Contents

ARTICLES

BOOK REVIEWS

REPORTS

Organon F 24 (3) 2017: 282-307

(Mock-)Thinking about the Same

ALBERTO VOLTOLINI[1]

ABSTRACT: In this paper, I want to address once more the venerable problem of *intentional identity*, the problem of how different thoughts can be about the same thing even if this thing does not exist. First, I will try to show that antirealist approaches to this problem are doomed to fail. For they ultimately share a problematic assumption, namely that thinking about something involves *identifying* it. Second, I will claim that once one rejects this assumption and holds instead that thoughts are *constituted* either by what they are about, their *intentional objects*, or by what determines their proposition-like *intentional contents*, one can address the problem of intentional identity in a different way. One can indeed provide a new solution to it that basically relies on two factors: a) what sort of metaphysical nature intentional objects effectively possess, once they are conceived as *schematic* objects *à la* Crane (2001, 2013); b) whether such objects really belong to the overall ontological inventory of what there is. According to this solution, two thoughts are about the same nonexistent intentional object iff i) that object satisfies the identity criterion for objects of that metaphysical kind and ii) objects of that kind belong to the overall ontological inventory of what there is, independently of whether they exist (in a suitable first-order sense of existence). As such, this solution is neither realist nor antirealist: *only if* condition ii) is satisfied, different thoughts can be about the same nonexistent *intentionale*; otherwise, they are simply constituted by the same intentional content (provided that this content is not equated with that *intentionale*). Third, armed with this solution, I will hold that one can find a suitable treatment of the specific and related problem of whether different people may mock-think about the same thing, even if there really is no such thing. Finally, I will try to show that this

[1] Received: 30 March 2017 / Accepted: 1 July 2017

✉ Alberto Voltolini
Department of Philosophy and Education Sciences
University of Turin
via S.Ottavio 20, I-10124 Turin, Italy
e-mail: alberto.voltolini@unito.it

treatment can be also applied to the case in which different thoughts are, according to phenomenology, about the same *intentionale* and yet this *intentionale* is of a kind such that there really are no things of that kind. For in this case, such thoughts are about the same *intentionale* only *fictionally*.

KEYWORDS: Intentional identity – intentional objects – intentional contents – schematic objects.

1. Introduction

In this paper, I want to address once more the venerable problem of *intentional identity*, the problem of how different thoughts can be about the same thing even if this thing does not exist. First, I will try to show that antirealist approaches to this problem, according to which there is no such thing and yet there is a legitimate sense in which the relevant thoughts are about the same thing, are doomed to fail. For they ultimately share a problematic assumption, namely that thinking about something involves *identifying* it. Second, I will claim that once one rejects this assumption and holds instead that thoughts are *constituted* either by what they are about, their *intentional objects*, or by what determines their proposition-like *intentional contents* (provided that these contents are not equated with such objects) one can address the problem of intentional identity in a different way. One can indeed provide a new solution to it that basically relies on two factors: a) what sort of metaphysical nature *intentionalia* effectively possess, once they are conceived as *schematic* objects *à la* Crane (2001; 2013); b) whether such objects belong to the overall ontological inventory. According to this solution, two thoughts are about the same nonexistent intentional object iff i) that object satisfies the identity criterion for objects of that metaphysical kind and ii) objects of that kind belong to the overall ontological inventory, independently of whether they exist (in a suitable first-order sense of existence). As such, this solution is neither realist nor antirealist: *only if* condition ii) is satisfied, different thoughts can be about the same nonexistent *intentionale*; otherwise, they are simply constituted by the same intentional content (provided that this content is not equated with that *intentionale*). Third, armed with this solution, I will hold that one can find a suitable treatment of the specific and related problem of whether

different people may mock-think about the same thing, even if there is no such thing. Finally, I will try to show that this treatment can be also applied to the case in which different thoughts are, according to phenomenology, about the same *intentionale* and yet this *intentionale* is of a kind such that there really are no things of that kind. For in this case, such thoughts are about the same *intentionale* only *fictionally*.

2. Intentional identity: the problem

From Geach (1980) onwards, a subtle problem has been dogging contemporary philosophy of language and mind: how can different people talk of and think about the (numerically) same thing, if this thing does not exist? In Geach's notorious example,

(1) Hob believes a witch has blighted Bob's mare, and Nob believes that she (the same witch) killed Cob's sow

many of us have the intuition that by uttering (1), one intends to talk of the same individual, so that (1) amounts to a *de re* report of different thoughts about it. We are therefore tempted to read (1) either strongly as:

(1a) $\exists\alpha$ (W(α) $\wedge$ Hob believes that B(α,b) $\wedge$ Nob believes that K(α,c))

or weakly as:

(1b) $\exists\alpha$ (Hob believes that W(α) $\wedge$ B(α,b) $\wedge$ Nob believes that K(α,c)).

Both (1a) and (1b) entail that there is an individual (1) talks of and the reported thoughts are about. (1b) is simply a weaker reading than (1a) for it does not force one to hold that the individual it talks of is a witch, but just that it is believed to be such. Yet how can this entailment be right, if the individual in question – the alleged witch – does not exist?[2]

[2] For this elegant way of introducing the problem, see García-Carpintero (2016, 334).

3. The antirealist solutions

Clearly enough, if one is a *realist tout court* about nonexistent things, the problem is easily solved. Different thoughts can be about the same nonexistent thing just as other thoughts are about the same thing that straightforwardly exists. In this respect, the fact that the former thing, unlike the latter thing, does not exist (in a suitable first-order sense of existence) makes no difference, since the realist is ontologically committed to it as well.[3] Yet the problem is serious for those who do not believe in nonexistent entities, *antirealists*. Now, several antirealist answers have been provided to the question of how different thoughts can be about the same nonexistent 'item' – I put this expression within quotes in order to indicate that *qua* antirealists, sustainers of this approach are not ontologically committed to such a nonexistent thing. Neither of them, however, appears to be fully convincing, as the following review will show.

To begin with, according to the so-called 'name-centric' approach,[4] different people think about the same nonexistent 'item' insofar as they share the same name-using practice involving a certain baptism, though unsuccessful because referentless, and the same communication chain. The necessary condition for that chain is a certain referential intention, i.e., the intention of referring to the same thing, if any, one's mentors referred to (e.g. Donnellan 1974; Taylor 2003; Sainsbury 2005).

Clearly enough, in order for this approach to work in the case of (1), one must name the nonexistent 'item' in question, the alleged witch in that case: "Hob believes that a witch, *namely* [a name follows], …". Yet this clarification notwithstanding, the approach does not seem to work. First of all, it fails to provide *necessary* conditions of intentional identity. As Friend (2014) has shown, in thinking e.g. about Santa Claus and about Father Christmas, people nowadays clearly think about the same nonexistent 'item'. Yet they fail to use the same name ("Santa Claus", "Father Christmas"). Granted, a supporter of this approach may invoke Mentalese and hold that such people mobilize the same Mentalese name. Yet this move fares no better. For, adds Friend, those people share no referential intention.

---

3 For an example of a realist solution, see Salmon (2002), who equates the relevant nonexistent thing with an abstract artifact non-spatiotemporally existing.

4 For this label see Friend (2014, 308).

Intending to refer to what one's mentors originally allegedly referred to by "Father Christmas", a certain personification of Christmas, is not the same as intending to refer to what one's informants originally allegedly referred to by "Santa Claus", an imaginary variant of an ancient bishop; those referential intentions mobilize different chains (cf. Friend 2014, 315).

At this point, fans of name-centrism may bite the bullet and say that in this case there is just a sort of intentional match: people nowadays using "Father Christmas" no longer want to defer to their mentors' referential acts, they just want to use it as they also use "Santa Claus". Yet even if this move worked, the approach fails to provide *sufficient* conditions of intentional identity. In thinking about a nonexistent 'item', even if one has the intention of referring to what her mentors allegedly referred to, she may fail to preserve such a reference. Another example by Friend shows this point. In writing *Hamlet*, Shakespeare may have intended to refer by "Hamlet" to what Kyd referred to. Yet he failed to refer to what Kyd actually referred to; namely, a real Danish prince actually named "Amleth" (as Friend prompts us to suppose; see Friend 2014, 316).

As an alternative antirealist approach, Friend herself has proposed the 'info-centric' approach (see Friend 2014). According to this approach, people think about the same nonexistent 'item' insofar as they mobilize the same *network* of notions, either because they share the same *mental file*, a certain mental repository of information, or because different such files, which may even be labelled differently, *suitably* overlap, by also being appropriately connected in that notional network.

To begin with, appealing to the idea of sharing a notional *network* rather than mere sharing a single file or mere having a file overlap has the advantage of overcoming a problem that antirealists who simply appeal to files' similarities in order to account for intentional identity unsuccessfully face (cf. Crane 2013, 161). The problem consists in the fact that not only repository similarity, but also repository identity *per se*, does not guarantee intentional identity.[5] Consider indeed the following counterexample to this simple account. Intuitively, in their being located in different though quasi-identical planets, Earth and Twin-Earth, Leverrier and Twin-Leverrier think about different nonexistent 'items', Vulcan and Twin-Vulcan, even if their mental files share the same information (witness the fact that they

[5] As Dennett (1968, 337) originally envisaged.

would notice nothing if they unconsciously swapped their planetary positions) and are also labelled the same ("Vulcan"). Now, the supporter of the info-centric approach may say that this counterexample to the simple account hits its target. For, since the Twins are disconnected in their being placed in different planets, they share no notional network.

Yet sharing the same notional network is again no *necessary* condition of intentional identity. For, as Everett (2013) has pointed out, people may think about the same nonexistent 'item' even if they do not share the same notional network. For example, different people may think about Mrs. Polonius, even if they attend different acts of different performances of *Hamlet* in which they grasp different ascribed features of what actually is one and the same nonexistent 'individual' – say, being Ophelia's mother, being Polonius' wife (cf. Everett 2013, 96).

On behalf of Friend, one may reply that the counterexample is misconceived. Appearances notwithstanding, those people share the same notional network. When no such network is actually shared, the fan of info-centrism acknowledges that the situation matches the one her detractor proposes in the 'Mrs. Polonius' case. Consider another case[6] involving two utterly independent 'Vulcan'-like baptisms, namely utterly independent baptisms mobilizing the same nominal form – "Vulcan" – yet allegedly focused on the same nonexistent 'celestial body', or anyway two utterly independent pointings allegedly ostending the same nonexistent 'item'. Definitely, in this case the baptisms or the pointings share no notional network. Thus, appearances notwithstanding, they do not focus on the same 'target'. Yet the 'Mrs. Polonius' case is different from this 'Vulcan' case, for unlike this case the people involved are trapped into the same network.

Yet this info-centric reply only prompts the further question of what makes two people be members of the same notional network. To begin with, a fan of info-centrism must hold that not only *causal*, but also *intentional* dependence of one's file on the relevant files of a practice's producers counts on this concern. For mere causal dependence is not enough. There may be cases of people thinking about different nonexistent fictional 'individuals' even if their thoughts are causally based either on a thought about the same nonexistent fictional 'individual' or even on a thought about

---

6 To reformulate Edelberg's (1992, 574-575) example. *Pace* García-Carpintero (2016, 342), this case is no counterexample to the info-centric approach.

the same concrete individual. Examples of the first kind of situation are given by uses of "Holmes" that trace back to Doyle's original use and nevertheless they are embedded in practices concerning completely different fictional 'Holmeses' (cf. Pautz 2008, 7). Examples of the second kind of situation are given by uses of the same name that trace back to an original use in which the name refers to a real individual – say, the very author of a certain novel – and nevertheless they are embedded in practices concerning completely different fictional 'individuals' (cf. García-Carpintero 2016, 343).

Yet if this is the case, then belonging to the same notional network is not even *sufficient* for intentional identity. For, as Friend herself admits, in thinking about a nonexistent 'item' one can exploit a certain intentional dependence on another file and yet fail to think, by means of her own file, about the same 'item' the other file focuses on. For example, Flaubert's writing "Madame Bovary c'est moi" shows that his 'Emma'-file intentionally depended on his own 'I'-file. Yet when writing *Madame Bovary*, Flaubert was obviously thinking about the nonexistent 'individual' *Emma*, not about himself (cf. Friend 2014, 324, 329).

As we have seen before, pretense may be involved in cases of intentional identity. Indeed, a third option in the antirealist camp is the 'pretense' approach. As Everett (2013) maintains, people think about the same nonexistent 'item' insofar as they belong to the same pretense practice of co-referring to, or better co-thinking about, the same 'individual'.

This way of putting things allegedly accommodates the aforementioned 'Mrs. Polonius' case. Although different people may attend different acts of different performances of *Hamlet* in which they grasp different ascribed features of what actually is one and the same nonexistent 'individual', they share the same pretense practice of make-believedly referring to, or even thinking about, the same 'individual'.

This option immediately prompts a question as to the conditions for being members of the same pretense practice. On behalf of this approach, one may suggest that a condition for the identity of a certain pretense practice is the existence of a *causal* relation between pretend uses of the same referential term.[7] If "referential term" is used here broadly enough, this

[7] For this suggestion, which traces back to Evans (1982), Everett (2000), and Kroon (2001), cf. Pautz (2008).

sounds at least a necessary condition for pretense membership. People involved in the 'Mrs. Polonius' pretense practice may not even grasp that the same name, "Mrs. Polonius", is mobilized, yet there is a causal relation between them that makes them make-believedly point to the same nonexistent 'individual' by means of the same demonstrative. The 'Mrs. Polonius' case suggests that this causal relation is also a sufficient condition, but this is controversial. Being causally induced to pretend something by the existence of a certain pretense practice hardly seems to be enough in order for the new practice to be the same as the old practice. Perhaps not even reinforcing this requirement with an *intentional* requirement is enough for being members of the same pretense practice, for people intending to play the same pretense game may be wrong. Probably, identity in story-telling is also required (cf. Voltolini 2006, 70-71).

