.

---

26 Cf. Sainsbury (2011) for a proposal to analyze fiction interpretation in terms of acceptance.

8 Horn II: existential imagination

Let's return to the Fairbrother sentence and our dilemma. We had an unfamiliar name in a fictional context and the question was how to accommodate the presupposition. In the previous section we considered the option of introducing a global vicarious anchor for referring to the fictional character and found it lacking. So now let's explore the option of a local accommodation.

8.1 Local accommodation

The first thing to note is that local accommodation is a standard option provided by the general presupposition resolution algorithm, so we don't have to treat fictional names differently from regular names to allow it. In fact, since global accommodation is ruled out on the basis of the considerations above (i.e., global anchoring entails existential commitment), van der Sandt's (1992) algorithm predicts that local accommodation would be the preferred resolution option to consider.

Moreover, names in other contexts are already known to allow non-global accommodation, as demonstrated by Bach's (1987) 'Aardvark' scenario:

(25)   If presidents were elected alphabetically, Aaron Aardvark could be our next president.

This counterfactual has a reading that is not about an actual person named 'Aaron Aardvark' but just means that in the relevant hypothetical situations with alphabetical elections, it could well happen that there is someone eligible with that name who would on that basis become president. As Geurts (1997) points out, this reading corresponds precisely to the local accommodation of the descriptive name presupposition inside the counterfactual consequent.

Applied to our leading example local accommodation leads to the following output ADS (ignoring the perceptual anchor for the book and the (vicarious) anchor for J.K. Rowling):

(26)   Barry Fairbrother did not want to go out to dinner. $\bigl[=(22)\bigr]$

$$
\left\{ \left\langle \text{IMG}, \begin{array}{|c|} \hline y \\ \hline \text{name}(y,\text{Barry Fairbrother}) \\ \hline \neg\ \boxed{\text{want}(y,\text{dinner}(y))} \\ \hline \end{array} \right\rangle \right\}
$$

On this reading, interpreting the sentence effectively causes the reader to imagine that there exists somebody named 'Barry Fairbrother' who does not want to go out to dinner.

This output is completely in line with that of the simple fairy tale scenario from section 6.1 in which the fictional name was first existentially introduced by an overt indefinite construction (*there was a princess named Isabella*. The fictional anchoring proposal, by contrast, treats reference to Fairbrother in the novel and to Isabella in the fairy tale as fundamentally different. While Fairbrother is represented by an anchor in the reader's mental state, Isabella was introduced locally by the straightforward DRT interpretation of the indefinite description. However, intuitively both names seem to fulfill the exact same functions in the continuations of their respective stories, viz., referring to fictional characters. In particular, both license counterfictional imagination, i.e., I can imagine that Fairbrother *did* want to go out to dinner, but also that princess Isabella lived in exile on a farm rather than a castle. In light of this observation, a defender of the fictional anchoring approach would thus need to invent an additional semantic mechanism whereby the indefinite construction in the fairy tale licenses the creation of a vicarious fictional anchor.

On the current approach there is a different split, viz., between reference to fictional entities like Isabella or Fairbrother on the one hand, and reference to (presumably) real entities like Napoleon or the glass of water I'm holding on the other. The latter type of reference involves (vicarious) global anchors, possibly faulty, but with referential commitment, while the former involves discourse referents existentially introduced inside an imagination compartment. The agent merely imagines that there exists a princess named 'Isabella' without believing (or accepting) her real existence. However, this split is not a lexical ambiguity. In the lexicon, every proper name is analyzed as a presupposition trigger. The split corresponds to a contextually driven difference in resolution of the presupposition.

8.2   Counterfictional imagination revisited

In a sense, the representations for fictional names that we end up with are akin to those postulated by the classic descriptivist approaches to fictional names (cf. Russell 1905, Quine 1948, Kaplan 1973, and, more recently, Currie 1990) that analyze *Frodo is a hobbit* as, roughly, *(in the fiction,) there exists someone named 'Frodo' who is a hobbit*. Friend's counterfictional imagination argument explicitly targets such descriptive approaches, and therefore, potentially, the current approach. As described in section 7.2, the basic idea was to apply

Kripke's test for rigidity in counterfactuals to the fictional domain in order to show that fictional names really should be analyzed as referential terms. Concretely, on the basis of reading Kafka I imagine that Gregor Samsa turned into a beetle, but then go on to imagine what it would be like if he, Gregor, had been named 'Josef' and turned into a horse instead of a beetle. How can both imaginations be about the same individual, Gregor Samsa, unless they are both *de re* attitudes about a fictional character?

On the current approach, what I imagine on the basis of the novel is the existential proposition that there exists someone named 'Gregor Samsa' who turned into a beetle. There is no anchor, my imagination is not singular/*de re*. The crucial intuition of the puzzle is that the second imagination is about the same individual, the fictional Gregor Samsa. On the fictional anchoring approach we would cash this out by having both imaginations depend on the same anchor. But in the ADT framework two distinct attitudes can also share discourse referents directly, without the mediation of anchors. In other words, the following ADS is well-formed and interpretable in the formal framework laid out by Kamp et al. (2003) and reconstructed in section 4 and the appendix:

$$
(27) \quad \left\{ \left\langle \text{IMG,} \begin{array}{|c|} \hline y \\ \hline \text{name(y,Gregor Samsa)} \\ \text{turn.into.beetle(y)} \\ \hline \end{array} \right\rangle, \left\langle \text{IMG,} \begin{array}{|c|} \hline \text{name(y,Josef)} \\ \text{turn.into.horse(y)} \\ \hline \end{array} \right\rangle \right\}
$$

Instead of assuming an anchor for a fictional Samsa on which both imaginations depend, we have one existential imagination, introducing a discourse referent y for the imagined Samsa, and another imagination that uses the same y and is thereby dependent on the first imagination.

