More on fictional names and psychologistic semantics: Replies to comments

Emar Maier
University of Groningen

I am extremely grateful for the thoughtful commentaries I received on my paper. They raise a wide variety of issues with my proposal, ranging from metaphysical worries, to psychological and linguistic ones. Moreover they also suggest various promising alternative approaches I would never have considered otherwise.

In this reply I have tried to take a few arguments or observations from each of the contributions and say something about those. In particular, I focus on: (i) the notion of truth, and its relation to Sandro Zucchi's distinction between fictional and metafictive discourse, mirrored by Dilip Ninan's distinction between authorial diktats and fictional reports, and subsumed by Rami and Zimmermann's four distinct uses of fictional statements; (ii) François Recanati's alternative model of the imagination; (iii) Bart Geurts's non-psychologistic approach to fiction in terms of shared commitments; (iv) Dilip Ninan's proposal for a simpler realist semantics for fictional names; and (v) Maria Aloni's remarks on the logic of my mental state model.

My aim has been not to refute the commentators' arguments but to acknowledge the importance of the objections raised and the questions asked. My remarks below should be seen as gestures in the direction in which I want to look for answers. Comprehensive answers will then, hopefully, be appearing in the course of my current research project, over the next 4-5 years. I hope the comments and my responses together will convince readers that the semantics and pragmatics of fictional discourse is an exciting research area, ripe for the picking.

1 Truth in psychologistic semantics

Many commentators (including Rami and Zimmermann, Zucchi, Aloni, and Ninan) point out that my approach doesn't make adequate predictions about truth and falsity of statements containing fictional names, or for any statements really. At first blush, this seems to be related to my going psychologistic.

Truth is by definition relative to the actual world, and the actual world plays no role in the description of narrow mental states, which forms the basis of my psychologistic semantics. In this semantics I literally swapped out the Tarskian definition of truth for a definition of what it means to capture a mental state (cf. the Appendix of the target paper). Still, some of my argumentation against realism and semantic anti-realism in the first sections of the paper revolved around the judgment that some fictional and metafictional statements can be true. So ultimately, to really do justice to those data, I owe you a notion of truth.

The problem of defining truth arises for dynamic semantics more generally: what does it mean to say that an assertion, when interpreted as a context change potential rather than a classical possible worlds proposition, is true? Well, one thing we can easily do in dynamic semantics is say what it means for a *context* to be true. If we think of contexts as sets of worlds, we simply say that C is true if it contains the real world, w_0. But what about the truth of a sentence or assertive utterance? Following Heim (1982), we might say that an utterance is true relative to C if updating leads to a true context, and false if updating turns a true context into a false one (though cf. Stokke 2012 for extensive discussion and refinements).

(1)   a.   a context C is true if $w_0 \in C$; false otherwise.
          b. φ is true relative to C if $C + \varphi$ is true; φ is false relative to C if C is true and $C + \varphi$ is false; and φ lacks a truth value relative to C if C is false.

The dynamic notions of truth in (1) can be transposed to the psychologistic domain.[1] We should focus on the belief-like components within our complex mental states, because other attitudes (like desires, intentions, and imaginations), are not intuitively truth-evaluable. Thus, we can say, for instance, that a mental state, in the form of an NBAS or, derivatively, an ADS, is true iff its beliefs and anchors are true.

(2)   a.   An NBAS **A** is true iff for all BEL- and ANCH-labeled attitudes Q: for all $i \in BG(Q)$, there is a g: $\langle w_0, g \rangle \in Q(i)$.
          b.   An ADS **K** is true iff every NBAS captured by **K** is true.

Next, an assertion φ relative to context **K** is true if updating the belief component in **K** with φ leads to a true ADS, and false or undefined otherwise, as in (1b). Note that updating in ADT is a somewhat complicated procedure

---

1 Just like I will transpose the Heimian notion of dynamic entailment as update invariance in section 6 below.

involving preliminary DRS construction and presupposition resolution, so we'll use the following shorthand notation: $\mathbf{K}+_{\mathrm{BEL}}\varphi$ denotes the output of first merging the preliminary DRS representation of φ with the (relevant/salient) belief component of $\mathbf{K}$ and then resolving all presuppositions in $\mathbf{K}$. We can now define truth:

(3)   An assertion of φ is true relative to mental state $\mathbf{K}$ iff $\mathbf{K}+_{\mathrm{BEL}}\varphi$ is true.

This kind of definition brings our framework to the same level as Heim's non-psychologistic dynamic semantics, in that we can say of assertions whether they are true of false relative to a true context (understood now as a mental state in which the beliefs are true). More can and should be said about this (for instance in light of the distinction between wide and narrow content here, and Stokke's refinements of (1b)), but as far as regular assertions are concerned I'll leave it at this. In the next section I discuss how – if at all – we might account for the truth of fictional statements within my psychologistic framework.