Yet this problem notwithstanding, the existence of a shared pretense practice is once again not a *necessary* condition of intentional identity. The Geach original case, for one, does not involve pretense (cf. also García-Carpintero 2016, 346). It is also hardly a *sufficient* condition. Let me reformulate an aforementioned worry in different terms. Let us suppose that in writing *Ulysses*, Joyce protracted the same kind of pretense that Homer initiated. Yet by allegedly pretending that in so writing he was thinking (under the different name "Leopold Bloom", but possibly under the same masculine pronoun) about the same thing Homer did (under the name "Odysseos"), did Joyce think about the same nonexistent 'individual' as Homer did?[8]

4. Why the antirealist solutions do not work

Clearly enough, the above options do not exhaust the possibilities an antirealist has of solving the problem of intentional identity. So, it may well

[8] García-Carpintero (2016) defends a 'presuppositional' approach to intentional identity that bears some similiarities to the 'pretense' approach. For a similar view, see also Howell (2015). Being more general than the latter, this approach yields necessary conditions of intentional identity. Yet to my mind it fails to provide sufficient conditions, for it unsuccessfully faces the same counterexample as the 'pretense' approach. I guess that Manning's (2015) 'simulationist' approach suffers from the same problem.

be the case that another antirealist option is put forward that does not present the same troubles as those presented by the previous options. Yet such troubles point to a common general worry that all antirealist solutions, whether actual or possible, raise. All of them in fact share the assumption that thinking about the same thing involves *co-identifying* (unsuccessfully, in the case of nonexistent 'items') the same thing.[9] Yet this assumption is worrisome. It is clear how to rank together mental arrows of identification when they actually and literally point towards the same target, an object that is out there. For in this case, in being such a target, that object attracts such mental arrows, by letting them somehow depend on it. Yet it is less clear how to rank together mental arrows of identification when there actually is no such target; speaking of arrows (co)-pointing towards something here seems to be only an obscure metaphor. If we share the Kripkean lesson on these matters as applied to thoughts, a descriptivist account of that co-identification is ruled out.[10] Moreover, any alternative non-descriptivist account seems to be just a rather artificial way of ranking together disparate identification practices. In this respect, speaking here of a "common focus" for the thoughts involved, as Geach did (see Geach 1980, 147), merely amounts to putting the cart before the horse, by again trying to meaning literally what is just an obscure metaphor, namely speaking of thoughts as being *codirected*. If there literally is no target, how can mental arrows be codirected upon 'it'?

5. An alternative approach

Let me try a different approach that does not share the above problematic assumption. Suppose that, as many externalists say (cf. McDowell 1994; 1998, McGinn 1989), thinking about something has not to do with identifying it, if thinking is instead *constituted* by that very something, the thing that (*inter alia*) *individuates* it.

---

9 As Friend explicitly acknowledges, by talking of the problem of intentional identity as a problem of *co-identification* (see Friend 2014, 308).

10 For troubles about a descriptivist account of the intentionality of thoughts, cf. Sainsbury & Tye (2012, chap. 1). Notoriously, Geach himself (1980) rejected a descriptivist solution of his puzzle.

In this paper, I cannot argue in favor of this new assumption,[11] yet I can show its utility for my present purposes. For in virtue of this assumption, I can find an alternative solution to the problem of intentional identity. Instead of looking for a solution that mobilizes a sort of identity in the thinking subjects' situation, as the antirealist solutions do, I may look for a solution that points to an identity in the objects that are thought-of by such subjects, provided that there are any such objects. First, thinking about the same thing, even if it does not exist, is guaranteed by i) the fact that it satisfies the *metaphysical* criterion of identity for that thing – that thing belongs to a certain *metaphysical* kind and things of that kind are identical iff (here follows what states the relevant identity criterion) – and ii) the fact that *there are* things of that kind, i.e., that we are *ontologically* committed to such things. For then, given that assumption, the thing in question coindividuates such thoughts. Second, clause ii) has the following consequences. In certain cases, the entity that coindividuates thoughts is precisely what such thoughts are about according to our phenomenology, i.e., to what appears to us,[12] sometimes even a nonexistent object of a certain kind, for we are ontologically committed to an object of that kind even if they do not exist (in a suitable first-order sense of existence). Yet in other cases, the entity that coindividuates thoughts is not what such thoughts are about according to our phenomenology, sometimes again a nonexistent object of a certain kind, for we are not ontologically committed to that object, independently of the fact that it does not exist (again, in a suitable first-order sense of existence).

In order to understand this point better, let me first of all agree with Crane's (2001; 2013) claim that from a *metaphysical* point of view, i.e., from the point of view concerning the *nature* of the thing of a certain kind (provided that there is any such thing), *intentional objects*, i.e., the objects thoughts are about, are *schematic* objects; that is, that they have no metaphysical nature *insofar as they are thought-of*. In other terms, no metaphysical conclusion as to the nature of intentional objects can be drawn from

---

11 I did it elsewhere: see Sacchi & Voltolini (2012); Voltolini (2015).

12 This phenomenological appearance may be meant as a *second-order belief* in the aboutness of one's thought (weak reading) or as a *feeling of directedness*, leading to a phenomenological conception of intentionality itself (strong reading). In this paper, I remain neutral on this issue.

the fact that thoughts have such objects. It just phenomenologically appears to us that they are what our thoughts are about. As Sainsbury (2010) comments, this means that *intentionalia* are no *exotica*, that is, they are not metaphysically bizarre entities that possess a *sui generis* metaphysical nature individuating all of them as *intentionalia* (e.g. Brentanian immanent entities, Meinongian entities, and the like).

Metaphysically speaking, this claim has a main consequence; namely, that *intentionalia* may have a thought-independent metaphysical nature and such a nature may be various. As Crane himself admits, different *intentionalia* have different thought-independent such natures (cf. Crane 2001, 17; 2013, 92). Yet the claim has also another yet *ontological* consequence, i.e., a consequence concerning the overall ontological inventory;[13] namely, that there are just those *intentionalia* whose thought-independent metaphysical nature is such that we are ontologically committed to objects of that nature. Now, given the previous assumption that a thought is constituted by what it is about, it further follows that only in that case a thought is individuated by its *intentionale*, for the latter constitutes the former. Otherwise, since a thought cannot be constituted by an *intentionale* if there is no such object, not only it is not really about such an object, but it is also individuated by the other things that actually constitute it. Such things determine for it a proposition-like *intentional content*, the semantically relevant element it turns out to involve; quite likely, a general content of the kind, there is a (unique) *F* that *G*s.[14] In this case, properly speaking there is no *intentionale* for two thoughts that merely share their intentional content. These thoughts just think the same insofar as they are constituted by the same intentional content, provided of course again that one has a suitable criterion of identity for that content. Since that content is proposition-

[13] For this conception of the relationship between metaphysics and ontology cf. e.g. Thomasson (1999).

[14] Since Crane does not endorse a conception of thoughts as being constituted by their *intentionalia*, he does not endorse this ontological consequence either. For him (see Crane 2001, 2013), thoughts are always about their *intentionalia*, independently of whether such *intentionalia* figure in the overall ontological inventory. Whereas for me, thought are about *intentionalia* just in case they figure in that inventory. As a further consequence, for Crane aboutness is a nonrelational property of thoughts, whereas for me it is a relational property of them. I have criticized Crane's approach elsewhere; see Voltolini (2009; 2013).

like, in order to assess that two thoughts think the same one will have to resort to one's favorite theory of propositions.

In order to clarify this issue, let me start by providing some examples. First, in reflecting on Elizabeth II, one may think both about her Majesty and about the number Two, as phenomenology tells one: it appears to one that one's thoughts are respectively about them. Both objects are *intentionalia* in a schematic sense: they have no metaphysical nature insofar as they are what those thoughts are about. They do have a thought-independent metaphysical nature, yet that nature is utterly different: her Majesty is a concrete entity (notably, a person), number Two is an abstract entity (notably, a number). Suppose also that in the overall ontological inventory there are not only *concreta*, persons in particular, for ontological skepticism about them seems to be self-defeating, but also *abstracta*, numbers in particular, as many of us are disposed to believe by relying e.g. on some indispensability argument. Thus, we are ontologically committed to both these *intentionalia*. Given the new assumption about constitution, both *intentionalia* respectively individuate our thoughts.

Yet second, in reflecting on Elizabeth II, one may think both about her Majesty and about her entelecheia, i.e., her teleological end or perfection, as phenomenology again tells one. Both objects are again *intentionalia* in a schematic sense. Yet their thought-independent nature is different: as we have already seen, her Majesty is a concrete entity (a person), yet her entelecheia is a (false) nonnatural posit endowed with causal powers insofar as it allegedly guides an organism's development. As we have seen, moreover, we accept persons in the overall ontological inventory. Yet we normally agree in rejecting non-natural posits endowed with causal powers, hence entelecheias as well. Thus, in the light of the new assumption about constitution, while Elizabeth II individuates the thought that according to phenomenology – correctly – is about her, her entelechia does not individuate the thought that according to phenomenology – incorrectly – is about it. Another entity indeed plays that job – plausibly, a proposition-like entelecheian intentional content, a content displaying the problematic determinations that makes entelecheias ontologically implausible.[15]

---

[15] In (2001, 30), Crane defended a similar strategy as to intentional content, but for the fact that for him that content is nonconceptual, hence nonpropositional. This may depend on the fact that in his strategy (see Crane 2001; 2013), nonexistent *intentionalia*

Once the above is the case, moreover, as to intentional identity I can say that our thoughts are coindividuated by the *intentionale* they share, just in case we are ontologically committed to objects with the specific nature of such *intentionale*, even if that nature makes them to be nonexistent entities, in a suitable first-order sense of existence. Otherwise, those thoughts are still coindividuated, yet not by their alleged *intentionale* of a certain nature, for there really is no such thing, but by entities of a different kind. More precisely, two thoughts are individuated by the same object they are about, just as phenomenology conveys to one, iff: i) the object the first thought thinks about according to one's phenomenology and the object the second thought thinks about according to one's phenomenology are metaphysically the same object; ii) there really is such an object. If there is no such object, then they are not really about it. Thus, although as to them there is no proper intentional identity, they are coindividuated by another kind of thing that there is; namely, what determines their proposition-like identical intentional content. Both thoughts indeed mobilize a content of the kind, there is a (unique) *F* that *G*s.

Once again, let me clarify matters by means of examples. Let me start with three hopefully uncontroversial cases. To begin with, suppose Jules is thinking about the set of the cordates and Jim is thinking about the set of the renates, as phenomenology respectively tells them. Do they really think about the same thing? Metaphysically speaking, first, the things that according to their phenomenology the two guys are thinking about are *abstracta* of a certain kind; namely, sets. Second, as is well known, sets have the following identity criterion: $\{x: Fx\} = \{x: Gx\}$ iff $\forall x(Fx \leftrightarrow Gx)$. According to this criterion, the set of the cordates and the set of the renates are the same set, for they are coextensional. Ontologically speaking, moreover, there are sets, or so most of us believe, for sets merely supervene on

never figure in the overall ontological inventory. Whereas I hold that the question of whether nonexistent *intentionalia* figure in that inventory is a case by case question, to be answered differently depending on whether the different kinds of nonexistent *intentionalia* that are at stake are accepted in that inventory – positively as to *ficta* and mere *possibilia*, negatively as to *impossibilia*. In a nutshell, in the positive case, instead of having a nonconceptual content we have a nonexistent intentional object that figures in the overall ontological domain, while in the negative case, since there is no such object, plausibly enough the intentional content one's thought possesses in that case is propositionally structured.

their members. Hence, Jules and Jim really think about the same thing, i.e., the very set that coindividuates their thoughts.

Mutatis mutandis, a similar story must be told as to thinking about numbers or about any other *abstracta*. In thinking about 2 and about II, as phenomenology respectively tells them, Jules and Jim are respectively thinking about the same thing that individuates both thoughts, i.e., a certain number, provided that: i) 2 and II are the same according to the identity criterion for numbers; ii) there are numbers (as we already acknowledged).

Finally, do Jules and Jim really think about the same thing when Jules is thinking about Elizabeth II and Jim is thinking about that lady over there (he is actually facing her Majesty), as their phenomenology respectively tells them? Metaphysically speaking, first, the things that according to their phenomenology the two guys are thinking about are things belonging to *concreta*, namely, things that *may* be spatiotemporal occupiers.[16] Second, *concreta* have the following identity criterion: *a* and *b* are the same *concretum* iff, in all worlds in which they exist, they have the same spatiotemporal profile. [17] According to this criterion, Elizabeth II and that lady are the same *concretum*. Ontologically speaking, moreover, there are *concreta*, as we have already acknowledged. Hence, Jules and Jim really think about the same thing, which coindividuates their thoughts.

Armed with these reflections, let me pass to scrutinize the controversial cases. First, suppose Jules is thinking about Sherlock and Jim is thinking about Holmes, as phenomenology respectively tells them. Do they really think about the same thing? Metaphysically speaking, first, the things that according to their phenomenology the two guys are thinking about are things belonging, if artefactualists about fictional characters are right,[18] to a particular kind of *abstracta*; namely, *ficta*, taken as abstract artefacts. Second, one may suppose that *ficta* have a precise identity criterion. For instance, one may say that *a* and *b* are the same *fictum* iff they share the same set of properties attributed to them in stories and they come out of the same make-believe process of a certain kind that there is a certain

16 For this conception, cf. Cocchiarella (1982). See also Priest (2016).

17 For this conception, see Cocchiarella (1982).

18 See paradigmatically Thomasson (1999).

individual.[19] According to this criterion, Sherlock and Holmes are the same *fictum*. Suppose further that we agree that there are *ficta*, for instance because dispensing with them is false parsimony, or because they are indispensable (dispensing with them entails dispensing with other entities we independently admit).[20] Granted, *qua fictum*, Sherlock aka Holmes does not exist. Yet since *qua fictum* it belongs to the overall ontological inventory, the fact that it does not exist, in a suitable first-order sense of existence, does not prevent Jules and Jim from really thinking about the same thing, which coindividuates their thoughts.

Second, suppose Jules is hallucinating Nessie and Jim is hallucinating that monster over there, as phenomenology respectively tells them. Granted, it may not appear to them that they are hallucinating, yet it appears respectively to them that their mental states are about Nessie and that monster. Do they really think of the same thing? Metaphysically speaking, first, the things that according to their phenomenology the two guys are thinking about are things that might have existed even if they do not exist (again in a suitable first-order sense of existence): mere *possibilia*. Hence again, they are allegedly thinking about things belonging to *concreta*, namely, things that *may* be spatiotemporal occupiers. Second, as we saw before, *concreta* have the following identity criterion: *a* and *b* are the same *concretum* iff, in all worlds in which they exist, they have the same spatiotemporal profile.[21] According to this criterion, if they share that profile in all worlds in which they exist, Nessie and that monster are the same *concretum*; namely, a certain mere *possibile*. Suppose further that we agree that in the overall ontological domain there are not only *concreta*, but also mere *possibilia*. For dispensing with mere *possibilia* would be false parsimony insofar as they are of the same metaphysical kind as actual *possibilia* (like Elizabeth II) that we already admit (cf. Lewis 1986; Voltolini 2007). Now granted, that mere *possibile* does not

---

[19] See Voltolini (2006). Caplan and Muller (2015) hold that *ficta* satisfy no criterion of identity: they are brutally identical. One may accept that basic entities satisfy no such criterion. Yet, as Caplan, Muller & Sanson (2017) admit, if *ficta* were no basic entities (as is probably the case), such a satisfaction would obtain.