Such cross-attitudinal dependencies are commonplace in ADT. In section 4.2 I analyzed the case of John, who believes that there's a ghost in his attic and wants it to be quiet as a desire referentially dependent on a *de dicto* belief. Belief dependence is surely the most salient form of cross-attitudinal dependence. In fact, since internal anchors are ultimately also interpreted as beliefs, all *de re* attitudes are belief dependent. What's not so widely discussed or acknowledged is that attitudes can depend on other attitudes than beliefs. For instance, I may want to buy a new smartphone in 2018 and imagine it having a flexible transparent screen – an imagination dependent on a *de dicto* desire. Or, I can hope that the hobbit I am reading about – represented as a discourse referent introduced in an imagination DRS – will escape from the armies of Mordor and return safely to the Shire – a hope dependent on a *de dicto* imagination. ADT allows us to represent and interpret all such

configurations: a discourse referent introduced in one attitude DRS is in principle accessible for all other attitude DRSs within the same ADS.[27]

Another feature that we're exploiting in (27), and in fact already in (24), is that we can have multiple distinct attitudes of the same type. We've already seen ADSs with multiple anchors, of course, but, as far as the narrow semantic interpretation is concerned, we could, in principle, just merge all anchors and beliefs together into a single belief DRS.[28] By contrast, in Friend's scenario we have two distinct imaginations that are not so reducible, because their contents are mutually inconsistent. Our ADT semantics, as formalized in the appendix, allows for this possibility. A complex mental state does not contain just a single doxastic, buletic, and imagination state, but has a structure with multiple beliefs, desires and imaginations, which may well be mutually inconsistent.

In sum, to capture the attested variety of attitudinal dependence in human mental states, an ADS may contain any number of labeled attitude DRSs, dependent and independent ones. The only well-formedness constraints are that every free variable in an attitude DRS is introduced somewhere in the main universe of some other attitude within the ADS, and that there are no circular chains of dependencies. Thus we may see complex networks where, say, a desire depends on multiple other desires, imaginations and beliefs,

---

27 Kamp et al. (2003) does consider imposing some global, structural restrictions on attitudinal dependence. For instance, it seems *prima facie* reasonable to stipulate that any attitude may depend on a belief or anchor, but, conversely, an anchor (or a belief) may not depend on, say, a desire. In previous work, following Heim (1992) and others, I postulated such an asymmetry, treating belief and anchors together as a basic background attitude that all non-doxastic attitudes can depend on (Maier 2015a). This not only served to keep the formalism simple, but also to explain some linguistic asymmetries in parasitic report sequences:

(i)   a.   John believes that Mary will come and he hopes Sue will come too.
            b.   *John hopes Mary will come and he believes Sue will come too.

On the other hand, examples like (9) (*Alice fears that there is a squirrel in her kitchen cabinets. She hopes to trap it alive*) become problematic if we build in this asymmetry from the start.

At this point I hold that, conceptually, anchors should not be allowed to depend on desires or imaginations, but I don't see strong reasons to categorically ban other types of cross-attitudinal dependence. In fact, my analysis of metafictional statements below relies on allowing imagination-dependent beliefs. I leave the formulation of potential constraints on cross-attitudinal dependencies within an ADS (and, thus, on cross-attitudinal anaphora resolution in ascriptions) for future research.

28 A possible benefit of representing beliefs as distinct DRSs is that that allows us to extend the framework by associating with each belief the source from which it derives, and a degree of confidence. Also, keeping distinct beliefs compartmentalized could eventually help us get around some problems of logical omniscience and belief revision. I will not pursue any of these orthogonal topics here.

and those in turn depend on yet other beliefs and anchors. As pointed out in section 4.5 it is precisely this generality that makes ADT's semantics so complicated.

8.3   From counterfactual to counterfictional: Remarks on the semantics of parasitic imagination

To oversee the full impact of the proposal in (27) we have to really delve into the semantics of dependent attitudes. In the semantics laid out in the appendix an attitude like imagination is characterized in the model by a two-dimensional object that specifies the agent's imagination alternatives relative to a certain background attitude. Thus, (27) means that the agent imagines someone named 'Gregor Samsa' who turned into a beetle, and also imagines, relative to that imagination as background, that he turned into a horse and was named 'Josef'. The main challenge is to justify conceptually (and formally) the new primitive notion of 'relative imagination'.

The basic idea is derived from Kaplan's (1989) two-dimensional semantics of statements: just like we can't determine the content of *I am a fool* independent of a context of utterance, so we can't always determine the content of an attitude without another attitude content as background. Technically, just as Kaplan's context parameter fixes the reference of *I* to the current speaker, while keeping the property of being the speaker out of the at-issue content, so the attitudinal background fixes the reference of the anaphoric dependencies (free variables) in the at-issue attitude without merging the background content and at-issue content into a single imagination.