2   Truth in fiction

Since the central idea behind my approach is that fictional statements are not to be interpreted as belief updates, but as imagination updates, it seems that the psychologistic definition of truth in (3) above won't work for fiction. But do we really need a notion of truth for fictional statements? Well, not if we follow Walton (1990) and analyze fictional statements as prescriptions to imagine (or better, as Rami and Zimmermann point out, as sanction-less *invitations* to imagine.) For we don't usually assign truth values to directive speech acts. This seems right. When I read *Lord of the Rings* I don't often stop and wonder if what Tolkien is writing is true. But then what to make of the intuition that 'Frodo is a hobbit' rings true, or at least more true than 'Frodo is an orc'?

At this point several commentators (including in particular Zucchi, Ninan, and Rami and Zimmermann) point out that I've neglected to draw a crucial distinction, viz. between on the one hand acts of storytelling and on the other hand reports about the content of a given fiction. The first type of what I simplistically called fictional statements is exemplified by Tolkien's written utterance of (4).

(4)   When Mr. Bilbo Baggins of Bag End announced that he would shortly be celebrating his eleventy-first birthday with a party of special magnificence, there was much talk and excitement in Hobbiton.

Ninan would call this an authorial diktat, Zucchi a fictional discourse. Rami and Zimmermann offer a more finegrained typology of uses of fictional discourse, roughly the first two of which (U1 and U2) seem to correspond to Tolkien's writing (4). Ninan and Zucchi agree that these authorial diktats, to use Ninan's original term, may indeed be thought of as truth-valueless directives, i.e. invitations to imagine. Rami and Zimmermann analyze diktats as (additionally) triggering belief updates – I'll return to their position shortly.

In any case, all commentators mentioned above agree that my utterance of (5) to another Tolkien reader is not so much an act of fiction-making or even storytelling, but rather a claim about the content of a certain fiction. Ninan calls this type of utterance a fictional report, Zucchi follows Currie's (1990) terminology and calls it a metafictive utterance:

(5)   Frodo was adopted by his cousin.

What I'm saying with (5) is that according to the Tolkien stories, Frodo was adopted by his cousin. Fictional reports then are really just regular assertions, and as such, interpreting (5) does not trigger a bout of imagination, but the true belief that the story in question in some way entails that Frodo was adopted by his cousin.

For what it's worth, my original reason for lumping fictional and metafictive utterances together in the original paper was a vague intuition similar to one that Zucchi helpfully attributes to Evans (1982), i.e., that in uttering (5) I'm continuing a game of pretense originating in reading Tolkien's works. The commentators' however provide several convincing arguments against this conflation.

The first argument for the diktat–report distinction is that, intuitively, authorial diktats have no truth value, but a fictional report like my utterance of (5) (or my original *Frodo is a hobbit*) does. Ninan and Zucchi suggest we model this by analyzing reports like (5) as prefixed with a silent 'fiction operator' ('in all worlds compatible with the fiction, φ', à la Lewis 1978), and then treating the whole as a regular assertion. Since I just defined notions of truth and falsity for assertions in section 1 above, this would indeed predict truth values for fictional reports but not for authorial diktats.

Interestingly, Rami and Zimmermann offer a dissenting view here. They sketch an alternative account in which the interpretation of an authorial diktat leads not – or not just – to an act of imagining, but to a bona fide belief:

> [. . .] the author who uses the sentence [(1a) 'Frodo is a hobbit born in The Shire'] literally as part of an act of telling or creating a story A, stipulates that there is some world of fiction that is

described by A and in which Frodo is a hobbit born in the Shire. However, if anything like this is a correct description of the conventional effects of such a fictional use of (1a), then the audience's natural and adequate cognitive reaction to it would be a belief that there is such a fictional world with the features described by (1a) rather than a mere act of imagination. (Rami and Zimmermann, this volume)

Since belief is truth-evaluable, it follows that, on Rami and Zimmermann's view, even authorial diktats may be said to have truth values. In fact, as I understand it, they will come out necessarily true, which would again set them apart from fictional reports.

Zucchi offers a second semantic argument against the conflation of authorial diktat and fictional report, based on the behavior of indexicals. He observes that indexicals in fictional reports are interpreted relative to their utterance context:

> While talking about *The Adventures of Sherlock Holmes* to the passenger next to me, I may point at London from the airplane window, as the city comes into view, and say truthfully:
>
> (6)   Holmes lives here. (Zucchi, this volume)

By contrast, in authorial diktats indexicals are interpreted relative to the story context, as Zucchi illustrates with a quote from Cain's story *The Butterfly*:

(7)   [...] Now I know what it is, I won't mind it any more, and tonight I'll get out of here.

The indexicals *I, tonight*, and *here* don't refer to the context in which Cain wrote this sentence, but to the context of the story's protagonist (who is at the same time also the fictional narrator). I take it that this general narrative indexical shift exemplified in (7) is distinct from indexical shifting in quotation and Free Indirect Discourse, which, I've argued elsewhere, can be reduced to a form of quoting a protagonist's thoughts or words. More needs to be said about the various kinds of indexicality and shifting in narratives.