[20] For these arguments, cf. Thomasson (1999) and Voltolini (2006) respectively.

[21] For a similar identity criterion for mere *possibilia* cf. Zalta (1983, 75).

exist. Yet since *qua possibile* it belongs to the overall ontological inventory, the fact that it does not exist, in a suitable first-order sense of existence, does not prevent Jules and Jim from really thinking about the same thing, which coindividuates their thoughts.[22]

To be sure, one might object that one can hardly apply this criterion to concrete cases involving mere *possibilia*, since *qua* mere *possibilia* Nessie and that monster have actually just a few properties among those that are predicated of them. Indeed, they lack all the so-called *existence-entailing* properties (cf. again Cocchiarella 1982). Neither Nessie nor that monster, for example, actually swims in Loch Ness, has a shiny skin, and attacks boats; they have such properties only in the merely possible worlds where they exist. So, how can one settle whether they are the same or not?

Yet the objector forgets that among the existence-entailing properties that are falsely ascribed at one and the same time to Nessie and that monster, there are also spatiotemporal properties, which are the properties mobilized in the identity criterion for mere *possibilia*. Now, if contrary to the above hypothesis, in the closest merely possible world where both exist, Nessie and that monster do not share such properties, then they are different entities. Suppose that, while facing Loch Ness, I hallucinate Nessie to be *here* while you hallucinate that monster to be *there*. This means that, in the closest merely possible world where both exist, Nessie has a certain spatiotemporal profile while that monster has another one. Hence, they are different entities. This is surely the case of the aforementioned Vulcan and Twin-Vulcan, which in the closest merely possible world where both exist have a spatiotemporally utterly different orbit.[23]

Third, suppose Jules is thinking about Twardy, the wooden cannon made of steel, and Jim is thinking about Cazmy, the steel cannon made of wood, as phenomenology respectively tells them. Do they really think about the same thing? Metaphysically speaking, first, the things that according to their phenomenology the two guys are thinking about belong to *impossible* objects. Second, *impossibilia* have the following identity criterion: *a* and *b* are the same *impossibile* iff, in the impossible worlds in which

---

22 For similar views, see Glick (2012) and Pagin (2014).

23 Thus *pace* Dennett (1968, 337), Jules and Jim do not have to share all their beliefs on a certain subject, but just the relevant spatiotemporal beliefs.

they exist, they share all their properties. According to this criterion, Twardy and Cazmy are the same *impossibile*. Yet ontologically speaking, as most people think, in the overall ontological domain there are no *impossibilia*, independently of the fact that they do not exist (actually, they cannot exist) in a suitable first-order sense of existence. For even allowing for impossible worlds[24] does not suffice for allowing for such things: impossible worlds are just one kind of impossible entities, impossible objects are another one. If this is the case, then their phenomenology notwithstanding, Jules and Jim are not really thinking about the same *intentionale*. Properly speaking, therefore, as to their thoughts there is no intentional identity. Yet they still think the same, for something else really (co)individuates such thoughts, i.e., something that really figures in the overall ontological inventory; possibly, a certain Twardian/Cazmyan propositional-like intentional content to the effect that there is a certain cannon that is both made of steel and made of wood.

This last case is very important. For as to *impossibilia*, a certain[25] amount of consensus holds as to the fact that an antirealist account must be adopted wrt a problem of intentional identity. Different people here think the same, yet this sameness has not to do with the kind of thing they seem to be thinking about, an *impossibile*, for in point of fact there is no thing of that kind.

Now, this antirealist account may be exported as to any *intentionale* of another kind for which it turned out that an antirealist approach is correct, for there is no such kind. Thus, not only if it were right *not* to believe in *ficta* and in mere *possibilia*, but also if it were right not to believe in sets, numbers and any other *abstracta*, a similar antirealist story as to the problem of intentional identity should be told in all such cases. Nevertheless, it remains that whatever cases support an antirealist stance, the relevant thoughts think the same thing not because they co-identify something, as antirealists traditionally believe (see the previous Section), but because they are constituted by the same ontologically relevant stuff. Simply, that stuff is not the thing is one is thinking about according to one's phenomenology, for there really is no such thing.

---

[24] As Priest (2016), Yagisawa (2008), and Berto (2012) do.

[25] Admittedly, a limited one: Impossibilists and Meinongians of any sort would disagree.

If this is the case, then a general solution to the problem of intentional identity can be found. First, one must check the nature of the entities people really think about when they are entertaining certain thoughts, insofar as by being really there in the overall ontological inventory, such entities constitute those thoughts. Second, if such entities are metaphysically identical, then those thoughts are really about the same thing, even if its having that very nature forces that thing not to exist, in a suitable first-order sense of existence. Third, if the things people seem to be thinking about are not really there in the overall ontological inventory, then their thoughts are not about the same *intentionale*, yet they may still be thoughts involving the same entities, i.e., what really constitutes such thoughts instead of that *intentionale*: a certain proposition-like intentional content.

This solution is neither realist nor antirealist *in general*. According to it, different thoughts can be about the same nonexistent object *only if* that object belongs to a kind such that there are things of that kind in the overall ontological inventory.

At this point, one may arise the following doubt. Since by following Crane I said that, *qua* object of one's thought, an *intentionale* has no metaphysical nature, what prevents one in the problematic cases from *not* ontologically dispensing with that *intentionale*, by metaphysically *identifying* it with an intentional content that really *belongs* to the overall ontological inventory? Thus, one might apply also to these cases the positive solution I have proposed for the problem of intentional identity. E.g. as to *impossibilia*, what prevents one from ontologically allowing for the relevant *intentionale*, by metaphysically identifying it not with an object that cannot exist, but with the relevant intentional content – a content that is impossible to satisfy – that is really there, so as to say that the relevant thoughts are really both about that content?

To be sure, such a move would save the intuition that intentionality is *uniform*, so that the thought's aboutness is to be taken at face value: *whenever* it seems to one that one's thought is about something, that thought is *really* so.[26] For there really is something that thought is about, simply its metaphysical nature is not the kind of nature one would have expected. E.g. in the case of *impossibilia*, whenever it seems to one that one is thinking about something, there really is something one's thought is about, yet it is

[26] On this intuition, cf. e.g. Spinelli (2016, 94).

not an impossible object but an ontologically palatable intentional content that is simply impossible to satisfy.

Yet unfortunately, there is a reason for not *utterly* endorsing such a welcome result.[27] In some cases, people seem to think about *different* objects. But if what they think about were intentional contents, it might implausibly turn out that they are thinking about the *same* entity. For instance, suppose that Jules seems to think about the set of sets that are not members of themselves, whereas Jim seems to think about the property of being a property that does not apply to itself. Intuitively, they seem to think about different things. Yet suppose one claimed that what they really think about are intentional contents and, owing to one's particular theory of propositions, such intentional contents are the same proposition. This would mean that, appearances notwithstanding, they are thinking of the same *intentionale*. Now, it may well be the case that phenomenology fails to be a good guide to metaphysics. One may believe that what one is thinking is an object of a *certain* metaphysical kind, whereas in point of fact it is an object of *another* metaphysical kind. Consider e.g. children's thinking of Santa Claus, who they assume to be a concrete entity whereas in point of fact it is a kind of fictional entity, a mythological entity. Yet phenomenology is a reliable guide to metaphysics at least as to the *number* of objects, if any, that are involved by one's thoughts. Now, in the case at issue one phenomenologically counts two objects – the paradoxical set and the paradoxical property – where metaphysically speaking just one entity would be at stake, a certain proposition. This suggests that in that case at least, it is better not to identify intentional objects with intentional contents, while going on saying that, since the alleged intentional object does not really figure in the overall ontological domain, instead of having that object, one's thought merely has an intentional content.

To complete the picture, let me conclude this section by saying that settling the kind of entity involved and its metaphysical relationship with the qualifying properties that are predicated of it also determines the kind of *de re* report of intentional identity that is true of it *just in case* that entity really belongs to the overall ontological inventory. As we saw before, reports like (1) can be read either in a strong or in a weak form, precisely depending on whether the entity they allegedly truly quantify

[27] In the final section, we will see that there is a *partial* way of endorsing it.

over possesses the property ascribed to it in the relevant report. Since *ficta*, if there are any, actually possess the properties that are predicated of them within stories, the kind of *de re* reports that turn out to be true of them is of the (1a)-form, the strong reading. For instance, we can truly read:

(2) Magica De Spell is a witch such that Hob believes that she hates Uncle Scrooge, whereas Nob believes that she likes Donald Duck

as:

(2a) $\exists\alpha(W(\alpha) \wedge$ Hob believes that $H(\alpha,U) \wedge$ Nob believes that $L(\alpha,D))$.

For Magica De Spell is a *fictum* that actually possesses the property of *being a witch* predicated of her in the Disney stories. Yet since mere *possibilia*, if there are any, do not actually possess the existence-entailing properties ascribed to them in the thoughts about them, the kind of *de re* reports that turn out to be true of them is of the (1b)-form, the weak reading. We can truly read (1), which is about a *merely possible* witch, merely as (1b).

6. How one can mock-think about the same thing, and a final consequence

On the basis of the above solution to the general problem of intentional identity, I can also provide a treatment of the particular problem of how in pretense one can think about the same thing, when in the overall ontological inventory there is no such thing. Once I endorse the above solution to the problem of intentional identity, this further problem becomes just a specific yet related problem, whose treatment provides no solution to the general problem of intentional identity, yet it may shed light to the above solution itself by integrating it.

To begin with, as Evans (1982) and Walton (1990) have shown us, pretense involves two forms of *as-if* activity. *Of* certain things that are there, one can make *as if* they had certain properties. But one can also make *as if*

there were things of certain kinds having certain properties, even if there are no such things.[28] Now, in so pretending that there are such things, one's thoughts about them are also pretended thoughts. Following Evans' terminology, let me call them *mock-thoughts*. Mock-thoughts are therefore to be distinguished from thoughts that are about nonexistent things, in a suitable first-order sense of existence. For unlike the latter things, the former things do not utterly figure in the overall ontological domain; in one's mock-thought, one merely pretends that such things figure in that domain.[29]

Now, in the light of the above solution of the problem of intentional identity, it is easy to say when two mock-thoughts are mock-thoughts about the same thing, by thus pretending to be about the same thing even if there is no such thing. In such mock-thoughts, one must not only pretend that there are certain things, but also that they are things of a certain kind, thereby pretending to satisfy an identity criterion for things of that kind. No identity in the kind of pretense one is engaged in is required any longer, as it was the case in the 'pretense' antirealist approach to the problem of intentional identity we saw in Section 3. For even if such kinds of pretense differ, the identity criterion people are pretending to adopt for the kind of things they mock-think about suffices for those mock-thoughts to be about the same thing.

Let me illustrate this point again by means of examples. First, in their respective pretend plays, Homer mock-thought about Odysseus and Dante mock-thought about Ulysses. Clearly enough, those plays did not by themselves affect the overall ontological inventory. Homer just pretended that there was such a thing as Odysseus, while Dante just pretended that there was such a thing as Ulysses. Nevertheless, did they mock-think about the same thing? Both pretended not only that there is a certain thing, but also that such a thing is both a *concretum* and satisfies the identity criteria for

---

[28] The two forms of pretense respectively occur in what Evans (1982) calls *conservative* and *creative* make-believe games. The former correspond in part to Walton's (1993) *prop-oriented* make-believe games.

[29] Mock-thoughts may even be unintentional, in the sense that one may not realize that one's mental activity merely amounts to a mock-thought. For those who reject fictional entities and think that in fiction one merely pretend that there are certain things, the Santa Claus case (along with all myth-involving cases) may be reinterpreted as a case in which children inadvertently pretend that there is a guy who does such and such.

concreta. If they had not pretended the latter at least implicitly, they would have rather pretended that such a thing is an impossible entity having incompatible determinations. But in their plays there is no sign of that form of pretense. Now, as we already know, two things are the same *concretum* iff they share the same spatiotemporal profile in all worlds in which they exist. Hence, two things are pretended to be the same *concretum* iff they are pretended to share that spatiotemporal profile. As a matter of fact, the pretended spatiotemporal profile of what Homer and Dante mock-thought is the same. Dante simply reprises Homer's narration by adding new events that involve Ulysses after his return home. Once he started navigating again, Ulysses bypasses the Hercules' pillars and eventually dies in the Southern hemisphere. Hence, both Homer and Dante mock-thought about the same thing. By the same reasoning, one may show that (Joyce's intentions notwithstanding) Homer and Joyce did not mock-think about the same thing. For obviously enough, Odysseus' pretended spatiotemporal profile and Bloom's pretended spatiotemporal profile do not coincide.

Second, *Hamlet* notoriously contains a play within a play, *The Murder of Gonzago*. In *Hamlet*'s nesting play, Shakespeare pretends that Gonzago, the hero of the nested play, is a fictional character. Granted, we know very little about Kyd's *Ur-Hamlet*. Yet let me suppose, for argument's sake, that it also contained a play within a play whose hero is Ur-Gonzago, let me call it *The Murder of Ur-Gonzago*. So in *Ur-Hamlet*'s nesting play, Kyd pretends that Ur-Gonzago is a fictional character. Do they mock-think about the same thing? Well, the mock-thoughts featuring the nesting plays involve fictional characters coming out of the nested plays (unlike mock-thoughts that would feature the nested plays, which just involve concrete individuals: in the above nested plays, *The Murder of Gonzago* and *The Murder of Ur-Gonzago*, both Gonzago and Ur-Gonzago are men, not fictional characters). Thus, both Shakespeare and Kyd pretend that there is a certain thing, that such a thing is a *fictum* and (at least implicitly) that it satisfies the identity criteria for *ficta*. As we have seen before, one may provide various identity criteria for *ficta*. For instance, one may say that two things are the same *fictum* iff they share the same set of properties attributed to them in stories and they come out of the same make-believe process that there is a certain individual. Hence, two things are pretended to be the same *fictum* iff they are pretended to share that very set and their

process of generation. As a result, one may show that (Shakespeare's intentions notwithstanding?) Shakespeare and Kyd do not mock-think about the same fictional character. For, since *Ur-Hamlet* does not coincide with *Hamlet*, either the properties respectively mock-attributed to Gonzago and Ur-Gonzago in their nesting plays or the respective mock-generation processes in those plays plausibly differ.