One key difference between Kaplan's reference fixing and our mental reference fixing is that the attitudinal background is not just a single context parameter, but rather a set of contexts (or actually an information state, to be precise), which complicates the reference fixing considerably. I pursue here an insight from Ninan (2008), who analyses belief-dependent imagination in terms of a two-dimensional attitude *Img*, i.e., a function from doxastic alternatives to sets of possible worlds.[29] Intuitively, *Img* maps any doxastic alternative c of an agent to the content of her imagination relative to a hypothetically fixed doxastic background $\{c\}$. In other words, $w \in Img(c)$ iff w is compatible with what the agent would imagine if she were a maximally opinionated agent who believes she inhabits context c. We can then say that

---

29 Ninan (2008, 2012) provides an alternative implementation using multi-centered worlds. Zeevat (1999) and Yanovich (2011) capture this type of background fixing in terms of mapping the variable to an individual concept. In principle, either of these mechanisms could probably be adapted to our needs.

an agent imagines proposition p, relative to belief state q, iff for each doxastic alternative $c \in q$, $Img(c) \subseteq p$.

For our purposes we need to generalize and "dynamify" Ninan's idea: instead of centered worlds or contexts we use possibilities (world–assignment pairs); instead of propositions and functions from contexts to propositions we use information states and functions from possibilities to information states (i.e., two-dimensional information states); and instead of just defining belief-relative imagination, we model every individual attitude in a mental state as a two-dimensional information state. The referential dependency structure of the mental state determines which attitudes are interpreted as backgrounds to which other attitudes within the state. The appendix makes this precise, but let's apply it to the evaluation of (27) to have a concrete example to illustrate what's going on.

We want to know if (27) correctly describes some mental state given by the model. First of all, this requires that we can identify in our mental state two imagination components, i.e. IMG-labeled (two-dimensional) information states, say $\langle \text{IMG}, Q_1 \rangle$ and $\langle \text{IMG}, Q_2 \rangle$, corresponding to the fiction-induced imagination and the counterfictional one, respectively. Let's assume that we have already checked that information state Q_1 indeed entails the fiction-induced imagination DRS K_1 in (27) and focus on the counterfictional one. To verify that the two-dimensional imagination Q_2 entails the counterfictional, dependent imagination DRS K_2 in (27), we have to compute $Q_2(\langle w, f \rangle)$, i.e., what the agent would imagine relative to background possibility $\langle w, f \rangle$, for each possibility $\langle w, f \rangle$ in the background state that this attitude depends on, i.e. Q_1. For each such imagination state $Q_2(\langle w, f \rangle)$ we then have to check that it entails K_2, or more precisely, that it entails the information state expressed by K_2 relative to embedding f (i.e. f is used to fix the reference of the free variable x in K_2). If these straightforward entailments between information states hold (for all background possibilities) then we conclude that Q_2 entails K_2, which concludes the evaluation.
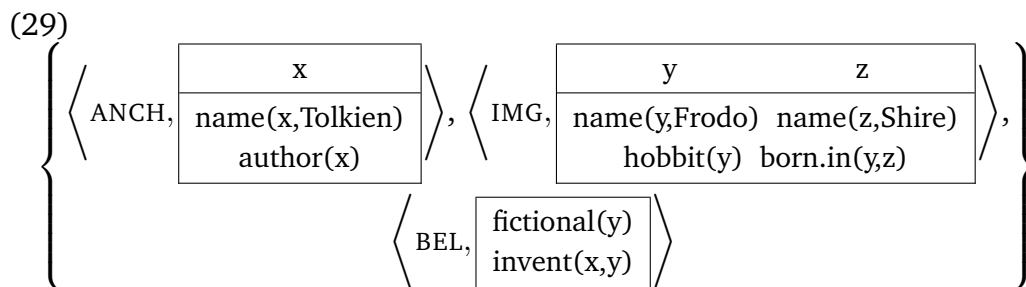
Summing up, we can account for counterfictional imagination by exploiting the idea of referential dependence between attitudes, rather than invoking conceptually dubious fictional anchors. The current proposal shows that the intentionality of fictional names, as brought out by Friend's argument, does not require a kind of anchoring that is faulty by design. All we need is that attitudes can depend on other attitudes within a mental state description. As I will show next, this proposal naturally extends to a solution of the paradox of fiction that we started out with.

9 The paradox of fictional names resolved

In section 1 we described the semantic paradox of fiction as the fact that we can consistently maintain, simultaneously, that Frodo is a flesh and blood creature, born in the Shire, and that Frodo is a fictional creature who doesn't actually exist. On the current approach, the key to resolving the tension between such fictional and metafictional propositions is that the first is something we *imagine*, based on reading the book, while the second is rather something we *believe*. In other words, typical utterances of the corresponding sentences below constitute different speech acts: (28b) is an assertion, a proposal for the interpreter to update her belief state, while (28a) is a fictional statement, a proposal for her to imagine a certain state of affairs.

(28) a. Frodo is a hobbit born in the Shire.
      b. Frodo is a fictional character invented by Tolkien.

As in the counterfictional scenario, the linguistic form of the statements in (28), in particular the use of the name *Frodo*, suggests that they express two propositions directed towards the same object. The fictional anchoring approach takes this suggestion at face value and would analyze the underlying mental state with a single vicarious, fictional Frodo-anchor and two *de re* attitudes referentially dependent on it. However, following my analysis of counterfictional imagination, we don't need such conceptually suspicious fictional anchors to do justice to this intuition of coreference. Instead, (28a) expresses an existential imagination, and (28b) a belief directly dependent on that imagination.