In closing let's explore what distinguishing between diktats and reports would mean within the ADT framework. First, if we analyze fictional reports as assertions, triggering belief updates rather than imagination updates, what do we do with the occurrences of fictional names in them? And how do

fictional reports (aka *metafictive* statements) relate to *metafictional* statements like (8)?

(8)   Frodo is a fictional character invented by Tolkien

As for the first question, we'd want the presupposition triggered by the fictional name *Frodo* in a fictional report like (5) to be bound by the discourse referent for Frodo introduced previously into the interpreter's imagination through some authorial diktat – my report is about the same Frodo that I'm imagining about when I read Tolkien. This would essentially make fictional reports metafictional, in the sense of involving an imagination-dependent belief, much like I assumed for statements like (8) in the paper. Yet, there are crucial differences between (5) and (8), as witness the fact that the intuitive truth of both (5) and (8) doesn't entail that Bilbo adopted a fictional character.

In sum, incorporating the distinction between authorial diktats and fictional reports is an important step towards accounting for the truth or falsity of some fictional statements. Future research will have to provide and account for a more detailed taxonomy of sentences containing fictional names, building on the finegrained distinctions proposed by Rami and Zimmermann and by others in the literature (e.g. paratextual vs. metafictional statements (García-Carpintero 2010), or internal and external para- and metafictional ones (Voltolini 2006)).

3   Imagination and pretense

The starting point of my analysis was to take seriously the role of the mental state of imagination in the interpretation of fiction. This is what led me to abandon the cherished Stalnakerian or dynamic conception of communication as updates on a common ground, and adopt a psychologistic conception instead.

Recanati seems to be more or less on board with my general 'cognitive turn', but, he argues, my model of the imagination is too simplistic. Where I analyzed imagination simply as a mode of attitude, on a par with desire, belief, fear, etc., Recanati argues that "imagination is not an attitude like the others, because it has a secondary character: imagination is the simulation of (inter alia) other, more basic attitudes." Following in the footsteps of Meinong, Recanati wants to treat imagination as a higher-order "pretense operator" applied to more basic, first-order attitudes. Thus, alongside beliefs

we have pretend-beliefs, i.e. the simulation of a belief in our imagination,[2] and next to desires we have pretend-desires. In fact, Recanati's pretense operator is not even restricted to such traditional propositional attitudes, it also applies to emotions and anchors (mental files). This should account for the "heterogeneity of imagination".

For the interpretation of fictional names, Recanati now invokes *pretend-anchors*, i.e. imaginative simulations of anchors. Such pretend-anchors can be *deployed*, within the pretense. Deploying a pretend-anchor amounts to the agent pretending to refer, which, finally, carries pretend-referential commitments. This allows Recanati to choose the first horn of my dilemma about the interpretation of fictional names, i.e. the interpretation of names always goes through anchors, only in the case of fictional names they are pretend-anchors.

I very much like the idea of capturing the heterogeneity of imagination and making sense of pretend-anchors. So, let me speculate a bit about what Recanati's proposal would look like in my Attitude Description Theory (ADT) formalism.

At first sight, Recanati's conception of pretense as a higher-order mental state is reminiscent of his earlier analysis of pretense as involving "hyperinsulated" mental states (Recanati 2000), in turn inspired by the "multiple mental models" theory of Perner (1991). Understanding pretense, in these theories, involves the cognitive ability to keep in mind a number of distinct mental representations, e.g., one based on perception of the real world, where Teddy is sitting in front of an empty plastic teapot, and another one based on a story being enacted, where Teddy is the father pouring himself a cup of coffee.

Perner has further developed this idea in an (informal) file card model of mental states, which is quite similar in spirit to Kamp's and my DRT-based models, as well as to Recanati's (2012) mental file model.

> Children [. . .] need to quarantine the pretend representations from their world knowledge. We can capture this within our file card model by creating sub-files, i.e., larger file cards containing the basic file cards [. . .]Reality is represented by the basic file cards [. . .] , and a large card representing the pretence. This needs to be a larger card because the pretence consists itself of entities and information about them. On the pretend card

---

2 Along the way, Recanati provides a neat deconstruction of the notion of acceptance that is central to the Stalnakerian conception of common grounds, especially as applied to fiction, viz., acceptance = pretend-belief (cf. my response to Geurts below for more on the link between fiction and acceptance).

are the cards representing the pretend entities with anchoring information (Perner et al. 2007:486)

Transposing file cards to ADT we would represent a fiction-induced imagination as a complete mini-ADS, labeled with a special higher-order pretense marker: $\text{PRET}(x) : \mathbf{K} \approx$ '$\mathbf{K}$ represents a complex pretense based on text x.'