At this point, on the basis of this treatment, I may supply the general solution I have put forward to the problem of intentional identity with an additional fictionalist account. This account holds just for the cases in which there really are no *intentionalia* figuring as the relevant thoughts' constituents, for in the overall ontological domain there really are no things of the metaphysical kinds such *intentionalia* belong to. I may indeed say that in such cases, while it is still really the case that the relevant thoughts think the same insofar as they are constituted by the same proposition-like intentional content, it is just a shallow pretense that such thoughts are about the same *intentionale*, as phenomenology suggests. As one may put it, what is phenomenologically the case is in such cases just what it is *fictionally* the case.

This way of putting things has an additional advantage. For it partially saves the *uniformity* intuition we considered in the previous Section that, in a sense, both thoughts that in my account are constituted by *intentionalia* and thoughts not so constituted are *about* something. For the former are such that they are *really* about something, yet the latter are such that they are *merely fictionally* about something.

Cases in which different thoughts are allegedly about the same *impossibile* are paradigmatic instances of this situation. Coming back to a previous example, when allegedly thinking about Twardy and Cazmy respectively, it is just fictionally the case that Jules and Jim think about the same *impossibile*, as phenomenology suggests. For it is instead really the case that they are both mobilizing in their thoughts a certain Twardian/Cazmyan proposition-like intentional content. Clearly enough, such cases definitely are not the only cases of this kind (situations involving thoughts allegedly about imaginary companions, idiosyncratic posits, indeterminate fictional characters are other clear instances).[30]

[30] For these examples, see Kroon (2011; 2013; 2015).

Acknowledgments

This paper was originally presented at the workshop on *Fiction and Intentionality*, University of Turin, October 6-7 2015, Turin. I thank all the participants for their stimulating questions. Moreover, I thank Fred Kroon, as well as two anonymous referees, for their criticisms and remarks.

References

BERTO, F. (2012): *Existence as a Real Property*. Dordrecht: Springer.

CAPLAN, B. & MULLER, C. (2015): Brutal Identity. In: Brock, S. & Everett, A. (eds.): *Fictional Objects*. Oxford: Oxford University Press, 174-207.

CAPLAN, B., MULLER, C., & SANSON, D. (2017): Counting Again. *Grazer Philosophische Studien* 94, 69-82.

COCCHIARELLA, N. (1982): Meinong Reconstructed *Versus* Early Russell Reconstructed. *Journal of Philosophical Logic* 11, 183-214.

CRANE, T. (2001): *Elements of Mind*. Oxford: Oxford University Press.

CRANE, T. (2013): *The Objects of Thought*. Oxford: Oxford University Press.

DENNETT, D. (1968): Geach on Intentional Identity. *The Journal of Philosophy* 65, 335-341.

DONNELLAN, K. (1974): Speaking of Nothing. *The Philosophical Review* 83, 3-31.

EDELBERG, W. (1992): Intentional Identity and the Attitudes. *Linguistics and Philosophy* 15, 561-596.

EVANS, G. (1982): *The Varieties of Reference*. Oxford: Clarendon Press.

EVERETT, A. (2000): Referentialism and Empty Names. In: Everett, A. & Hofweber, T. (eds.): *Empty Names, Fiction, and Puzzles of Non-Existence*. Stanford: CSLI, 37-60.

EVERETT, A. (2013): *The Nonexistent*. Oxford: Oxford University Press.

FRIEND, S. (2014): Notions of Nothing. In: García-Carpintero, M. & Martí, G. (eds.): *Empty Representations*. Oxford: Oxford University Press, 307-332.

GARCÍA-CARPINTERO, M. (2016): Pretend Reference and Coreference. In: Capone, A., Kiefer, F. & Lo Piparo, F. (eds.): *Indirect Reports and Pragmatics*. Dordrecht: Springer, 333-358.

GEACH, P. T. (1980): Intentional Identity. In: *Logic Matters*. Oxford: Blackwell, 146-153.

GLICK, E. (2012): A Modal Approach to Intentional Identity. *Noûs* 46, 386-399.

HOWELL, R. (2015): Objects of Fiction and Objects of Thought. In: Brock, S. & Everett, A. (eds.): *Fictional Objects*. Oxford: Oxford University Press, 41-70.

KROON, F. (2001): Fictionalism and the Informativeness of Identity. *Philosophical Studies* 106, 197-225.

KROON, F. (2011): The Fiction of Creationism. In: Lihoreau, F. (ed.): *Truth in Fiction*. Frankfurt: Ontos, 203-222.

KROON, F. (2013): The Social Character of Fictional Entities. In: Barbero, C., Ferraris, M., & Voltolini, A. (eds.): *From Fictionalism to Realism*. Cambridge: Cambridge Scholars, 87-110.

KROON, F. (2015): Creationism and the Problem of Indiscernible Fictional Objects. In: Brock, S. & Everett, A. (eds.): *Fictional Objects*. Oxford: Oxford University Press, 147-173.

LEWIS, D. (1986): *Counterfactuals*. Oxford: Blackwell.

MANNING, L. (2015): No Identity without an Entity. *Pacific Philosophical Quarterly* 96, 279-305.

MCDOWELL, J. (1994): *Mind and World*. Cambridge (Mass.): Harvard University Press.

MCDOWELL, J. (1998): Having the World in View: Sellars, Kant and Intentionality. *The Journal of Philosophy*, 431-490.

MCGINN, C. (1989): *Mental Content*. Oxford: Blackwell.

PAGIN, P. (2014): Intersubjective Intentional Identity. In: García-Carpintero, M. & Martí, G. (eds.): *Empty Representations*. Oxford: Oxford University Press, 91-113.

PAUTZ, A. (2008): Fictional Coreference as a Problem for the Pretense Theory. *Philosophical Studies* 141, 147-156.

PRIEST, G. (2016): *Towards Non-Being: the Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.

SACCHI, E. & VOLTOLINI, A. (2012): To Think Is to Have Something in One's Thought. *Quaestio* 12, 395-422.

SAINSBURY, M. (2005): *Reference without Referents*. Oxford: Oxford University Press.

SAINSBURY, M. (2010): Intentionality without *Exotica*. In: Jeshion, R. (ed.): *New Essays on Singular Thoughts*. Oxford: Oxford University Press, 300-317.

SAINSBURY, M. & TYE, M. (2012): *Seven Puzzles of Thought and How to Solve Them*. Oxford: Oxford University Press.

SPINELLI, N. (2016): What It Is to Be an Intentional Object. *Disputatio* 8, 93-112.

TAYLOR, K. A. (2003): *Reference and the Rational Mind*. Stanford: CSLI.

THOMASSON, A. L. (1999): *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

VOLTOLINI, A. (2006): *How Ficta Follow Fiction*. Dordrecht: Springer.

VOLTOLINI, A. (2007): How to Allow for *Intentionalia* in the Jungle. *Russell* 27, 86-105.

VOLTOLINI, A. (2009): Consequences of Schematism. *Phenomenology and the Cognitive Sciences* 8, 135-150.

VOLTOLINI, A. (2013): There Are *Intentionalia* of Which It Is True That Such Objects Do Not Exist. *International Journal of Philosophical Studies* 21, 394-414.

VOLTOLINI, A. (2015): What's in a (Mental) Picture? In: Torza, A. (ed.): *Quantifiers, Quantifiers, and Quantifiers*. Dordrecht: Springer, 389-406.

WALTON, K. L. (1990): *Mimesis as Make-Believe*. Cambridge (Mass.): Harvard University Press.

WALTON, K. L. (1993): Metaphor and Prop Oriented Make-Believe. *European Journal of Philosophy* 1, 39-57.

YAGISAWA, T. (2010): *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.

ZALTA, E. (1983): *Abstract Objects*. Dordrecht: Reidel.

Organon F 24 (3) 2017: 308-338

Fictional Names, Fictional Characters and Persons Referred to in Narrative Fiction

PETR KOŤÁTKO[1]

ABSTRACT: The paper is based on a strict distinction between the notion of a person referred to by a fictional name, as uttered within a text of narrative fiction, and the notion of a fictional character. The literary functions of such a text require the reader to interpret the occurences of a fictional name as records of utterances of that name by the narrator, referring to that individual which has been assigned that name at the beginning of the chain to which these utterances belong. This, according to the author's view, provides proper basis also for interpretation of various kinds of extratextual use of fictional names. A literary character is, on the contrary, an element of a construction of a literary work and is identified by a set of requirements (e.g. of the kind mentioned above) imposed by the text's literary functions on the reader. The author attempts to justify the assumption that the referential function of fictional names so understood is to be interpreted as directed to the actual world (rather than to an artificial world created by the writer), to specify the (rather limited) role reserved for pretense within this approach, to explain the implications of this account of fictional characters for the dispute between realists and anti-realists in this field etc.

KEYWORDS: Fictional name – literary character – fictional world – pretense.

[1] Received: 3 April 2017 / Accepted: 27 July 2017

✉ Petr Koťátko
Institute of Philosophy
Czech Academy of Sciences
Jilská 1, 110 00 Prague 1, Czech Republic
e-mail: kotatko@flu.cas.cz

1. The notion of literary functions

What follows is an extended version of a paper originally delivered at a conference in Bratislava in October 2016.[2] Let me start in the same way as in that presentation, namely by a short comment on the conference title: *Semantics of fictional discourse*. I believe that it makes a good sense under a rather modest reading: there certainly are types or areas of discourse to which the term "fictional discourse" is quite naturally applicable and they certainly deserve careful semantic analysis. But the title can also be read as suggesting that there is a unique area of discourse called "fictional", for which we are equipped with a commonly shared bunch of intuitions, in which proper names function in specific way, pretense plays specific role, the illocutionary force of utterances is modified in specific way etc. I am afraid that this picture is itself a kind of fiction, perhaps useful, perhaps misleading, perhaps both, depending on a given context. What I can see are types of situations, like reading a literary text of narrative fiction, following theatre performance, speaking about literary or dramatic characters, listening to somebody's telling a joke etc., in which linguistic utterances fulfil specific functions which impose specific demands on the interpreters – and I cannot fail noticing that these functions and these demands are dramatically different. Without trying to interfere into the projects of my distinguished colleagues, I take it as a good reason for restraining my own aspirations – and my way of doing so in this paper will consist in focusing on *literary texts of narrative fiction*, the way we are supposed to interpret them and the discourse linked to them.

Now, even within this restricted field, I don't think that we should start with discussing problems like the status of literary characters, their identity conditions, their completeness or incompleteness, the role of fictional names etc. According to my opinion, the basic question providing proper framework for addressing such issues is: what does the reader have to do (to assume, to accept, to imagine) in order to allow the text of narrative fiction to fulfill its literary functions? Moreover, I believe that substantial part of the talk about literary characters, including metatextual claims like

---

[2] I am grateful to the organizers, Marián Zouhar and his colleagues from the *Institute of Philosophy of the Slovak Academy of Sciences*, for the invitation to that exceptional meeting, and to the participants for inspiring criticism.

"Emma Bovary ruined her husband", is precisely a talk about the demands just mentioned: about what we have to assume in order to allow Flaubert's text to fulfill its literary functions.

Perhaps I should say a bit more to explain the prominent role I assign to the texts' literary functions and to the demands they impose on the readers. Let me start with the trivial assumption that to read a text as a literary work of narrative fiction is to approach it as a bearer of certain literary functions and to make the interpretative moves required by these functions – the moves which will allow the text to function as such and such a piece of literature for us as its readers. The only thing which distinguishes literary functions from any other functions the text may have is that they together constitute the literary work represented by the text – indeed, I think it makes a good sense to approach the literary work as a structured complex of the text's literary functions. The same function which counts as literary within such a complex, because it, in this framework, contributes to the constitution of a literary work, need not count as literary in another context. For instance, one of the functions of Balzac's novel *The Splendors and Miseries of Courtesans*, making it (together with other functions) precisely that piece of literature it is, certainly is to show the situation in France in the restoration times, or, more specifically, to show what are the chances of a gifted but poor young man with high ambitions and not too strong moral scruples, trying to succeed in Paris in those times. It is symptomatic for Balzac's project of *Studies (or: Scenes) of Parisian life* that this function is supported by an extensive historical material and will be blocked or at least seriously undermined if we are not ready to accept this material as reliable – and on this basis not just to pretend to believe, but to believe that Balzac's narrator's detailed descriptions of the structure of the police apparatus, of the role of the bills of exchange within the financial system etc. are true. Another function which can certainly be ascribed to *Splendors* is to engage the reader's imagination, sensitivity and moral intuitions in following the tragic story of an ambitious young man – and a necessary condition for this function being efficient is to assume (this time in the *as if* mode)[3] the existence of this person and to accept as a matter of fact (in the

[3] In what folows I will use the appendix "AI" to indicate the *as if* mode, e.g. in the form "to believe$_{AI}$", "the assumption$_{AI}$" etc. Sometimes, in particular when referring to

same mode) that he did the things described in the text. This includes a series of cooperative moves: for instance we should enable the sentence "Lucien signed the bill without hesitation", as it appears in Balzac's text, to fulfill its specific literary function, namely to establish the fictional fact that Lucien de Rubempré signed the relevant bill without hesitation – which requires that we assume$_{AI}$ that precisely this happened in the actual world.[4]

The fact that Balzac's text imposes such requirements on the reader makes Lucien one of the *characters* of Balzac's novel. Correlatively, the author can be said to have *created the character* named "Lucien" precisely in the sense that he wrote a text whose literary functions require the moves just mentioned. This formulation, as you can see, does not include any explicit reference to the author's intentions and also does not ascribe any kind of pretense to him: it just speaks about the text's requiring some pretense from the *reader* (we will return to this point in Section 7). And finally, the character named "Lucien" can be said to *exist* (iii) in that sense that there *exists* (ii) a text with literary functions which require us to assume that there *exists* (i) a person with this name. The last sentence includes three occurrences of the term "to exist" which, apparently, should not be taken as referring to the same mode of existence. Let me take them in the reverted order:

---

other authors, I will use more common terms "to pretend", "to make-believe" or "to imagine", taking them as synonymous with "to believe$_{AI}$".

[4] I hope that these two examples of literary functions (together with the next one related to Beckett, in Section 2) will make it clear why I don't believe that it makes sense to strive at a general definition of a literary function, over and above the trivial remarks made at the beginning of this paragraph. Proposals with analogical aspirations made in the theory of fiction (like attempts to define fiction in terms of prompting an imaginative response) not only face obvious counterexamples but raise doubts about the prospects and value of any such enterprise (for a critical discussion see Friend 2008). In what follows I will focus on the function of presenting a story as told by a narrator, which (as we will occasionally see) can itself bear an inexhaustible variety of literary functions, each of them being potentially constitutive in that sense that it contributes to making a piece of narrative fiction the literary work it is. (This remark has been inspired by on an objection raised by one of the reviewers.)