(29)

$$
\left\{
\left\langle \text{ANCH,}
\begin{array}{|c|}
\hline
x \\
\hline
\text{name(x,Tolkien)} \\
\text{author(x)} \\
\hline
\end{array}
\right\rangle,
\left\langle \text{IMG,}
\begin{array}{|c c|}
\hline
y & z \\
\hline
\text{name(y,Frodo)} & \text{name(z,Shire)} \\
\text{hobbit(y)} & \text{born.in(y,z)} \\
\hline
\end{array}
\right\rangle,
\left\langle \text{BEL,}
\begin{array}{|c|}
\hline
\text{fictional(y)} \\
\text{invent(x,y)} \\
\hline
\end{array}
\right\rangle
\right\}
$$

Conceptually, what we see here is that the agent imagines the existence of a hobbit named 'Frodo' and believes, relative to that imagination, that he, the imagined hobbit, is a fictional creature invented by Tolkien.

Instead of postulating reference to fictional entities, via fictional anchors or otherwise, this solution assumes the notion of an imagination-dependent belief. Just as I can have a desire about something that I believe to exist,

or have a (counterfictional) imagination about something I read about in a fiction, I submit that we can also have beliefs about merely imagined entities – and that is precisely what metafictional statements like (28b) express.[30]

In closing, let me point out that this account of metafictional attitudes covers a range of related cases familiar from the philosophical literature, including negative existentials, (30a), and the fact that we can compare fictional characters to real individuals, (30b), and to fictional characters from other fictions, (30c) (cf. Friend 2007, Zalta 2000):

(30)   a.   Frodo doesn't exist
       b.   Frodo is braver than me
       c.   Frodo is braver than Luke Skywalker

On my view, these are all expressions of metafictional attitudes. We've just seen how reading *Lord of the Rings* leads me to imagine the existence of a hobbit named 'Frodo', which introduces a corresponding discourse referent y inside an imagination box. We can use this discourse referent to represent other attitudes referentially dependent on that existential imagination. For instance, (30a) expresses the belief that y does not exist.[31] Similarly, (30b) expresses the belief that y is braver than me, the actual center of consciousness (represented with the special discourse referent i). As for (30c), finally, I assume that watching *Star Wars* leads me to imagine the existence of a brave Jedi named 'Luke', represented as a new discourse referent in an imagination box distinct from the one involving Frodo. Dependent on these two existential

---

30 In the formal system defined in the appendix this means that we have a belief modeled by a (non-vacuously) two-dimensional information state. I should point out that applying the Ninan-inspired intuitive explanation of two-dimensional attitudes is not very helpful in this case: it tells us to fix the agent's information state to one of the possibilities compatible with her imagination, i.e. with the book, and then look at what the agent would believe. As Ninan (2008) himself already points out for two-dimensional imagination, "this 'decision procedure' idea shouldn't be taken too seriously; it's really just a heuristic for getting an intuitive grip on what I think we should regard as a primitive notion in our theory of imagining: the notion of a centered world's being compatible with what someone imagines relative to one of her centered belief worlds." In line with that assessment I think we should regard 2D beliefs as semantic primitives as well.

31 'exist' here is just a regular predicate meaning something like real-world, physical existence. So, in the end, we do need a model-theoretic ontology that includes existing and non-existing entities. The current account, however, avoids the pitfalls of straightforward Meinongian realism identified in section 2. I don't predict that *Sam carried Frodo* entails that Sam carried an abstract or fictional object, and I don't assume any ambiguity between fictional and metafictional name usage, or distinguish different types of predication.

imaginations I may then form new attitudes – imaginations, beliefs, desires, dreams, etc. – using the discourse referents introduced in either.[32]

10 Conclusion

Fictional proper names pose a notoriously difficult puzzle for truth-conditional semantics. Since Sherlock Holmes never existed, *Sherlock Holmes* does not refer and, by compositionality, statements containing the name have no truth value, or even, if we assume that names are rigid designators, express no classical proposition.

To solve this puzzle I proposed a formal semantic approach based on the pragmatic account of fiction as 'prescriptions to imagine' (Walton 1990). The account was couched in a version of the psychologistic semantic framework introduced by Kamp (1990, 2003, 2015). This framework combines an explicit theory of the representation of mental states (ADT), with a model of communication in terms of distinct production and interpretation algorithms that map a portion of a mental state representation to an utterance, and vice versa.

I analyzed the interpretation of fictional statements as dynamic updates on an imagination component of the interpreter's mental state, while plain assertions (including 'metafictional' ones) correspond to updates on a belief component. Moreover, proper names – regular, empty, or fictional – are uniformly analyzed as presupposition triggers.

Departing from Kamp and philosophers in related 'mental file' frameworks, I argued against the interpretation of fictional names via global fictional anchors. Instead, fictional names are interpreted inside the fiction-induced imagination component of the ADS, effectively leading to an existential/descriptive interpretation of the name: reading about Sherlock Holmes leads me to imagine that there exists a detective named 'Sherlock Holmes'.

I demonstrated that the general ADT formalism for the representation of mental states is expressive enough to straightforwardly capture common counterexamples to other approaches, including counterfictional imaginations, transfictional comparisons, negative existentials, and metafictional beliefs, while keeping the fictional proper name represented locally, inside the imagination. The central feature of the ADT formalism that I relied on is the possibility of different attitude components sharing discourse referents

---

32 The formalism for interpreting multiple simultaneous dependencies I develop in Maier (2016) requires that the dependencies themselves are mutually consistent. It is not clear to me what should happen in the case of a metafictional attitude depending on two distinct and incompatible fictions, so I leave this for another occasion.

and thus referentially depending on each other. This allows us to correctly represent metafictional attitudes, such as believing that Holmes is fictional or imagining what would happen if he were a linguist rather than a detective, as attitudes referentially dependent on the fiction-induced imagination that there is a detective named 'Sherlock Holmes'.