(9) $\left\{ \text{PRET}(x): \left\{ \begin{array}{c} \left\langle \text{ANCH}, \begin{array}{|c|} \hline x \\ \hline \text{book(x)} \\ \hline \end{array} \right\rangle \\ \left\langle \text{ANCH}, \begin{array}{|c|} \hline z \\ \hline \text{name(z,Samsa)} \\ \hline \end{array} \right\rangle \\ \left\langle \text{BEL}, \begin{array}{|c|} \hline \text{beetle(z)} \\ \hline \end{array} \right\rangle \end{array} \right\} \right\}$

Before delving into the model-theoretic implications, note that this Perner-inspired diagram indeed seems to represent imagination as a higher-order operator, and that within the pretense we can have any variety of basic attitudes like belief, but also desires and anchors.

The ADS structure in (9) also lends itself to formalizing the idea of quarantine or hyperinsulation, viz. by stipulating that discourse referents introduced within a pretense are not accessible from the outside. However, the kind of phenomena discussed in my paper actually warn against this. Counterfictional imagining, transfictional comparison and metafictional belief all require discourse referents set up in one pretense to be accessible to another pretense or belief. A natural way to implement this within the existing ADT architecture would be to, so to speak, distribute the pretense operator over the contents of the complex pretend representation:

(10) $\left\{ \begin{array}{c} \left\langle \text{ANCH}, \begin{array}{|c|} \hline x \\ \hline \text{book(x)} \\ \hline \end{array} \right\rangle \\ \left\langle \text{PRET}(\text{ANCH}, x), \begin{array}{|c|} \hline z \\ \hline \text{name(z,Samsa)} \\ \hline \end{array} \right\rangle \\ \left\langle \text{PRET}(\text{BEL}, x), \begin{array}{|c|} \hline \text{beetle(z)} \\ \hline \end{array} \right\rangle \end{array} \right\}$

I think this kind of representation corresponds roughly to the picture Recanati sketches in his commentary.

In future work I would like to make model-theoretic sense of the pretense operator in (10), and then explore the consequences of this different con-

ception of imagining. For instance, could we use this special treatment of imagination as a meta-attitude to shed new light on the various philosophical and linguistic puzzles about *de re/de se* imagination, and counterfactual attitudes and their reports more generally?

4   Common grounds and shared commitments

Geurts objects to the psychologism in my semantics. Although no one denies, he says, that a speech act has an effect on the mental state of the hearer, it has proven very fruitful in pragmatics to focus on the social aspect of communication, i.e. its effects on the common ground. Geurts argues that redefining the common ground in terms of commitments will allow for a proper socio-pragmatic account of fiction, obviating the need for a relapse into psychologism.

I concede that my rejection of the common ground approach in the paper was a bit hasty, and I like the exercise of rethinking Stalnakerian pragmatics in terms of commitments. Interestingly, though, I think a careful reading of Stalnaker's own remarks about the common ground offer an even more conservative approach, without the need for a whole new commitment architecture.

Deviating from Stalnakerian orthodoxy, Geurts views communication as the negotiation of joint commitments. Crucially, commitments are not mental states but interpersonal relationships, i.e. the basic notion is a three place relation: x commits himself to y to act in accordance with φ. From there, Geurts redefines the common ground as the set of joint commitments. The key feature Geurts relies on for his analysis of fiction is that the commitment-based common ground, unlike the standard conception of common ground as common belief, is "doxastically neutral," i.e. "a and b may be jointly committed to φ without believing φ to be true."

Stalnaker and others have also recognized that what we take for granted in a conversation, i.e., what we presuppose in the pragmatic sense, can diverge from what we believe. In various scenarios we may act as if we believe something and communication proceeds as usual. Stokke (2013) for instance discusses lying and Stalnaker (1970) himself already touches on pretense:[3]

> A speaker may presuppose what is untrue to facilitate communi-
> nication, as when an anthropologist adopts the presuppositions
> of his informants in questioning him. Most innocent of all are

---

3 Sainsbury (2011) works out this Stalnakerian analysis of fiction in some detail. Maier (forthcoming) explores the relation between lying and fiction in a Stalnakerian setting.

cases of fiction and pretending: speaker and audience may
conspire together in presupposing things untrue. (Stalnaker
1970:39–40)

Stalnaker accommodates this insight by replacing belief with acceptance as
the basic attitude on which the notion of common ground is defined. He
characterizes this propositional attitude of acceptance informally as follows:

To accept a proposition is to treat it as a true proposition in
one way or another – to ignore, for the moment at least, the
possibility that it is false. (Stalnaker 1984:79)

Like Geurts's commitment, acceptance is doxastically neutral – we can choose
to accept things we don't believe. From this notion we can construct a suitable
notion of common ground. Stokke (2013) extracts the following precise
acceptance-based definition from Stalnaker (2002):

(11)   A proposition p is common ground iff both speaker and hearer accept
         p, and believe that they accept p, and believe that they believe that
         they accept p, . . .

Acceptance and commitment are similar in that they are both voluntary
and doxastically neutral, in contrast to belief. And both, presumably, give
rise to a notion of common ground that is suitable for modeling linguistic
communication as a social activity.