(i) As readers of *Splendors*, we are supposed to assume$_{AI}$ that there exists *certain person* (identified in a way we will discuss later), obviously not as a literary character but in the most ordinary mundane sense – in which for instance you could claim that there is a Rolls-Royce in your garage and I would pretend to believe you.

(ii) We assume the existence of *a text* which functions (or has a potential to function) as a literary work. The explanation of this mode of existence will depend on our position concerning certain not quite trivial issues, like the type-token distinction, the nature of linguistic conventions (if we take text as a sequence of expressions interpreted according to semantic conventions of some language) and the text-work relation. And the latter will involve us, among other things, into disputes concerning the role of the empirical author's intentions and of the socio-cultural setting in the identification of the literary work (this issue will be briefly opened in Section 9).

(iii) The existence of *a literary character* is parasitic upon the existence of a text with certain literary functions, but not only that: it simply follows from the specification of these functions and it consists in these functions' requiring certain moves from the readers, in the first place accepting certain existential assumption in sense (i).

Now, if we admit that the text, its literary functions and their parameters are abstract entities, we should not have any problem with admitting that literary characters are abstract entities as well:[5] but we should be careful not to read into this claim more than what has been just said. To identify these entities is to specify certain moves required by literary functions of a literary text, similarly like to specify the meanings of expressions of a language spoken by some population is to describe the way they are used and interpreted in that population. As far as I can see, there is no need to make

[5] This corresponds to Amie Thomasson's claim that "to accept that Austen wrote certain sentences in a novel pretending to refer to one Emma Woodhouse (not referring back to any actual person), but deny that she created a fictional character, is a mere distortion of ordinary usage" (Thomasson 2003, 149). I would just replace the reference to the author's pretense by reference to reader's pretense required by functions of the text (cf. Section 7).

further moves providing either meanings or literary characters with some more substantial metaphysical status.

There are various kinds of objections which can be raised by the realists concerning fictional characters: I will occasionally react to some of them in connection with some specific issues, but let me mention at least one of them right now. I mean the fact that we can quantify over fictional characters, e.g. in sentences like: "There are more craven than brave fictional characters", which implies "There are fictional characters" (the example is borrowed from Friend 2007, 147). According to my understanding, what we claim when uttering such a sentence is this: "There are more cases in which a piece of fiction requires us to assume$_{AI}$ the existence of a craven person (as referred to in the fiction) than cases in which a piece of fiction requires us to assume$_{AI}$ the existence of a brave person." The notion of fictional character has not disappeared here, it is encoded in the clause "requires us to assume$_{AI}$ the existence of", and hence in the reference to requirements imposed on us by pieces of fiction. It should be clear that in this reading we do not interpret such quantified claims as made in the pretense-mode: they speak about pretense, but with full-blooded, unrestricted assertive force.

2. The principle *F*

I have suggested that the basic question we should start with, in order to create proper framework for the discussion about the role of proper names in the texts of narrative fiction, about the status of literary characters etc. is: what does the reader have to do in order to allow the text of narrative fiction to fulfill its literary functions?

Here is the general reply I propose as a starting point for discussing more specific topics:

/F/ The literary functions of a text of narrative fiction require that the reader approaches$_{AI}$ its sentences as records of utterances of an inhabitant of the actual world – the narrator, who tells us what happened in this world. [6]

[6] Obviously, this scheme applies only to narrative fiction in strict sense: the fictional content need not be presented to the reader as narrated, but, for instance, as a content of

Within this scheme, the interpretation of a text and its components, like fictional names, is pinned down to the *actual world* and to the narrator's utterances (taking place in this world – as the reader is supposed to assume$_{AI}$). Within the approach I am going to defend, this double linkage is crucial for the identification of the entities referred to in the texts of narrative fiction, for the solution of the completeness/incompleteness problem concerning these entities and for the explanation of completeness of the propositions expressed by sentences occurring in literary texts.

One might be tempted to add that the role of the reader further includes the presumption of the primary narrator's[7] *credibility*, which does not require justification, but can be withdrawn if the narrator proves to be (in some respect) unreliable.[8] But it need not be the case and the narration can be construed so that it does not give us any chance to rely on what the narrator says – and correspondingly, the narrator himself does not raise any claim for our confidence. For instance, let us imagine that the literary text presents the world as a universal chaos and the narrator, his narrative tools and his narrative performance are construed as part and product of this chaos, as it is in Samuel Beckett's late texts (cf. in particular Beckett 1979). In such a case, one of the basic literary functions of the text, namely to allow the reader to experience various aspects of the universal chaos, requires that the reader approaches the narrator as in principle unreliable – and Beckett's narrator himself repeatedly points to his unreliability concerning not only the truthfulness of his claims but also the meaningfulness of his utterances. In short, while Balzac's narrator proves his perfect condition on every page and certainly deserves the title "narrator in good shape", Beckett's narrator is the best example of the literary construct I suggest to call "narrator in decay" (cf. Koťátko 2016). This example shows that the question of the narrator's reliability should be solved within the

---

letters collected and published by an editor or as a content of somebody's consciousness to which we are given direct access. Cf. Chatman (1978, chap. 4).

[7] A narrator is called "primary" if her narrative performance is not embedded in a narrative performance of some other narrator.

[8] For instance, Felix Martínez-Bonati assumes that accepting this assumption is a necessay condition for the meanings of the narrator's utterances being fulfilled with images which constitute the fictional world of the literary work. Cf. Martínez-Bonati (1981, 31, 34, 129 et al.).

interpretation of particular texts rather than *a priori* on the level of the theory of narrative fiction. And I believe that several other issues frequently discussed by theorists of fiction should be approached in the same way, for instance the author–narrator distinction or the problem of the completeness–incompleteness of the narrated world and its inhabitants. Let me say a few words about the latter issue.

It seems indisputable that the literary functions of, let us say, Balzac's *Splendors*, require us to approach$_{AI}$ Ester, Lucien, abbé Herera alias Vautrin etc. as human beings which are, precisely like us, complete in all respects obligatory for this kind of entities – and to assume$_{AI}$ that the narrator provides us with an incomplete description of these complete entities, precisely like we do it in everyday conversation when speaking about our neighbors. And, since the literary functions of the text require us to assume$_{AI}$ that the people as well as the places, events etc. referred to are complete, they *are* complete, in the only relevant sense in which the question of their completeness or incompleteness can be raised.[9] If you ask how Balzac could have succeeded to create such complete entities, granted that he could provide us only with incomplete sets of descriptions, the reply is quite simple: he did so by writing a text whose functions require us to approach$_{AI}$ the entities referred to as complete. And, granted our principle *F*, approaching$_{AI}$ people, places, events etc. spoken about in Balzac's text as complete does not require any special move: it is included in our interpreting the text as speaking about the actual world – provided that we regard this world as complete (in that sense that any possible state of affairs either is or is not a fact in this world).

But precisely this last assumption cannot be generalized for all kinds of narrative fiction. It seems fairly right to claim that the persons referred to in Beckett's *Trilogy* are incomplete beings – indeterminate in substantial respects, including their personal identity: since it belongs to the literary functions of Beckett's text that it confronts us precisely with such a picture of human beings and of the world they inhabit. "My heroes are falling to bits," says Beckett, when explaining his difference from Kafka (cf. Shenker

---

[9] Cf. Stacie Friend's remark that "for the anti-realist there is no contrast between what Anna (Karenina) is really like and how we imagine her to be" (Friend 2007, 152). I would just modify the last clause to "how we are required (by the text's literary functions) to imagine her to be".

1953, 3), and the reader is not given any reason to assume that she could assemble complete human beings from these bits, if she were allowed to collect all of them (i.e. if she had access to all the relevant information missing in the text). This confrontation of Balzac and Beckett is intended to show that the question of completeness or incompleteness of the persons referred to in narrative fiction cannot be solved on the level of a general theory of literary fiction: it should be raised over and over again within the interpretation of particular texts.[10]

3. The actual world and the narrative fiction

So, I conclude that the general specification of the moves required from the reader by literary functions of a text of narrative fiction, as it is presented in our principle *F*, should not be complemented by any appendix concerning issues like the assumption$_{AI}$ of the primary narrator's credibility or the assumption$_{AI}$ of the completeness of the entities spoken about in the text. But there is a much more controversial aspect of the principle *F* than its apparent need for complementation, which deserves special attention because of its important implications. I mean the assumption that a text of narrative fiction directs our thought (as well as our imagination and our sensitivity) to the *actual world*, and that it does so quite straightforwardly (rather than through analogies, allusions etc.). According to this assumption, the author does not construct a new, artificial world: instead, she creates a text whose literary functions are anchored in and directed to the actual world and typically (but not necessarily) require us to assume$_{AI}$ that this world in some respects differs from what we believe to be the case outside the scope of the *as if* operator.[11]

This may look like a revolt against the familiar jargon of possible worlds, but let us recall the lesson given to us by the author who introduced

---

[10] Let me add, with gratitude, that I have benefited a lot from my discussion with Göran Rossholm on this topic (cf. Rossholm 2015).

[11] In other words: instead of saying "We readers imagine that what is actual is a story-world" (Currie 2003, 147) one should say "We readers imagine that the story takes place in the actual world".

possible worlds apparatus into modern semantics. In the opening paragraphs of *Naming and Necessity*, Saul Kripke attempts to calm down our philosophical temperament and imagination, which could be encouraged by the term "possible world" – by insisting that possible worlds are simply "total 'ways the world might have been'" (Kripke 1972, 18). And he makes us sure that if we have (any kind of) problems with the term "possible worlds", nothing will be lost if we speak about "possible states or histories of the world" instead. Following this advice, we can replace the popular claim that the author creates a new world (the fictional world of her novel) with a much less spectacular claim that the author invites us to imagine and $\text{accept}_{\text{AI}}$ as actual an alternative state of the world we live in.[12] I believe that in this way we can reduce the danger mentioned also by Kripke, namely the possibility that our theoretical work will collapse to solving problems generated by the apparatus we have chosen.[13]

For instance, if we resign on the creationist vocabulary of fictional worlds, we will not have to make intricate theoretical moves to explain how fiction and the thought and imagination generated by it works, in contrast to how our thought and imagination works when directed to the actual world – and then make other no less intricate moves to solve the problems generated by this construction, e.g. to reconcile our doctrine with the hardly resistible intuition that the authors of narrative fiction are trying to say something about how things go in our world. For example, that Balzac is trying to show us in *Splendors*, among other things, the rules governing the life of the high Paris society in the restauration time and that Beckett is trying to show us in his *Trilogy* what space does the world in which we live leave for meaningful action and mutual understanding.

In general, if we accept the principle *F* as the starting point and correspondingly assent to the equation:

---

[12] I have defended this approach in polemics with representatives of the fictional worlds theory, in English e.g. in Kot'átko (2014). I find my views very close to the position recently presented in Friend (2016).

[13] "Certainly the philosopher of 'possible worlds' must take care that his technical apparatus does not push him to ask questions whose maningfulness is not supported by our original intuitions of possibility that gave apparatus its points" (Kripke 1980, 18).

> the world the narration is about = the world in which the narration takes place = the world in which we follow the narration = the actual world of our life,

then we, as the readers of narrative fiction, as well as the theorists of fiction can simply adopt a set of intuitively plausible (if not trivial) assumptions which would otherwise require special moves from the reader (like "recentering" and "transportation", cf. e.g. Ryan 2010) and special justification from the theorist. For example, we can take for granted that the sentences uttered by the narrator and by the people she speaks about are sentences of one of the natural languages spoken in our world, for instance sentences of French, rather than sentences of some fictional language spoken in the fictional world of, let us say, *Madame Bovary*. No transportation of the conventions of French (as they evolved in the actual world from Latin and other roots) to another world is needed for justifying our reliance on our competence in French when interpreting the text.

4. Fictional names

Within this framework we can automatically $\text{assume}_{\text{AI}}$ that the names we find in a text of narrative fiction function in the same way as the names we use in everyday conversation. For the theorist studying the semantics of narrative fiction this implies that if she accepts, for instance, Kripke's causal theory of names for ordinary discourse, she should automatically apply it also to fictional texts. Granted the principle *F*, there is no space left for such a theorist for hesitating between the causal theory and its rivals. And since this is also my case, I am committed to the following principle:

/R/ The occurrence of an expression which behaves like a proper name in the text of narrative fiction indicates that the reader should $\text{suppose}_{\text{AI}}$ that in this stage of narration the narrator utters a proper name to refer to that individual which has been assigned that name at the beginning of the chain to which this narrator's utterance belongs.

This formulation is, admittedly, not too elegant: for most purposes it should be enough to say that we take_{AI} the narrator as uttering a name to speak

about the person referred to by that name in his community – if we abstract from the fact that the same name (identified purely phonologically or orthographically) has usually more bearers. Needless to stress, the principle *F* implies that the names used in everyday communication to refer to countries, cities, mountains, statesmen etc. do not change their referential function if they appear in a fictional text: since the universe of discourse remains the same as in everyday conversation, namely the actual world.

On the level of theoretical debates, the way in which proper names function in fiction is indeed open for discussion (and I have profited a lot from exchanges concerning this issue with Manuel García-Carpintero and Zsófia Zvolenszky, cf. e.g. García-Carpintero 2015, Zvolenszky 2015). But the reader, I suppose, spontaneously approaches the names in the text of narrative fiction in the way she is accustomed to from everyday communication: that means that she assumes (although in the *as if* mode) that the persons spoken about were given their names quite independently of the narrator's performance, were then continuously referred to by means of those names and the narrator simply joins this practice. This $\text{assumption}_{\text{AI}}$ is applied automatically (which means that it doesn't require justification), unless it is blocked by some special narrator's move. For instance, the narrator can remark that "for understandable reasons" she changed the names of the protagonists. Alternatively, the narrator may in a metafictional, self-disclosing remark explicitly present her hero as a fictional construct with a fictitious name. So, in the First Chapter of the novel *Waverley* the narrator (on this occasion, I would say, coinciding with the empirical author, i.e. Walter Scott), overtly introduces the reader into his considerations which led him to choosing the name *Waverley* for his hero.[14] Obviously, in such remarks the occurrences of a name in the literary text are not presented as part of a chain of uses of that name, originating in an act of baptism independent of the narrator's performance. But we, as cooperative readers accepting rules of the game, are still invited to $\text{approach}_{\text{AI}}$ the name and its occurrences in this way, because it has been introduced into the game precisely with this function, despite the ostentatious and self-disclosing form of this introduction.