Appendix: The syntax and semantics of ADT

(31)   Primitive symbols of DRS language

    a.   A set of predicates $Pred = \{P, Q, \text{walk}, \text{john}, \text{beat}, =, \ldots\}$

    b.   A set of discourse referents $DRef = \{x, y, x_1, \ldots\}$

(32)   Syntax of DRS language

    a.   If $x_1 \ldots x_n$ are discourse referents and P an n-place predicate, then $P(x_1 \ldots x_n)$ is a DRS condition

    b.   If $x_1 \ldots x_n$ are discourse referents and $\psi_1 \ldots \psi_m$ are DRS conditions, then $\langle \{x_1 \ldots x_n\}, \{\psi_1 \ldots \psi_m\} \rangle$ is a DRS (notation: $U(K) = \{x_1 \ldots x_n\}$ and $Con(K) = \{\psi_1 \ldots \psi_m\}$)

    c.   If K, K' are DRSs, then $\neg K$ and $K \to K'$ are DRS conditions

(33)   Free variables of a DRS or DRS condition:

    a. $FV(K) = \bigcup\{FV(\psi) | \psi \in Con(K)\} \setminus U(K)$

    b. $FV(P(x_1 \ldots x_n)) = \{x_1 \ldots x_n\}$

    c. $FV(\neg K) = FV(K)$

    d. $FV(K \to K') = FV(K) \cup (FV(K') \setminus U(K))$

    e. K is proper if $FV(K) = \emptyset$, improper otherwise.

(34)   Intensional model: $M = \langle D, W, \mathbb{I} \rangle$, in which

    a. D is a non-empty domain of individuals

    b. W is a non-empty domain of possible worlds

    c. $\mathbb{I}$ is an interpretation function, $W \times Pred \to \bigcup_{n \in \mathbb{N}} D^n$

(35)   Embeddings:

    a.   an embedding is a partial function from $DRef$ to D

    b. F is the set of all embeddings

    c. g extends f: $f \subseteq_X g := f \subseteq g \wedge Dom(g) = Dom(f) \cup X$

(36)   Embeddings verifying a DRS/condition (relative to a model M, suppressed in notation):

    a. $g \models_w K$ if for all $\psi \in Con(K)$: $g \models_w \psi$

    b. $g \models_w P(x_1, \ldots x_n)$ if $\langle g(x_1) \ldots g(x_n) \rangle \in \mathbb{I}_w(P)$

    c. $g \models_w \neg K$ if there is no $h \supseteq_{U(K)} g$ with $h \models_w K$.

d. $g \models_w K \to K'$ if for all $h \supseteq_{U(K)} g$ with $h \models_w K$, there is an $h' \supseteq_{U(K')}$ h such that $h' \models_w K'$.

(37) Possibilities and information states:

    a. A pair $\langle w, f \rangle \in W \times F$ is a possibility

    b. $s \subset W \times F$ is an information state if there is a $Y \subset DRef$ such that: $\forall \langle w, f \rangle \in s [Dom(f) = Y]$

(38) Information state interpretation of proper DRS K:

    a. $[\![K]\!]^{\mathsf{is}\,f} = \{ \langle w, g \rangle | g \supseteq_{U(K)} f \text{ and } g \models_w K \}$

    b. $[\![K]\!]^{\mathsf{is}} = [\![K]\!]^{\mathsf{is}\,\emptyset}$

(39) Further definitions/notations regarding information states:

    a. **I** is the set of all information states

    b. $\Lambda = \{ \langle w, \emptyset \rangle | w \in W \}$ (the empty information state)

    c. $s \triangleright s'$ ("s contains a least as much information as s'") iff for all $\langle w, f \rangle \in s$ there is a $g \subseteq f$ such that $\langle w, g \rangle \in s'$.

    d. $s \uplus s' = \{ \langle w, f \cup f' \rangle | \langle w, f \rangle \in s \text{ and } \langle w, f' \rangle \in s' \text{ and } f \cup f' \text{ is a function} \}$ (information state merge)

(40) A pre-ADS is a finite set of pairs (labeled attitude DRSs) of the form $\langle l, K \rangle$ in which l is a mode label ($\in \{\text{ANCH}, \text{BEL}, \text{DES}, \text{IMG}\}$) and K a DRS.

(41) Referential dependence inside a pre-ADS **K**:

    a. For any $\langle l, K \rangle, \langle l', K' \rangle \in \mathbf{K}$: $\langle l', K' \rangle$ referentially depends on $\langle l, K \rangle$ ($\langle l, K \rangle \prec_{\mathbf{K}} \langle l', K' \rangle$) iff $FV(K') \cap U(K) \neq \emptyset$

    b. A pre-ADS **K** is well-founded iff $\prec_{\mathbf{K}}$ is well-founded (iff there are no "loops", i.e. no infinite sequences of labeled attitude DRSs $L_0, L_1 \dots \in \mathbf{K}$ with $L_{i+1} \prec_{\mathbf{K}} L_i$).

    c. dependencies of a labeled attitude DRS L inside a pre-ADS **K**: $Deps(L, \mathbf{K}) = \{ L' | L' \prec_{\mathbf{K}} L \} \}$

    d. A pre-ADS **K** is proper iff for all $\langle l, K \rangle \in \mathbf{K}$: $FV(K) \subseteq \bigcup_{\langle l', K' \rangle \prec_{\mathbf{K}} \langle l, K \rangle} U(K')$.