But acceptance and commitment are not identical. Acceptance is just a
belief-like propositional attitude, a two-place relation between an individual
and a proposition, and thus inherently simpler than the three-place commit-
ment. In other words, with acceptance, the social aspect only comes in at the
level of the definition of common ground. It would be an interesting project
for future research to compare the resulting common ground models in more
detail.

In any case, shifting from (common or individual) belief to acceptance or
commitment, or imagination for that matter, is but the first step in analyzing
fiction. As Geurts observes, we still need some way of separating fictional
common grounds from the regular common ground. And this is where the real
action is in analyzing fiction, whether psychologistically, or intersubjectively
in terms of common grounds. For instance, just as we saw in section 3 how
Recanati and Perner had suggested that mental representations of pretense
are insulated or quarantined off of beliefs about the actual world, Eckardt
(2014) suggests a version of the common ground analysis where fictional
common grounds are kept completely separate from the official common

ground. But, as Lewis (1978) and Bonomi & Zucchi (2003) argue, lots of information about the real-world carries over into our representations of fictional worlds. And examples like the ones I discuss, involving metafictional statements and counterfictional imagination, show that information also flows in other directions, from fiction to beliefs, and from a fiction to another pretense, respectively. Whether it is shared commitments (as proposed by Geurts or García-Carpintero 2015), or joint acceptance (as first suggested by Stalnaker), or individual imagination (as I proposed in the target paper) that will prove to be the best way to model these types of information flows, is a question for future research.

5 Referring to mere possibilia

As Ninan observes, my account relies heavily on the notion of parasitic attitudes. For example, counterfictional imagination and metafictional belief involve mental states parasitic on a fiction-induced imagination. Ninan proposes an account of these kinds of cases that makes similar predictions to mine, with similar metaphysical assumptions, but without relying on parasitic attitudes. Although I believe we need parasitic attitudes independently, for dealing with the semantics of reports like *Hans wants the ghost in his attic to be quiet* (cf. §4.4 of the target paper), it is worthwhile examining if we also need this (rather heavy) machinery for fiction.

The core of Ninan's proposal is this:

> Fictional names refer to fictional entities. Fictional entities are mere possibilia: they exist at other worlds but not at the actual world. (Ninan, this volume)

At first sight this seems relatively straightforward. *Frodo* is a referential expression, referring directly to the flesh and blood hobbit Frodo, who, like me and most other individuals, doesn't exist in all possible worlds. As it happens, Frodo doesn't exist in the actual world, just like I don't exist in (many of) the *Lord of the Rings* worlds.

Interestingly, Ninan's elegant proposal coincides with a view that Kripke has tried to reject in *Naming and Necessity*. I will reconstruct Kripke's counterargument and then discuss how this relates to my own view.

Here is Kripke's original, brief formulation of the argument:

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

distinct possible people, and even actual ones such as Darwin or Jack the Ripper, might have performed the exploits of Holmes, but there is none of whom we can say that he would have been Holmes had he performed these exploits. For if so, which one?

I thus could no longer write, as I once did, that 'Holmes does not exist, but in other states of affairs, he would have existed.' [(Kripke 1963)] The quoted assertion gives the erroneous impression that a fictional name such as 'Holmes' names a particular possible-but-not-actual individual. (Kripke 1980:158)

Ninan's view as quoted earlier aligns with the one under attack here, as it indeed involves 'Holmes' naming a particular possible-but-not-actual individual.

Kripke's counterargument is somewhat hidden in this brief passage, but he elaborates on it in a later lecture on the subject of fictional names (Kripke 2011). The idea is that there is no way of knowing which individual is Holmes. In contrast to existing individuals, there is no fact of the matter as to who Holmes is – mere possibilia do not have well-defined identity criteria. All we have in the case of a fictional entity is the descriptions provided by the author, but multiple ontologically distinct individuals could in principle satisfy those.

Let me try to put it another way. Let's start with Kripke's account of the semantics of genuinely referring names. The usage of a genuinely referential term can in principle always be traced back, through a causal–historical chain, to some individual, who does, as a matter of fact, possess certain properties. Thus, we can truthfully say 'Kripke exists' or 'Kripke wrote *Naming and Necessity*' and thereby ascribe properties to a particular individual, viz the individual at the end of the chain. We can also make *de re* modal statements about such an individual ('Kripke might not have existed' or 'Kripke might not have written N&N'). Empirical research could even lead us to question whether Kripke has some of the properties we typically associate with him, i.e., it could still turn out that Kripke was a fraud and in fact stole or bought the text of N&N and put his name on it.

By contrast, for a name like *Holmes* there simply is no causal–historical chain leading us from the usage of the name to a particular entity. Hence, there is no empirical grounds on which we can make sensible *de re* modal claims, or on which we might conclude that we were mistaken, e.g. finding out that he was not really a detective, or didn't actually play the violin.