[14] I use the word "hero" as a term for the person spoken about by the narrator, rather than as a synonym for "character".

5. Identification of the persons referred to in fiction and identification of fictional characters

But let's focus on more straightforward cases like, e.g., the name "Emma", as it appears in the text of *Madame Bovary*. The interpretation$_{AI}$ of its occurrences as records of the narrator's utterances of that name, in which the narrator joins a chain of uses located in the actual world, provides us with a simple principle of the identification$_{AI}$ of the person we are thinking about under the name "Emma" when reading Flaubert's text. It is the person uniquely satisfying the description (*D*) "the person to whom the name 'Emma' has been assigned at the beginning of the chain to which these narrator's utterances belong". The world to which this description is to be applied is fixed in advance as the actual world – by our locating the narrative performance and the entities referred to by the narrator into this world. So we are not confronted with any problem of the kind discussed by Gregory Currie (in Currie 2003): namely that such a description identifies various individuals in various "story worlds" of *Madame Bovary* (that means in those possible worlds in which everything said in the novel by the primary narrator is true). And we are not forced to conclude, following Currie, that the expression "Emma Bovary", as it appears in Flaubert's text, does not work as a name of an individual but rather as a name of "Emma-role", which is a function from possible worlds to individuals.

The identificatory force of the description *D* is parasitic upon the reference to the narrator's utterances and to the general mechanism of functioning of names. This, I believe, makes this way of identification safe against the problems potentially connected with the so called "individuation just by names": here we have a case of individuation by names as uttered by particular speaker in particular circumstances. So, when equipped with our parasitic description *D*, we can be sure that we are thinking and speaking (in the *as if* mode) about "the right Emma", despite the fact that the same name plays a prominent role also in Jane Austen's famous novel (Stacie Friend mentions this problem, addressing it to antirealists concerning fictional characters; see Friend 2007, 145).[15] To be sure, the identity of the

[15] Similarly for the cases in which the narrator refers to a person by means of a description. The identification based on the occurrence of the description "the man in the

referents of the name "Emma", taken$_{AI}$ as used by Flaubert's narrator, and the name "Emma", taken$_{AI}$ as used by Austen's narrator, is not thereby excluded: this possibility is just left open, as it should be. (Before starting to read Flaubert's text we cannot exclude that its literary functions require that the utterances of "Emma" to be found there are interpreted$_{AI}$ as referring to the same person as the utterances of "Emma" in Austen's text.) Similarly, this approach leaves open the possibility that different names, taken$_{AI}$ as uttered by different narrators, should be interpreted$_{AI}$ as referring to the same person, if it is required by the literary functions of at least one of the texts in question – as it is e.g. with Homer's "Ὀδυσσεύς", Vergilius' "Ulysses" and Dante's "Ulisse" (cf. the discussion of this case in Friend 2014; I will return to the problem of co-identification later). Finally, this approach excludes actual persons bearing the name "Emma Bovary" as well as persons satisfying non-parasitic descriptions based on Flaubert's novel (like "the only daughter of a farmer from Les Bertaux, wife of a doctor from Yonville" etc.) from being acceptable candidates for the status of the person the novel is about – unless we have a reason to believe that the literary functions of Flaubert's text require us to relate the narrator's utterances to any such person.

But this does not mean that the parasitic description *D* is just an artificial construct designed to keep apart fiction and non-fiction (where it is needed) or to solve other potential problems of the theory of literary fiction: since this parasitic way of identification is frequently applied in everyday conversation as well. For instance, if I witness a conversation in which the participants use the name "John Smith" and I am not certain who of the hundreds of bearers of this name is spoken about, I can still identify the object of conversation quite precisely by means of parasitic description "the person referred to in this conversation as 'John Smith'". In many cases I would, for practical reasons, prefer having a device of identification less bound to particular utterances made by particular speakers in particular situations, like "the present dean of the Philosophical Faculty of Charles

corner" in a text of narrative fiction has the form of an extended metalinguistic description: "the person uniquely satisfying the description 'the man in the corner' as used by the narrator of this utterance". Here, the reference to the narrator's utterance eliminates problems with multiple occurrences of this description on various places of the text, or in other texts (cf. Friend 2007, 146, 149).

University". But this does not question the identificatory force of the parasitic description available to me, that is its ability to pick out precisely one individual (provided that the conversation I witness has a determinate subject). And similarly in case of reading Flaubert's novel. I assume$_{AI}$ that there is a person uniquely satisfying the parasitic description "the person referred to by these narrator's utterances of the name 'Emma'" (which can be unpacked in the Kripkean way mentioned above), and it is the person identified in this way to whom I attach various non-parasitic descriptions I collect when reading Flaubert's text. According to this picture, the burden of the identificatory task lies on the parasitic description bound to the narrator's utterances and to the general mechanism of referential functioning of names, rather than on the bunch of non-parasitic descriptions provided by the text. Hence we can disagree about Emma's properties without putting in doubt the presumption that we are speaking about the same: about properties of the person we assume$_{AI}$ as the referent of the narrator's utterances of the name "Emma". Unlike Stacie Friend I would not put it so that "we intersubjectively identify characters even if we disagree about them" (Friend 2007, 146); instead, I would say that we intersubjectively identify$_{AI}$ a person and connect two different characters with Flaubert's novel. Let me explain this point.

It should be clear that what I have been speaking about until now is not a way of identifying Emma as a fictional character:[16] I have described a way in which the reader can think about Emma as about a *real human being* of flesh and bones, rather than as about a fictional character. The informational content involved in the parasitic identification based on the description *D* is indeed extremely poor, but this should not disturb us: our question is whom we are thinking about, not how she is like. As attentive readers, we *learn* something about her from every new page of the book, but we *think* about her from the very first occurrence of her name in the text. On the other hand, to identify Emma as a *fictional character* is, within this

[16] It should not be confusing to use the same name "Emma Bovary" both for the literary character and for the young lady we are suposed to assume$_{AI}$ as existing in the actual world, we just should keep in mind that they are not bearing the name in the same way. We assume$_{AI}$ that there exists a real person referred to by this name in the same way in which we are referred to by our names; and we use the same name for the literary character constituted (among other things) by this assumption.

account, to identify certain parameter of the literary functions of Flaubert's text, precisely speaking, of the requirements imposed by these functions on the reader.[17] Basically, this amounts to providing a list of assumptions the reader is required to make_{AI} in order to let the text fulfill its literary functions for her. In Emma's case, the list will include:

(1) the assumption that there exists precisely one person referred to by the narrator's utterances of the name „Emma" (namely the person to whom that name has been assigned at the beginning of the chain to which these utterances belong);

(2) the assumption that that person (the person uniquely satisfying the description specified above) married a young doctor called Charles Bovary;

etc.

The fact that the literary functions of Flaubert's text require the reader to $\text{accept}_{\text{AI}}$ these assumptions constitutes Emma as a literary character with all its specific features, i.e. makes it what it is: such and such literary construct distinct from all other characters known to us from literature, drama, film or computer games, and of course distinct from theoretical constructs, legal constructs, etc. As far as I can see, there is no need to endow it with some metaphysically more substantial way of being in order to solidify its ontological status.

So, when I say that the parasitic description "the person referred to by the narrator's utterances of the name 'Emma'" provides the reader with a way of thinking about Emma, I do not mean that what is identified in this way is Emma as a fictional character. And, of course, I do not want to claim that there is some real person identified in this particular way. What I mean is that the availability of such a description makes the reader's thoughts about a person called "Emma" quite determinate, complete or saturated at least with respect to the identification of their object. Put in terms of the

[17] This obviously means to identify a component of the actual world: the fictional character called "Emma Bovary" exists in the actual world due to the fact that there is a text whose literary functions require us to $\text{assume}_{\text{AI}}$ the existence of a woman with that name, not as a literary character but as a being of flash and bones (cf. Section 1). So, I would say precisely the opposite to Anthony Everett's claim: "Fictional characters exist in stories, not in the real world" (Everett 2013, 132).

adverbial account of reference (cf. e.g. Kriegel 2007), one can say that what acquires in this way a determinate content is the reader's "thinking Emma-wise" or, equivalently, the reader's "Emma-ward-esque thoughts". This way of speaking does not commit us to there being any entity thought about or spoken about when one thinks or speaks about Emma.[18]

On the other side, even if the referential mechanism we are supposed to $\text{connect}_{\text{AI}}$ with the name "Emma" does not relate Flaubert's text, nor our thoughts accompanying our reading the text, to any real or fictional entity, it plays an essential role in our $\text{assumption}_{\text{AI}}$ that the text *is* related to a person called "Emma" existing in the actual world. It might be illuminating to compare this with the general aspirations of the non-relational theory of reference represented in the most inspiring way by Mark Sainsbury and Uriah Kriegel (cf. Sainsbury 2012, Kriegel 2007 and criticism in Koťátko 2014). Reference by means of fictional names (as they are used in texts of narrative fiction) is indeed non-relational, because it does not establish relation to any kind of entities referred to. But at the same time it confirms the primacy of the relational notion of reference, because the very determinedness of the reference by means of fictional names includes the assumption (although made in the *as if* mode) that it relates us to some real entities: that there are real persons, places, situations, events etc. the narration is about. Hence even the fictional reference, if understood in this way, undermines the universal pretensions of the non-relational account of reference. And, to be sure, our thoughts and claims about Emma *as a literary character* (like "Emma is the most famous Flaubert's character") are standard relational thoughts and claims about certain parameters of Flaubert's text.

This may seem to contradict to our common way of speaking. When we are asked to characterize some literary character, e.g. Emma Bovary, we usually say things like: it was a charming, sensitive, ambitious, self-centered, frivolous etc. young woman – saying thereby things which cannot be

[18] I take the following Stacie Friend's words as pointing in the same direction: "Yet there is a reason to think that an anti-realist account is required independently of issues to do with fiction and fictional characters, since there are a wide variety of domains in which we seem to be talking or thinking about the same thing even when there is no thing we are talking or thinking about. Once we have such an account, it is not clear why we need fictitious objects" (Friend 2007, 154).

straightforwardly ascribed to a text or its parameters but only to a real human being. But there is no confusion or category mistake behind this way of speaking: it is just one example of an indirect predication, in which the property specified in the predicate term is not straightforwardly ascribed to the referent of the subject-term, but to another entity related to it in some easily identifiable way.[19] In our case we characterize certain construct (literary character) by enumerating some of the moves constituting the construct, namely the assumptions$_{AI}$ required by the literary text, which together produce an image of a young woman bearing the name "Emma Bovary" with the properties mentioned, presented$_{AI}$ as an image of a real human being.[20] The enumeration of properties ascribable only to concrete individuals can serve as a characteristics of a literary character only within this (typically implicitly assumed) framework.

6. Fictional names: extra-textual use

We have approached fictional characters as certain parameters of the literary functions of the texts of narrative fiction and admitted that this justifies their classification as abstract entities. We have found no motivation for postulating any other, metaphysically more substantial (and philosophically more appealing) kind of abstract entities to play the role of fictional characters, but the reason might have been that we have neglected some contexts in which people speak about fictional characters – some important levels of fictional discourse.[21] But I think this is not the case, at least with respect to the kinds of use of fictional names frequently discussed under the title "paratextual" and "metatextual" and also with respect to negative

---

[19] Similarly: when saying about some symphony that it is noisy we mean that its standard performances are noisy, when saying that some sentence is clever we mean that the thought it expresses is a result of a clever way of thinking etc.

[20] Another way of putting this would be to say that Emma as a character *encodes* (rather than *exemplifies*) the properties of the kind mentioned, borrowing the well-known terminology from Zalta (1988). I owe a lot to Marián Zouhar's discussion about the merits and limits of Zalta's *encoding–exemplifying* distinction in Zouhar (2016).

[21] This part of the paper has been inspired by recent discussion with Zsofia Zvolenszky (cf. Zvolenszky 2015).

existential claims, in which fictional names appear. The interpretation of these cases I am suggesting does not add anything new to the approach I have been advocating above: hence I will be quite brief (with some exceptions).

(1) *Paratextual use*:
"Emma ruined her husband."

This can be quite naturally paraphrased by means of some fiction-operator, like:

(1') In Flaubert's novel, Emma ruined her husband.

But the real point at issue is how to interpret this paraphrase: the principle *F* suggests to unpack it in the following way:

(1'') The literary functions of Flaubert's novel require us to assume$_{AI}$ that there exists a person referred to by the narrator as "Emma" and that that person ruined her husband.

That certainly does not commit us to the existence of any abstract entity called "Emma", over and above our approaching the literary character called "Emma" as a parameter of the functions of the text. Obviously, it makes sense to say: Emma is one of the persons we have to assume$_{AI}$ as existing, in order to make sense of Flaubert's novel. But this does not include any hypostasis: it amounts to saying that the literary functions of Flaubert's text impose such and such demands on us. Stacie Friend has pointed out (in Friend 2007, 143) that if we take the name "Emma Bovary" (her example is "Candide") as an empty term, the sentences containing it will not be able to express a complete proposition whether or not they are prefixed by a fiction operator.[22] But within the approach suggested here, the term "Emma Bovary" does not come out as empty: when reading the text, i.e. in our confrontation with the name's textual use, we are supposed to interpret$_{AI}$ it as a device of the narrator's reference to a real person. And

[22] And correlatively: "If statements apparently about Hamlet are not really about anything, if nothing we say about Hamlet is true, what is the point of talking about him? The anti-realist owes us an explanation of the function of fictional discourse" (Friend 2007, 143).

if it is used within a talk about a piece of literary fiction, it refers to certain parameter of that fiction, in a way which is made explicit in our analysis.

(2) *Negative existential claims*:
"Emma does not exist."

Here I opt for the following metalinguistic paraphrase:

(2') The word "Emma", *as it appears in Flaubert's text*, does not have any referent in the actual world (i.e. it is not a proper name of any real person).

This paraphrase is, due to the clause in italics, perfectly compatible with the possibility that (a) there is (in the actual world) a person called "Emma", (b) there is a person satisfying all the Emma-descriptions we find in Flaubert's text and (c) there exists another text of narrative fiction in which the name "Emma" is (similarly like "Rouen" in Flaubert's text) used to refer to a real entity.

(3) *Metatextual claims*:
"Emma is a fictional entity."

The interpretation I am suggesting combines elements of the paraphrases of (1) and (2):

(3') The expression "Emma", *as it appears in Flaubert's text*, does not have any referent in the actual world, but the literary functions of the text require that we $\text{assume}_{\text{AI}}$ the opposite.