(42) An ADS is a pre-ADS that is proper and well-founded.

We want to semantically interpret these syntactic entities by matching them up with semantic entities, provided by the model, that capture the content of complex mental states in set-theoretic terms. The content of an individual attitude, say a particular belief or desire of an agent, is modeled as a two-dimensional information state (a function from possibilities to information states):

(43)   A 2D information state is a function Q such that:
   a. $Q : W \times F \to \mathbf{I}$.
   b.   If $\langle w', f' \rangle \in Q(\langle w, f \rangle)$ then $f' \supseteq f$.

Generalizing and extending Ninan (2008), the idea behind these 2D information states is that to determine whether $i' \in Q(i)$ we put the agent in the singleton information state $\{i\}$ (by presenting the agent with a lot of information, excluding various other possibilities until only one is left) and ask whether – given that information – the possibility i' is compatible with the agent's Q-attitude (e.g., a specific desire).

Complex mental states are given in the model as sets of such labeled 2D information states, plus a dependency relation, specifying that a certain imagination, say, referentially depends on certain background beliefs and desires. Extending Kamp's ISBAS terminology, I'll call these sets NBAS (Ninan-Based Attitudinal States).

(44)   A pre-NBAS is a pair $\langle \mathbf{A}, \prec^* \rangle$ with:
   a. $\mathbf{A}$ is a set of pairs of the form $\langle l, Q \rangle$ in which l is a mode label and Q a 2D information state
   b. $\prec^*$ is a well-founded relation on $\mathbf{A}$.

(45)   Dependencies of a labeled attitude $\langle l, Q \rangle$ inside a pre-NBAS:
$Deps(\langle l, Q \rangle, \langle \mathbf{A}, \prec^* \rangle) = \{ \langle l', Q' \rangle \in \mathbf{A} | \langle l', Q' \rangle \prec^* \langle l, Q \rangle \}$

By abuse of notation I'll henceforth use metalanguage variables Q, Q', Q_i to denote both the pair consisting of a label plus a 2D information state, and the corresponding plain 2D information states, simply ignoring the label whenever convenient. Similarly, I'll use K, K', K_i to denote both labeled and plain DRSs.

As in the ADS syntax, an NBAS should be well-founded, i.e. the chains of dependencies must be ultimately grounded in independent attitudes. Intuitively, these independent attitudes should correspond to plain information states. Formally, they are still two-dimensional but the first dimension is vacuous.

(46)   A pre-NBAS $\langle \mathbf{A}, \prec^* \rangle$ is an NBAS if it satisfies the following requirement: for every $Q \in \mathbf{A}$: if $Deps(Q, \langle \mathbf{A}, \prec^* \rangle) = \emptyset$, then $Q(\langle w, f \rangle) = Q(\langle w', f' \rangle)$ for all possibilities $\langle w, f \rangle$ and $\langle w', f' \rangle$ with $Dom(f) = Dom(f')$.

The psychologistic variant of the Tarskian definition of truth is a definition of when a certain syntactic formula (i.e., an ADS) correctly captures (part of) an agent's mental state (i.e., the NBAS provided by the model). Roughly, an

ADS captures/describes an NBAS iff every labeled attitude DRS corresponds to a labeled 2D information state of the same mode (requirement (47a)) that entails it. The tricky part is the definition of entailment, which should be specified relative to the attitude's relevant background state within the ADS (i.e., BG in requirement (47c), defined in a separate definition, (48)). The precise definition is a bit more complicated due to the following technicalities: (i) an NBAS may contain many more attitudes than an ADS (hence the g in (47)); (ii) both ADS and NBAS contain discourse referents, but these need not coincide (hence the r in (47)); (iii) the dependency relations in the NBAS must parallel those in the ADS (hence the requirement in (47b)).

(47)　An ADS $\mathbf{K}$ captures an NBAS $\langle \mathbf{A}, \prec^* \rangle$ iff there is a variable renaming function r on $\mathbf{K}$ and a one-to-one function g from $\mathbf{K}$ to $\mathbf{A}$ such that:

　　a.　For every $\langle l, K \rangle \in \mathbf{K}$ there is a 2D information state Q such that $g(\langle l, K \rangle) = \langle l, Q \rangle$ (i.e., g preserves mode labels).

　　b.　For every $K, K' \in \mathbf{K}$: if $K \prec_{\mathbf{K}} K'$ then $g(K) \prec^* g(K')$ (i.e., g preserves dependency structure).

　　c.　For every $K \in \mathbf{K}$, $\langle w, f \rangle \in BG(g(K), \langle \mathbf{A}, \prec^* \rangle): (g(K))(\langle w, f \rangle) \triangleright$ $[\![r(K)]\!]^{\text{is}\,f}$ (i.e., $g(K)$ entails K relative to the relevant background state).

What's left is the definition of the relevant background state for any labeled 2D information state in an NBAS. The idea is that the background of an independent attitude is the empty information state, while the background of a dependent attitude is the sum of the attitudinal states it depends on. However, these attitudinal states may in turn depend on yet other backgrounds, which introduces recursion into the definition.