We end up with a fundamental asymmetry between the actual world and other possible worlds, in that we can meaningfully say (12a) but not (12b):

(12) a. Kripke exists, but in some other possible worlds he doesn't.
      b. Holmes doesn't exist, but in some other possible worlds he does.

Kripke concludes that *Holmes* simply is not a genuinely referential term – at least not in Ninan's sense in which the name is supposed to pick out a merely possible individual.[4]

If we buy Kripke's argument – which of course not everybody does, cf. e.g. Ben-Yami (2010) – we must abandon Ninan's proposal. But then what about my own proposal? After all, as Ninan observes, his metaphysical assumptions closely follow mine:

> Perhaps this means they [=fictional entities] are in the domains of other worlds, but not in the domain of the actual world; perhaps it means that they are in the domain of every world, but exist concretely at other worlds, non-concretely at the actual world. Maier should have no objection to these metaphysical claims, for his own account relies on some such view. The principal difference so far between this account and Maier's own account is that, on this [=Ninan's] account, names refer to these fictional entities, rather than inducing existential quantification over them. (Ninan, this volume)

Both Ninan and I, as well as the various time-slices of Kripke considered above, agree that existence is a property, and one that individuals typically possess only contingently. Kripke's argument targets precisely the point where Ninan and I part ways, viz. the way in which fictional names refer.

For Ninan, fictional name reference is just standard, direct/rigid reference: $[\![\text{Frodo}]\!]_w = \text{Frodo}$, if he exists at w; undefined otherwise, i.e., the kind of definition that Kripke finds unintelligible (for fictional names, that is). On my account, the name *Frodo* triggers the existential presupposition that there is someone by that name, and this presupposition gets bound or accommodated within the imagination. As far as purely fictional statements are concerned, the resulting interpretation I predict is roughly equivalent to classical descriptivist analyses, like, say, Kaplan's (1973) or Currie's (1990): in all imagination worlds there is some hobbit named *Frodo*, who . . . . There

---

4 I should add that Kripke (2011) later endorses a different kind of realist position in which *Holmes* actually refers, to an abstract entity. Since Ninan seems to be on board with the counterarguments I mention in the paper against this type of abstract-object-realism, I'll leave my Kripke exegesis at this.

is no rigid designation here,[5] no single possible individual who *is* Frodo. At best, we could say that Frodo here corresponds to an individual concept, but then one that is not even weakly rigid.

6   The logic of mental states

Aloni argues that my Attitude Description Theory (ADT) does not yield a very useful logic. In particular, she notes that certain intuitively valid inferences involving beliefs about fictional entities or even real entities are not predicted to be valid. For instance, from (13) we want to be able to infer (i) that the agent believes that Tolkien is an author and (ii) that she believes that Tolkien invented Frodo.

(13)
$$\left\{ \begin{array}{c} \left\langle \text{ANCH}, \boxed{\begin{array}{c} x \\ \hline \text{name(x,Tolkien)} \\ \text{author(x)} \end{array}} \right\rangle, \left\langle \text{IMG}, \boxed{\begin{array}{c} y \\ \hline \text{name(y,Frodo)} \end{array}} \right\rangle, \\ \left\langle \text{BEL}, \boxed{\text{invent(x,y)}} \right\rangle \end{array} \right\}$$

Inferences about beliefs were not discussed in the paper, so Aloni starts by proposing a straightforward definition:

(14)   An agent with mental state described by ADS **K** believes that φ iff every NBAS captured by **K** has a belief component that entails φ.

This definition is still incomplete in various respects. First of all, even if **K** has but a single belief, the corresponding NBAS may have many. Following Aloni I'll ignore this and just consider only **K** and NBAS with a single belief component. Next, in the case at hand, the relevant belief component in the NBAS will be referentially dependent on (at least) two other components. This means we'll want to define the relevant entailment notion in the definiens relative to the belief's background.

Finally, what exactly is the role of φ in the definiens in (14)? Intuitively, when we ask if the agent's mental state entails the belief that Tolkien is an author, are we saying that the agent has the general belief that there is someone named Tolkien who is an author, or the singular belief, about the actual real-world Tolkien, that he's an author, or perhaps just the subjective

---

5 Regular proper names of course do refer rigidly to their bearers, by virtue of the mechanisms of internal and external anchoring. Recall that my uniform analysis of fictional and other names consists in the uniform treatment of all names as presupposition triggers, not in associating truly referential readings to all occurrences of names.

belief corresponding to the agent's interpretation of *Tolkien is an author*, or the belief that the agent would express thus?[6]

Given my psychologistic framework, a definition in terms of the subjective interpretation of an utterance seems reasonable. Following our definition of truth in section 1 above we'll use $\mathbf{K} +_{\text{BEL}} \varphi$ to denote the interpretation of φ as a belief update relative to a mental state $\mathbf{K}$.