The clause in italics plays here the same role as in (2): it makes our claim compatible with some possibilities which no sound interpretation should exclude. As this paraphrase shows, I do not share the view that metatextual uses of fictional names require introducing fictional characters as abstract entities into our ontology – over and above what has been said about the existence of fictional characters as parameters of literary functions of texts of narrative fiction (cf. e.g. van Inwagen 1977; Kripke 2011; Zvolenszky 2015).

In this reading, the metatextual claim (3) implies the negative existential claim (2), since (3) is here interpreted as claiming that Emma is a *mere*

fictional character – which is, I suppose, the intuitively most natural reading. But let us imagine that somebody utters the sentence

(3a) Napoleon is a fictional character of *War and Peace*

without intending to imply that the name "Napoleon", as it occurs in *War and Peace*, should not be read as referring to the real emperor we have heard about in the school lessons of history. I think this use of (3a) makes a good sense and the appropriate paraphrase would then be:

(3a') The text of *War and Peace* requires the reader to assume$_{AI}$ that there exists a real person referred to by occurrences of the name "Napoleon" in that text.

The assumption behind this suggestion is that even if the name "Napoleon" is uttered in *War and Peace* with its standard referential function, it makes still good sense to speak about Napoleon also as about Tolstoy's fictional character. Then we should distinguish:

(a) the person referred to by utterances of Napoleon in the text of *War and Peace* – which is the real emperor referred to by historians as "Napoleon I";

(b) the fictional character named "Napoleon", identified by a set of assumptions$_{AI}$ required by the text from the readers, including:

  (i) There exists a person referred to by the narrator's utterances of the name "Napoleon".

  (ii) That person is identical with the emperor referred to by historians as "Napoleon I" (in other words: the narrator and the historians participate in the same chain of uses of the name "Napoleon").

  (iii) That person defeated Russians and Austrians at Austerlitz.

  (iv) That person saw and commented the half-dead body of Andrei Bolkonski on the Austerlitz battlefield.

  Etc.

It might be objected against our analysis of the cases (1), (2) and (3) that it turns claims about non-linguistic objects to claims about language.

But it depends on how you decide to specify the topic of these claims: I understand them as claims about the world including (as part of the specification of their content) metalinguistic elements. The negative existential claim says that the world is such that it does not include any person referred to by certain name, as it is used in certain text. And the metatextual claim adds to this that the world nevertheless includes a piece of fiction which requires the acceptance$_{AI}$ of an opposite existential assumption.

(4) *Claims about intertextual identity*:
(a) "Dante's Ulisse is identical with Homer's Odysseus."
(b) "Dante's Ulisse is different from Homer's Odysseus."

Before commenting on these claims let me say a few preliminary words related to Stacie Friend's presentation of this case (in Friend 2014). The question is: is Dante's Ulisse (referred to in the 26th Canto of *Inferno*) identical with Homer's Odysseus? As Stacie Friend rightly points out, this issue is purpose-relative and context-sensitive (cf. Friend 2014, 321). In other words, this question can be understood as introducing quite different problems to be solved: hence we should not be surprised that we can be given incompatible, and yet intuitively plausible and fully justified replies. From my point of view we have to distinguish two readings of the question. One concerns the identity or non-identity of the *person* we are required to assume$_{AI}$ as being referred to by occurrences of "Ulisse" in Dante's text with the *person* we are required to assume as being referred to by occurrences of "Ὀδυσσεύς" in Homer's text. The other reading concerns Dante's and Homer's *literary characters*. So, we have in fact two different questions, which I would articulate in the following way:

(1) Do the literary functions of Dante's *Inferno* require that we take$_{AI}$ the occurrences of the name "Ulisse" as referring to the same person as the occurrences of the name "Ὀδυσσεύς" in *Odysseia*?
(2) Is Dante's Ulisse the same character as Homer's Odysseus?

I suppose everybody will agree that the reply to the first question is "yes", while the reply to the second is "no". Dante's literary character named "Ulisse" is an element of the literary construction of *The Divine Comedy* and to identify this element is to specify the requirements it imposes on the

reader. In particular, Dante's text requires the reader to accept$_{AI}$ a series of assumptions, including:

(a) There exists (in the actual world) a person referred to by the occurrences of the name "Ulisse" in the text.

(b) That person is identical with the person referred to by the occurrences of the name "Ὀδυσσεύς" in Homer's *Odysseia*.

(c) That person suffers in the eighth circle of the hell in one flame with Diomedes.

Etc.

This list gives a clear reply to both questions raised above: it is certainly different from the list of requirements connected with the name "Ὀδυσσεύς" in Homer's *Odysseia*. There are some overlaps (for instance the readers of both texts are required to believe$_{AI}$ that the person referred to by the relevant name spent some time in the company of an enchantress called "Kirke"), but there are also incompatible requirements (in one case we are required to believe$_{AI}$ that the hero returned to Ithaca, met again his wife, son, old dog etc., in the other case we are required to believe$_{AI}$ that the hero never returned home and instead died on the sea). At the same time, the *Inferno*-list includes requirement that the reader identifies$_{AI}$ his hero with that of *Odysseia*: otherwise she would not make proper sense of the 26th Canto of *Inferno*.

Analogically, imagine a commercial in which a body-builder dressed in a (synthetic) piece of lion skin, armed with a monstrous mace and calling himself "Hercules" suddenly appears in a supermarket and loudly demands his favorite yoghurt. Again, it seems clear that we are supposed to assume that it is the same *person* as the one referred to in Greek myths by the name "Ἡρακλῆς", in other words, that the same man who killed the Nemean Lion, Learnean Hydra, Stymphalian Birds and countless other creatures is now asking for his yoghurt – otherwise the intended effect would not work. But at the same time, the *character* of the sketch differs from the *character* of the myth, precisely because of the yoghurt affair (i.e. for similar reasons as in the Ulysses' case).

Finally, let's take the *Pierre Menard* case from Borges' famous story. Here we will probably agree with Borges' narrator and numerous

commentators that the main *character* of Cervantes' novel *El ingenioso hidalgo Don Quixote de la Mancha* and the main *character* of Pierre Menard's novel bearing the same title are different, as is the whole construction of the novels, despite their texts being word by word identical. But unlike in Ulysses' case, I don't see any reason why we should, as readers of Menard's novel, assume$_{AI}$ that the name "Quixote", as uttered by Menard's narrator, refers to the same *person* as the name "Quixote", as uttered by Cervantes' narrator. What is clear is just that the construction of Menard's novel includes reference to Cervantes' novel: the identity of their texts and the contrast of the literary projects is indeed something the Menard's reader is supposed to be aware of and to appreciate.

(5) *Intentional transitive constructions*:
"I pity Emma."

I think that the proper interpretation of this case should start with the question: how would I explain the claim I intend to make in uttering this sentence, when asked e.g.: "What are you speaking about? Whom do you claim to pity?" The proper reply would of course depend on the kind of challenge behind this question. In our present context, I would probably say something like this: "Flaubert's novel will make sense to me as a piece of narrative fiction only if I imagine, or pretend to believe, that there exists a person referred to by the narrator as 'Emma', who did such and such things and to whom such and such things happened. I have imagined that and it made me feeling pity." It should be clear that what I pity here is a person of flesh and bone I am supposed to imagine as existing, not a literary character: the literary character named "Emma" is an ingenious literary construct which I can only admire. Stacie Friend has pointed out, when speaking about sentences like (5), that according to realists "the best explanation of these phenomena is that we are thinking and talking about fictional objects" (Friend 2007, 147). But a natural response to a realist claiming this would be that we certainly don't pity fictional objects. We, in conformity with the requirements imposed on us by the functions of literary texts, assume$_{AI}$ that there are certain persons who behaved in certain ways and certain things happened to them – and this induces certain emotional response.

7. The role of pretense

It should be clear that neither the paratextual statement nor the metatextual statement nor other kinds of statements we have been just discussing are, under this interpretation, made in the *as if* (or *pretense*) mode. They are full-blooded statements about Flaubert's novel which, if made sincerely, manifest beliefs (rather than make-beliefs) of the speaker. Stacie Friend has pointed out that "Whether or not we accept realism, we must allow pretense a significant role in explaining thought and discourse about fictional characters" (Friend 2007, 154). I agree that this role is significant, but would like to add that it is also quite narrow:[23] in my account, pretense is involved just as a requirement imposed on readers by the literary functions of the text.[24] When speaking about fictional characters in (for instance) metatextual claims we do not continue in this pretense and there is no reason for doing so: we make straightforward, serious claims about the literary functions of a text. And, as I have already noted (in Section 1), when speaking about the author's creative achievement, we do not have to ascribe to her any pretense either. It is quite sufficient to say that the author creates a text with certain parameters – text whose literary functions require (among other things) that the reader makes certain moves in the mode of pretense. When doing this the author does not make any assertions, nor pretends to be doing so, nor performs any other acts whose proper reception would require that the charitable audiences adopt a special stance involving the "disengagement from certain standard speech act commitments, block-

[23] For an account of fictional discourse based on a radically extended application of the notion of pretense see Everett (2013).

[24] According to Friend, the anti-realist account of various kinds of claims made within fictional discourse requires slips between various games of make-believe, and hence we need "a way to distinguish those games" (Friend 2007, 153) – which antirealists still owe us. But the approach I am proposing does not require any such slips, since, as I believe, there are no games of make-believe played here. Friend suggests that according to anti-realists, "in talking about fictional characters we engage in the pretense, established by authors of fiction, that there are such and such persons, places and things" (Friend 2007, 153). This is certainly not my position: in fictional discourse we seriously speak about fictional characters, without pretending that they are anything else than components of the construction of literary works.

ing inferences from a fictive utterance back to the speaker or writer, in particular inferences about beliefs" (Lamarque & Olsen 1994, 46). The point is that the author does not occupy any "official" position within the basic interpretative scheme or division of roles specific for reading narrative fiction. It is the narrator to whom the reader is supposed to $\text{assign}_{\text{AI}}$ the utterances of sentences she finds in the text and the performance of the speech acts indicated in these utterances, among them full-blooded assertions, imposing on the $\text{assumed}_{\text{AI}}$ speaker all the commitments connected with this speech act type. This includes that it makes good sense to approach the narrator as making unjustified statements, as lying, as making insincere promises and using all kinds of communicative tricks we know from everyday communication – otherwise the construction of an unreliable narrator would be impossible. So, there is no reason and no space for general "disengagement from standard speech act commitments" related to the author or to the narrator: like in everyday communication, this move is applicable only in special cases, e.g. when the text includes signals that the narrator is joking.

Within this approach, the starting point are the functions of the text and the characteristics of the author's creative act is derived from the specification of these functions. The reverted order is much more popular, owing to authors like John Searle, Saul Kripke or Gareth Evans (cf. Searle 1975; Kripke 2013, 17; Evans 1982, 353): here the (intentional) characteristics of the author's creative acts is an essential part of the explanation of the functions of literary texts, in particular of the role played by proper names within these texts. So, the author is said to *pretend* that she uses expressions like "Emma Bovary" to refer to real human beings (John Searle and many others) or, equivalently, to *pretend* that the conditions of reference connected with the names she utters are satisfied in the actual world (Saul Kripke). I take it as an attraction of the approach I have argued for that it does not require any speculations about the mode of the author's creative act. Instead, we specify the functions of literary texts of narrative fiction and add that if the author's creative act is purposeful, it includes the intention to produce a text with these functions. Pretense enters into this picture on another place: the author creates a text which will fulfill its functions for the readers only if they accept some assumptions in the mode of pretense.

When doing so, the author can be said to "make the first move in a game of make-believe", in that sense that she makes a move which (if successful)

will prompt acts of make-believing on the part of the readers. But this move need not have the form of "a deliberate initial pretense" (Evans 1982, 353). The author may simply intend to create a text which will function in such and such a way and approach this project in purely "constructivist" manner, which will not require her involvement in the form of make-believing that such and such things happened, or "pretending to have knowledge of things and episodes" (Evans 1982, 353).[25] Nobody will deny that a well-trained liar can produce in her audiences a belief that *p* without herself believing that *p*. Why not to admit that a writer can deliberately produce in her readers a make-belief that *p* without herself make-believing that *p*?

8. Confrontation with the fictive utterance theory of fiction

Let me compare this approach briefly with the one proposed by the fictive utterance theorists of fiction. So, according to David Davies:

> Whereas it is a condition for assertion that the speaker intends the audience to believe what she states, in fictive utterance the author intends that her audience make-believe what is narrated (Davies 2001, 265; more recent discussion and development of this position see in Davies 2012).

I think that both claims – about assertive utterances and about fictive utterances should be revised in an analogical way. First, I believe that it is hopeless to try to define assertion in terms of speaker's intentions, since it is always easy to find clear counterexamples – cases in which the speaker does not have some of the required intentions but there is no reason to deny that she makes a full-blooded assertion. For instance, the condition mentioned by David Davies is certainly not necessary: I can utter the sentence "Jane isn't at home" with the intention to create in the audience the belief

[25] Or let us imagine a writer who seriously takes herself as describing real events but does so with high literary aspirations, intending to create a great novel. Let us say that she succeeds (the result is appreciated as a piece of high literature) but her beliefs concerning the events described come out as false. Will anybody claim that in order to be justified in regarding her as an author of a piece of narative fiction, we will have to re-evaluate her original atttitudes, classifying them as make-beliefs?

that Jane is at home, since I hope that the audience will take me, in this particular case, as trying to deceive her (for instance because she assumes that I want to prevent her from meeting Jane). But even then I certainly assert (as I intended) that Jane isn't at home – unless the context is such that it blocks the projection of the conventional meaning of the sentence uttered into the meaning of my utterance. Hence, rather than defining assertion (and other speech act types), in terms of conditions which must be fulfilled on the part of the speaker, it should be defined in terms of its communicative functions. In our case, these functions certainly include that my utterance *commits me* to believing that Jane isn't at home and to intending to create or activate the same belief in the audience.[26] Analogically, the specific status of the occurrences of sentences within literary texts of narrative fiction should be explained by specifying their function within this context, rather than by specifying the author's intentions. I have suggested (cf. the principle *F* in Section 2) that their function require the reader to interpret$_{AI}$ them as records of the narrator's utterances related to the actual world.

The ambitions of the fictive utterance theory of fiction are not limited to literary texts of narrative fiction: but when we apply it to a reader of such a text who interprets an occurrence of a sentence *S* expressing (as uttered in given context) a proposition *p* and indicating the assertive force, we get the following confrontation:

D.D.: The interpreter believes that when uttering the sentence *S* the author intends to produce in her a make-belief that *p*. On this basis (i.e. on the basis of the recognition of this intention) and as part of her interpretative cooperativeness, the interpreter make-believes that *p*.