(48)　$BG(Q, \langle \mathbf{A}, \prec^* \rangle) =$

$$\begin{cases} = \Lambda, & \text{if } Deps(Q, \langle \mathbf{A}, \prec^* \rangle) = \emptyset; \\[2ex] = \bigcup \{ Q_1(i_1) \uplus \ldots \uplus Q_n(i_n) \mid i_1 \in BG(Q_1, \langle \mathbf{A}, \prec^* \rangle) \ldots i_n \in BG(Q_n, \langle \mathbf{A}, \prec^* \rangle) \} \\ \qquad\qquad\qquad\qquad\qquad\qquad , \text{if } Deps(Q, \langle \mathbf{A}, \prec^* \rangle) = \{ Q_1 \ldots Q_n \} \end{cases}$$

References

Anderson, John & Christian Lebiere. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Asher, Nicholas. 1986. Belief in Discourse Representation Theory. *Journal of Philosophical Logic* 15(2). 127–189. http://dx.doi.org/10.1007/BF00305490.

Bach, Kent. 1987. *Thought and Reference*. Oxford: Oxford University Press.

Bauer, Matthias & Sigrid Beck. 2014. On the Meaning of Fictional Texts. In D. Gutzmann, J. Köpping & C. Meier (eds.), *Approaches to Meaning*, 250–275. Leiden: Brill.

Bonomi, Andrea & Sandro Zucchi. 2003. A pragmatic framework for truth in fiction. *Dialectica* 57(2). 103–120. http://dx.doi.org/10.1111/j.1746-8361.2003.tb00259.x.

Brasoveanu, Adrian & Jakub Dotlacil. 2015. Incremental and predictive interpretation. *Semantics and Linguistic Theory (SALT)* 25. 57–81. http://dx.doi.org/10.3765/salt.v25i0.3047.

Cohen, Ariel & Manfred Krifka. 2014. Superlative quantifiers and meta-speech acts. *Linguistics and Philosophy* 37(1). 41–90. http://dx.doi.org/10.1007/s10988-014-9144-x.

Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge: CUP.

Eckardt, Regine. 2014. *The Semantics of Free Indirect Speech. How Texts Let You Read Minds and Eavesdrop*. Leiden: Brill.

Edelberg, Walter. 1992. Intentional Identity and the Attitudes. *Linguistics and Philosophy* 15(6). 561–596.

Elbourne, Paul. 2010. The existence entailments of definite descriptions. *Linguistics and Philosophy* 33(1). 1–10. http://dx.doi.org/10.1007/s10988-010-9072-3.

Farkas, Donka & Kim Bruce. 2009. On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1). 81–118. http://dx.doi.org/10.1093/jos/ffp010.

Fauconnier, Gilles. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: CUP.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon. http://dx.doi.org/10.1093/0198244274.001.0001.

Frank, Anette & Hans Kamp. 1997. On Context Dependence In Modal Constructions. *Semantics and Linguistic Theory (SALT)* 7. 151–168. http://dx.doi.org/10.3765/salt.v7i0.2790.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100(1). 25–50.

Friend, Stacie. 2007. Fictional Characters. *Philosophy Compass* 2(2). 141–156. http://dx.doi.org/10.1111/j.1747-9991.2007.00059.x.

Friend, Stacie. 2011. The great beetle debate: A study in imagining with names. *Philosophical Studies* 153(2). 183–211. http://dx.doi.org/10.1007/s11098-009-9485-4.

Geach, Peter. 1967. Intentional Identity. *The Journal of Philosophy* 64(20). 627–632.

Geurts, Bart. 1997. Good news about the description theory of names. *Journal of Semantics* 14(4). 319–348. http://dx.doi.org/10.1093/jos/14.4.319.

Geurts, Bart. 1999. *Presuppositions and Pronouns*. Amsterdam: Elsevier.

Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14(1). 39–100. http://dx.doi.org/10.1007/BF00628304.

Heim, Irene. 1982. *On the Semantics of Definite and Indefinite Noun Phrases*. Amherst: University of Massachusetts dissertation. http://semanticsarchive.net/Archive/Tk0ZmYyY.

Heim, Irene. 1983. On the projection problem for presuppositions. *West Coast Conference on Formal Linguistics (WCCFL)* 2. 114–126.

Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics* 9(3). 183–221. http://dx.doi.org/10.1093/jos/9.3.183.

Kamp, Hans. 1990. Prolegomena to a Structural Account of Belief and Other Attitudes. In A. Anderson & J. Owens (eds.), *Propositional attitudes: The role of content in logic, language, and mind*, 27–90. Stanford: CSLI.

Kamp, Hans. 2011. Representing De Se Thoughts and their Reports. http://nasslli2012.com/files/kamp_2011.pdf.

Kamp, Hans. 2015. Using Proper Names as Intermediaries between Labelled Entity Representations. *Erkenntnis* 80(2). 263–312. http://dx.doi.org/10.1007/s10670-014-9701-2.

Kamp, Hans, Josef van Genabith & Uwe Reyle. 2003. Discourse Representation Theory. In D. Gabbay & F. Guenthner (eds.), *Handbook of philosophical logic*, vol. 10, 125–394. Heidelberg: Springer. http://dx.doi.org/10.1007/978-94-007-0485-5_3.

Kaplan, David. 1973. Bob and Carol and Ted and Alice. In J. Hintikka, J. Moravcsik & P. Suppes (eds.), *Approaches to natural language*, Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-010-2506-5.

Kaplan, David. 1989. Demonstratives. In J. Almog, J. Perry & H. Wettstein (eds.), *Themes from Kaplan*, 481–614. New York: Oxford University Press.

Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4(2). 169–193.

Kratzer, Angelika. 1981. The notional category of modality. In H. Eikmeyer & H. Rieser (eds.), *Words, Worlds, and Contexts*, 38–74. Berlin: De Gruyter.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Kripke, Saul. 2011. Vacuous names and fictional entities. In *Philosophical troubles*, 52–74. Oxford: Oxford University Press.