(15)   An agent with mental state described by an ADS $\mathbf{K}$ believes φ iff every NBAS captured by $\mathbf{K}$, is also captured by $\mathbf{K} +_{\text{BEL}} \varphi$

With this definition in place let's take a closer look at Aloni's two challenges in turn, as each poses a different problem. I hope that spelling out some concrete derivations below will help elucidate the model-theoretic underpinnings of my proposal.

Challenge 1: Tolkien is an author

Updating our $\mathbf{K}$ with *Tolkien is an author* results in the following slightly different output ADS:

(16)
$$\left\{ \left\langle \text{ANCH,} \begin{array}{|c|} \hline \text{x} \\ \hline \text{name(x,Tolkien)} \\ \text{author(x)} \\ \hline \end{array} \right\rangle, \left\langle \text{IMG,} \begin{array}{|c|} \hline \text{y} \\ \hline \text{name(y,Frodo)} \\ \hline \end{array} \right\rangle, \left\langle \text{BEL,} \begin{array}{|c|} \hline \text{invent(x,y)} \\ \text{author(x)} \\ \hline \end{array} \right\rangle \right\}$$

According to Aloni, (16) is a strictly more informative ADS than (13). To see this, we'd have to construct an NBAS captured by the original ADS, but not by the updated ADS. This would be an NBAS with a dependent belief state Q_3 in which x is not an author, but which depends on a different state, Q_1 in which x is an author named Tolkien. Now, the whole point of using two-dimensional attitudes was to make such inconsistent dependencies possible, so indeed the semantics I gave does not exclude such NBAS, and therefore, Aloni is right that the original ADS in (13) does not entail the belief that Tolkien is an author.

Luckily, a straightforward solution exists. As I argue in the target paper, the content of an anchor is something that the agent holds true, even if, as

---

6 The latter two subjective belief options are roughly equivalent. They differ subtly from the singular belief option in that they don't require that Tolkien actually exist, as the relevant agent's Tolkien anchor may turn out to be faulty.

Recanati points out, only dispositionally. Perhaps – using Recanati's terminology – when we report the agent as believing that Tolkien is an author we're reporting not just on the content of the occurrent thought event, but more generally on what they hold true in the dispositional sense (or, what they are committed to, as Geurts would say). A straightforward implementation of this idea would be to just merge all or some of the beliefs and anchors in the ADS into a new belief, and then apply the previous definition. For instance:

(17)   An agent with mental state captured by ADS **K** dispositionally believes that φ iff the result of merging the belief DRS with the anchors and beliefs it depends on into a single BEL-labeled attitude gives a well-formed ADS that entails that the agent believes that φ.

Challenge 2: Tolkien invented Frodo

Aloni's second example makes a different point. On the current reconstruction of belief-entailment, we would check whether the ADS is invariant under the interpretation of *Tolkien invented Frodo* as a belief update. This seems to be the case. The name *Tolkien* binds to x, introduced in the anchor, and the name *Frodo* to y, introduced in the imagination, leaving only 'invent(x,y)' in the belief, which was already there, so we seem to be making the right prediction here. Unless, of course, there is something more fundamentally wrong in the underlying ADT semantics, which is what Aloni suggests.

To evaluate Aloni's concern, let me demonstrate how the semantics is supposed to work by proving that in any NBAS captured by **K**, the ADS in (13) above, the relevant belief component contains the information that x (i.e. Tolkien) invented y. As far as I can tell, the semantics works as it should, but I admit that some crucial definitions were not presented in sufficient detail in the Appendix to properly appreciate this. This defect of my presentation is remedied below.

Let's make our exercise more concrete. Let **A** be some NBAS captured by **K**. Say $\langle \text{ANCH}, Q_1 \rangle, \langle \text{IMG}, Q_2 \rangle, \langle \text{BEL}, Q_3 \rangle \in \mathbf{A}$ correspond to the three components of **K**. Let's further assume, for simplicity, that **A** consists of just these three components, and that the discourse referents in the NBAS and ADS are aligned, so that no renaming is necessary. What I will show is that, relative to its background, Q_3 contains the information that x invented y, or more precisely,

(18)   For all $\langle w, f \rangle \in BG(Q_3)$, $Q_3(\langle w, f \rangle) \rhd \{\langle w', f \rangle | f(\text{x}) \text{ invented } f(\text{y}) \text{ in } w\}$.

The tricky part of the proof is the computation of the relevant background state for Q_3 out of the dependencies, Q_1 and Q_2. The complications are due to the fact that my definition allows that dependencies may themselves have further dependencies, recursively. In this case, however, Q_1 and Q_2 are independent, so intuitively, the relevant background should correspond to the information state merge of Q_1 and Q_2.