Lewis, David. 1978. Truth in fiction. *American Philosophical Quarterly* 15(1). 37–46.

Maier, Emar. 2014. Mixed Quotation: The Grammar of Apparently Transparent Opacity. *Semantics and Pragmatics* 7(7). 1–67. http://dx.doi.org/10.3765/sp.7.7.

Maier, Emar. 2015a. Parasitic attitudes. *Linguistics and Philosophy* 38(3). 205–236. http://dx.doi.org/10.1007/s10988-015-9174-z.

Maier, Emar. 2015b. Reference, Binding, and Presupposition: Three Perspectives on the Semantics of Proper Names. *Erkenntnis* 80(2). 313–333. http://dx.doi.org/10.1007/s10670-014-9702-1.

Maier, Emar. 2016. Referential Dependencies Between Conflicting Attitudes. *Journal of Philosophical Logic* 1–27. http://dx.doi.org/10.1007/s10992-016-9397-7.

Meinong, Alexius. 1904. Über Gegenstandstheorie. In A. Meinong (ed.), *Untersuchungen zur Gegenstandstheorie und Psychologie*, 1–51. Leipzig: Barth.

Ninan, Dilip. 2008. *Imagination, Content, and the Self*: MIT dissertation.

Ninan, Dilip. 2012. Counterfactual attitudes and multi-centered worlds. *Semantics and Pragmatics* 5. http://dx.doi.org/10.3765/sp.5.5.

Perry, John. 1980. A Problem about Continued Belief. *Pacific Philosophical Quarterly* 61. 317–332.

Perry, John. 2001. *Reference and Reflexivity*. Stanford: CSLI.

Quine, Willard V O. 1948. On What There Is. *Review of Metaphysics* 2(1948). 21–38.

Radford, Colin. 1975. How can we be moved by the fate of Anna Karenina? *Proceedings of the Aristotelian Society* 49. 67–93. http://www.jstor.org/stable/4106870.

Recanati, François. 2012. *Mental files*. Oxford: Oxford University Press.

Roberts, Craige. 1987. *Modal subordination, anaphora, and distributivity*. Amherst: UMass dissertation. http://scholarworks.umass.edu/dissertations/AAI8710499/.

Roberts, Craige. 1989. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy* 12(6). 683–721. http://dx.doi.org/10.1007/BF00632602.

van Rooy, Robert. 2000. Anaphoric relations across attitude contexts. In K. von Heusinger & U. Egli (eds.), *Reference and Anaphoric Relations*, 157–181. Dordrecht: Springer.

Russell, Bertrand. 1905. On Denoting. *Mind* 14(56). 479–493. http://www.jstor.org/stable/2248381.

Sæbø, Kjell Johan. 2012. Reports of Specific Indefinites. *Journal of Semantics* 30(3). 267–314. http://dx.doi.org/10.1093/jos/ffs015.

Sainsbury, Mark. 2005. *Reference without Referents*. Oxford: OUP. http://dx.doi.org/10.1093/0199241805.001.0001.

Sainsbury, Mark. 2011. Fiction and Acceptance-Relative Truth, Belief and Assertion. In F. Lihoreau (ed.), *Truth in Fiction*, 38–137. Ontos Verlag.

Salis, Fiora. 2013. Fictional Names and the Problem of Intersubjective Identification. *Dialectica* 67(3). 283–301. http://dx.doi.org/10.1111/1746-8361.12031.

van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377. http://dx.doi.org/10.1093/jos/9.4.333.

Schoubye, Anders. 2013. Ghosts, murderers, and the semantics of descriptions. *Noûs* 47(3). 496–533. http://dx.doi.org/10.1111/j.1468-0068.2011.00836.x.

Searle, John. 1975. The Logical Status of Fictional Discourse. *New Literary History* 6(2). 319–332. http://dx.doi.org/10.2307/468422.

Searle, John. 2001. Modals and Illocutionary Forces – Reply to Zaefferer. *Revue Internationale de Philosophie* 55(216). 286–290. https://www.cairn.info/revue-internationale-de-philosophie-2001-2-page-277.htm.

Shan, Chung-chieh. 2007. Causal reference and inverse scope as mixed quotation. *Amsterdam Colloquium* 16. 199–204. http://www.illc.uva.nl/AC/AC2007/uploaded_files/proceedings-AC07.pdf.

Stalnaker, Robert. 1970. Pragmatics. *Synthese* 22(1-2). 272–289.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge: MIT Press.

Stone, Matthew. 1997. The Anaphoric Parallel between Modality and Tense. http://repository.upenn.edu/ircs_reports/79.

Thomasson, Amie. 1999. *Fiction and Metaphysics*. Cambridge: CUP.

Walton, Kendall. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge: Harvard University Press.

Wechsler, Stephen. 2010. What 'you' and 'I' mean to each other: Person indexicals, self-ascription, and theory of mind. *Language* 86(2). 332–365. http://dx.doi.org/10.1353/lan.0.0220.

Yanovich, Igor. 2011. The problem of counterfactual de re attitudes. *Semantics and Linguistic Theory (SALT)* 21. 56–75. http://elanguage.net/journals/index.php/salt/article/view/21.56.

Zalta, Edward. 1983. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Berlin/Heidelberg: Springer.

Zalta, Edward. 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, 117–147. Stanford: CSLI.

Zeevat, Henk. 1999. Demonstratives in Discourse. *Journal of Semantics* 16(4). 279–313. http://dx.doi.org/10.1093/jos/16.4.279.