Perhaps what threw Aloni off was that I did not explicitly address what independence means, formally. The simplest way to implement it would be to allow dependent attitudes (2D information states) and independent ones (standard information states). In another paper (Maier 2016), I proposed an alternative, more uniform approach: a 2D information state Q is independent iff its dependence is trivial, in the sense that (i) its background is Λ, the so-called empty information state $(= \{\langle w, \emptyset \rangle | w \in W\})$, and (ii) $Q(i) = Q(i')$ for all $i, i' \in \Lambda$. For readability we could then denote $Q(i)$, for arbitrary $i \in \Lambda$, as $\overline{Q}$. We then stipulate that all information states at the bottom of dependency chains in an NBAS are independent in this sense.

Now that we know what it means for Q_1 and Q_2 to be independent, we can apply definition (48) from the Appendix to calculate the background of Q_3:

(19) $\qquad BG(Q_3) = \bigcup \{Q_1(i) \uplus Q_2(i') | i, i' \in \Lambda\} = \overline{Q_1} \uplus \overline{Q_2}$

The notion of information state merge, $\uplus$, defined in (39d) in the Appendix, is the semantic analogue of DRS merge, i.e. we combine the information in two information states (typically, reducing the number of worlds while extending the assignments). In this case:[7]

(20) $\qquad \overline{Q_1} \uplus \overline{Q_2} = \{\langle w, f \cup f' \rangle | \langle w, f \rangle \in \overline{Q_1} \text{ and } \langle w, f' \rangle \in \overline{Q_2}\}$

It follows that the background for Q_3 contains all the information contained in the merge of the anchor and the imagination DRSs. Semantically:

(21) $\qquad BG(Q_3) \rhd \{ \langle w, f \rangle \mid Dom(f) = \{x, y\} \text{ and } f(x) \text{ is an author named}$
$\qquad\qquad\qquad\qquad\qquad \text{Tolkien in } w, \text{ and } f(y) \text{ is named Frodo in } w \}$

Now, as I point out in the paper, it's still an open question what to do if the merge of two relevant background states would be inconsistent, but that doesn't seem to be the case here. Moreover, Q_3 doesn't depend on two independent states with different domains separately, but on a single – if

---

7 Since the universes of the corresponding ADS components are disjoint, we can assume that the domains of $\overline{Q_1}$ and $\overline{Q_2}$ are likewise disjoint, and hence the requirement in (39d) that $f \cup f'$ is a function is vacuous.

somewhat peculiar – mixed state, viz. $BG(Q_3)$, which contains information about both Tolkien, x, and Frodo, y.

The content of the belief, Q_3, itself is to be computed relative to possibilities from this background (whose domain subsumes $\{x,y\}$). Applying the matching condition for the dependent belief, gives us that, for each background possibility $\langle w, f \rangle$, $Q_3(\langle w, f \rangle) \rhd [\![K_{\text{BEL}}]\!]^{\text{is}^f}$, from which the truth of (18) follows directly.

References

Ben-Yami, Hanoch. 2010. Could Sherlock Holmes Have Existed? *Croatian Journal of Philosophy* 10(3). 175–181.

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.

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.

Evans, Gareth. 1982. *Varieties of Reference*. Oxford: Oxford University Press.

García-Carpintero, Manuel. 2010. Fictional singular imaginings. In Robin Jeshion (ed.), *New Essays on Singular Thought*, 273–299. Oxford: OUP.

García-Carpintero, Manuel. 2015. Contexts as Shared Commitments. *Frontiers in Psychology* 6. http://dx.doi.org/10.3389/fpsyg.2015.01932.

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

Kaplan, David. 1973. Bob and Carol and Ted and Alice. In K. J. J. Hintikka, J. M. E. Moravcsik & P. Suppes (eds.), *Approaches to Natural Language*, Dordrecht: Springer Netherlands.

Kripke, Saul. 1963. Semantical Considerations on Modal Logic. *Acta Philosophica Fennica* 16(1963). 83–94.

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

016-9397-7.

Maier, Emar. forthcoming. Lying and Fiction. In Jörg Meibauer (ed.), *Oxford Handbook of Lying*, Oxford: OUP. https://philpapers.org/rec/MAILAF.

Perner, Josef. 1991. *Understanding the representational mind.* Cambride, MA: MIT Press.

Perner, Josef, Bibiane Rendl & Alan Garnham. 2007. Objects of desire, thought, and reality: Problems of anchoring discourse referents in development. *Mind and Language* 22(5). 475–513. http://dx.doi.org/10.1111/j.1468-0017.2007.00317.x.

Recanati, François. 2000. *Oratio Obliqua, Oratio Recta*. Cambridge: MIT Press.

Recanati, François. 2012. *Mental Files*. Oxford: OUP.

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

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

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

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy* 25(5-6). 701–721. http://dx.doi.org/10.1023/A:1020867916902.

Stokke, Andreas. 2012. Truth and Context Change. *Journal of Philosophical Logic* 43(1). 33–51. http://dx.doi.org/10.1007/s10992-012-9250-6.

Stokke, Andreas. 2013. Lying and Asserting. *Journal of Philosophy* 110(1). 33–60.

Voltolini, Alberto. 2006. *How ficta follow fiction*. Dordrecht: Springer.

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