P.K.: The interpreter make-believes (= believes$_{AI}$) that there exists a real person (the narrator) who asserts that *p* (by uttering *S*). On this basis and as part of her interpretative cooperativeness, the interpreter make-believes that *p* (provided that she approaches the narrator as reliable).

---

[26] I have defended the view that speech act types (and correlatively utterance meanings) should be defined in terms of commitments to certain attitudes imposed on the speaker rather than in terms of actual speaker's attitudes e.g. in Koťátko (1998).

9. Determination of literary functions

This should not be read as implying that the author's intentions have no relevance for literary interpretation. I have pointed out (in Section 7) that the empirical author does not occupy any official position in the interpretative scheme set up by the literary functions of texts of narrative fiction. Its main coordinates are fixed by assumptions$_{AI}$ concerning the narrator, her utterances and the actual world as the supposed universe of discourse and as the sphere in which the narrative performance is supposed to take place. This does not exclude that our knowledge about the empirical author and our hypotheses concerning her intentions may play an important role within the broad basis of interpretation, from which we approach the text and identify its literary functions (and thereby the literary work represented by the text). It can and need not be so, depending on our understanding the very enterprise of reading a literary text. If we take it as an opportunity for communication with the empirical author, the respect to her (probable) intentions will be the main constraint. If we approach the literary text as an artifact anchored in certain socio-cultural environment and as a medium for our intellectual and emotional intercourse with that environment, the literary conventions, the image of the world, the hierarchy of values etc. prevailing (according to our knowledge) in that environment will be the main constraint. And obviously, both approaches can be efficiently combined. But if we approach the literary text primarily or exclusively as a source of aesthetic pleasure, or as an inspiring challenge for our intellect, imagination, sensitivity, moral intuitions etc., we may feel free to opt for any reading which, according to our view, maximizes this potential of the text, without any respect to its author's intentions or its socio-cultural coordinates. A radical version of this kind of dealing with a literary text is mentioned in the end of Borges' story on Pierre Menard, as the so called "new art of reading": an example is to read *Odysseia as if* it were written in Rome in Augustus' time (at the turn of the era), hoping that this will generate an exceptionally powerful aesthetic experience. As usual, each approach can be enthusiastically defended or rejected, either as guilty in intentional fallacy, or in conventional fallacy, or simply as cynical – but the decision is in the last instance always on the reader. As far as I can see, there are no *apriori* principles or transcendental conditions of reading a literary text which would disqualify some of the interpretative attitudes

mentioned, in that sense that the reader who opts for it does not approach the text as a piece of literature.

References

BECKETT, S. (1979): *The Beckett Trilogy* (*Molloy*, *Malone Dies*, *The Unnamable*). London: Pan Books.

BORGES, J. L. (1998): Pierre Menard, Author of the Quixote. In: Borges, J. L.: *Collected Fictions*. (Trans. by A. Hurley). London: Penguin Books, 88-95.

CURRIE, G. (2003): Characters and Contingency. *Dialectica* 57, No. 2, 137-148.

DAVIES, D. (2001): Fiction. In: Gaut, B. & Lopes, D. (eds.): *The Routledge Companion to Aesthetics*. London: Routledge, 263-274.

DAVIES, D. (2012): Fictionality, Fictive Utterance and the Assertive Author. In: Currie, G., Koťátko, P. & Pokorný, M. (eds.): *Mimesis: Metaphysics, Cognition, Pragmatics*. London: College Publications, 61-85.

DOLEŽEL, L. (2000): *Heterocosmica. Fiction and Possible Worlds*. Johns Hopkins University Press.

EVANS, G. (1982): *The Varieties of Reference*. Oxford: Clarendon Press.

EVERETT, A. (2013): *The Nonexistent*. Oxford: Oxford University Press.

FRIEND, S. (2007): Fictional Characters. *Philosophy Compass* 2, No. 2, 141-156.

FRIEND, S. (2008): Imagining Fact and Fiction. In: Stock, K. & Thomson-Jones, K. (eds.): *New Waves in Aesthetics*. New York: Palgrave Macmillan, 150-169.

FRIEND, S. (2014): Notions of Nothing. In: García-Carpintero, M. & Martí, G. (eds.): *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press.

FRIEND, S. (2016): The Real Foundation of Fictional Worlds. *Australasian Journal of Philosophy* 95, No. 1, 29-42.

GARCÍA-CARPINTERO, M. (2015): Is Fictional Reference Rigid? *Organon F* 22, Supplementary Issue, 145-168.

CHATMAN, S. (1978): *Story and Discourse: Narrative Structure in Fiction and Film*. Ithaca: Cornell University Press.

KOŤÁTKO, P. (1998): Two Notions of Utterance Meaning. *Proceedings of the Aristotelian Society* 98, No. 1, 225-240.

KOŤÁTKO, P. (2014): Representations and Relations. *Organon F* 21, No. 3, 282-302.

KOŤÁTKO, P. (2016): Narrator in Decay. In: Koblížek, T. & Koťátko, P. (eds.): *Chaos and Form. The Echoes of Beckett in Literature, Theatre and the Arts*. Prague: Litteraria Pragensia.

KRIEGEL, U. (2007): Intentional Inexistence and Phenomenal Intentionality. *Philosophical Perspectives* 21, 307-340.

KRIPKE, S. (1972): *Naming and Necessity*. Oxford: Blackwell.

KRIPKE, S. (2011): Vacuous Names and Fictional Entities. In: *Philosophical Troubles. Collected Papers. Vol 1*. Oxford: Oxford University Press, 52-74.

KRIPKE, S. (2013): *Reference and Existence: The John Locke Lectures for 1973*. New York: Oxford University Press.

LAMARQUE, P. & OLSEN, S. H. (1994): *Truth, Fiction, and Literature*. Oxford: Clarendon Press.

MARTÍNEZ-BONATI, F. (1981): *Fictive Discourse and the Structures of Literature: A Phenomenological Approach*. Ithaca: Cornell University Press.

ROSSHOLM, G. (2015): Contribution to Fictional Epistemology. *Organon F* 22, Supplementary Issue, 133-144.

RYAN, M.-L. (2010): Fiction, Cognition, and Non-Verbal Media. In: Grishakova, M. & Ryan, M.-L. (eds.): *Intermediality and Storytelling*. Berlin & New York: deGruyter.

SAINSBURY, M. (2012): Representing Unicorns: How to Think about Intensionality. In: Currie, G., Koťátko, P. & Pokorný, M. (eds.): *Mimesis: Metaphysics, Cognition, Pragmatics*. London: College Publications, 106-131.

SEARLE, J. (1975): The Logical Status of Fictional Discourse. *New Literary History* 6, No. 2, 319-332.

SHENKER, I. (1956): A Portrait of Samuel Beckett, the Author of Puzzling Waiting for Godot. *New York Times*, May 6, Sect. 2.

THOMASSON, A. (1999): *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

THOMASSON, A. (2003): Fictional Characters and Literary Practices. *British Journal of Aesthetics* 43, 138-157.

VAN INWAGEN, P. (1977): Creatures of Fiction. *American Philosophical Quarterly* 14, No. 4, 299-308.

VOLTOLINI, A. (2013): A Syncretistic Ontology of Fictional Beings. In: Koblížek, T., Koťátko, P. & Pokorný, M. (eds.): *Text and Work: the Menard Case*. Praha: Litteraria Pragensia, 89-108.

ZALTA, E. N. (1988): *Intensional Logic and the Metaphysics of Intentionality*. Cambridge (Mass.) & London: MIT Press.

ZOUHAR, M. (2017): Fiktívne objekty a ich vlastnosti. [Fictional Objects and Their Properties.] *Organon F* 24, Supplementary Issue, 132-152.

ZVOLENSZKY, Z. (2015): Inadvertent Creation and Fictional Characters. *Organon F* 23, Supplementary Issue, 169-184.

Organon F 24 (3) 2017: 339-365

The Truth about Sherlock Holmes

FREDRIK HARALDSEN[1]

ABSTRACT: According to *possibilism*, or non-actualism, fictional characters are possible individuals. Possibilist accounts of fiction do not only assign the intuitively correct truth-conditions to sentences in a fiction, but has the potential to provide powerful explanatory models for a wide range of phenomena associated with fiction (though these two aspects of possibilism are, I argue, crucially distinct). Apart from the classic defense by David Lewis the idea of modeling fiction in terms of possible worlds have been widely criticized. In this article, I provide a defense of a possibilist account against some lines of criticism. To do so, I assume that names for fictional characters are directly referential and a possible-worlds model that accommodates transworld identity. On this background, I argue, it is possible to construct an elegant model of fictional discourse using familiar models of information exchange in ordinary discourse, and I sketch how this model can be used to i) make a natural distinction between fictional and counterfactual discourse, ii) account for creativity, and iii) sustain a natural definition of truth-in-fiction that avoids certain familiar objections to possibilism. Though I set aside questions about the metaphysical commitments of a possible-world interpretation here, there is accordingly reason to think that the battle over possibilist treatments of fiction will have to be fought over metaphysical foundations rather than technical shortcomings.

KEYWORDS: Creativity – fiction – non-actualism – possibilism – truth-in-fiction.

[1] Received: 20 April 2017 / Accepted: 20 July 2017

✉ Fredrik Haraldsen
SASPRO/Marie Curie Fellow
Institute of Philosophy
Slovak Academy of Sciences
Klemensova 19, Bratislava 831 64, Slovakia
e-mail: fredrikhar@yahoo.no

1. Introduction

Realists about fictional characters are committed to an ontology of fictional characters. A benefit of realism is the ability to make relatively straightforward sense of talk about such characters; though 'Trump is human' and 'Sherlock Holmes is human' differ in truth-value, the sentences seem *prima facie* to have the same kind of semantic content, and if 'Trump' refers to Trump and 'Trump is tall' expresses a singular proposition whose truth-conditions depend on how things are with Trump, one may expect the same to apply to 'Holmes'. *Prima facie*, then, 'Holmes' refers to Holmes, and sentences involving Holmes express singular propositions that are assigned truth conditions the same way sentences containing 'Trump' are.

Realists disagree about the nature of these characters, however. According to *possibilism*, or non-actualism, fictional characters are *possible objects*. Although Holmes does not actually exist, he could have, and does exist in some possible world different from the actual one. Moreover, we understand sentences containing 'Holmes' at least in part by grasping what things are like in worlds relative to which those sentences are true. In the present article I provide a partial defense of possibilism. However, to do so we should distinguish two distinct goals a possibilist analysis could aim to achieve. First, a possibilist analysis might aim to provide a semantic theory for sentences in fiction, one that assigns intuitively correct truth-conditions to such sentences. Now, I will assume that sentences containing 'Holmes' express singular propositions and that 'Holmes' directly refers to the merely possible object Holmes, who exists in multiple worlds where he instantiates different properties. A challenge to this view that I briefly discuss below is to explain how we can refer to the right individual (see Kripke 2011; Thomasson 1999). To provide a semantic theory that identifies truth-conditions for each sentence in a fiction, however, we need not worry for instance about what counts as true in a fiction or whether there is a set of relevant worlds that can serve as truth-makers for the *whole* fiction.

A second goal is to develop a model of fiction as a type of *discourse*, that can help explain how we interpret, and share interpretations of, fictional narratives. Doing so requires for instance accounts of what counts as being the case in a story, what distinguishes storytelling from counterfactual discourse and how possible worlds can be used to reflect the *creative*

aspect of storytelling. The primary aim of this article is to sketch a framework that can support the second goal. I do not intended to offer a comprehensive theory of literary interpretation, of course; questions about the role of the narrator or rhetorical devices such as irony or metaphor, for instance, are beyond the scope of this article and probably beyond what we can expect to model with possible worlds alone. Rather, my goal is to identify some basic tools that can later be supplemented in various ways.

To provide such a framework we need an account of what it takes for a claim to count as *true* in a fiction. A first stab may be:

(1) *In fiction f, 's' is true* iff *in the (relevant) possible worlds reflecting f, s*.

Two points are worth making here. First, we need to specify what it takes for worlds to be *relevant*. For the moment, think of the relevant worlds as the *closest* ones in which the events described in *f* take place. I return to this issue below. Second, (1) assumes that utterances of sentences such as 'Holmes is a detective' are (in the relevant contexts) elliptical for 'in the fiction, Holmes is a detective'. We could, however, interpret such sentences as non-elliptical and literally false but rather understand fiction talk for instance as engendering a shift in the *context of assessment* (following e.g. Kölbel 2002; MacFarlane 2005). Either option will in principle work for my purposes.

Possibilism does not only predict intuitively correct truth-conditions for sentences containing names for fictional characters, but in many cases also their intuitively correct *truth-values*, and does so in many ways better than other realist theories – Meinongian views, Platonism or abstract artifact theories – or irrealist views. Possibilism predicts that 'Holmes does not exist' and 'there is no Holmes' are true, insofar as existence ascriptions involve implicit restriction to the actual world. 'Holmes is a detective' has both a true and a false reading, depending on the presence or scope of an 'in-the-fiction' operator. Moreover, possibilism seems able to account for issues such as embedded fictions, intertextual comparisons ('Holmes is smarter than Poirot'), meta-fictional claims ('Holmes is a fictional character') and even make sense of speculations about characters' motivations or psychology, also when these are not explicitly detailed, insofar as there may well be a fact of the matter in the closest worlds where a character acts

as described in the fiction. Possibilism even gives us a basis for explaining how it is possible to *learn* from fiction – not only facts about the actual world related in the fiction, but about morality or the human condition. Of course, I will not have space to discuss all these features in detail here. But even the apparent ease with which possibilism can accommodate the intuitive data is some evidence in its favor.

There are, however, important questions I will set aside here; I mention some of them in Section 2. In Section 3 I sketch a possibilist model of fiction and explain how it accommodates i) creativity and ii) the distinction between fictional and counterfactual talk. In Section 4 I sketch and defend a promising theory of *truth-in-fiction*.

2. Referring to Holmes

There are several issues I do not have room to discuss. I merely note some of them here:

1) To provide a theory of truth for fiction it is natural to assume that we need possible worlds to be fairly metaphysically robust. To what extent we are committed to *modal realism* is a question I leave open.

2) I will assume that names are directly referential. A commitment to direct reference sits uncomfortably with a counterpart theory of modality, which appears to require some form of descriptivism (see Sullivan 2005, for discussion). A counterfactual claim about Aristotle must, on a direct reference view, be a claim about *Aristotle*, and not his potentially descriptively indistinguishable counterparts (on pain of a change in *semantic content*). To determine whether 'Aristotle could have failed to teach Alexander' is true, then, we don't search the space of worlds to discover one in which an identifiable Aristotle fails to teach Alexander. Rather, we consider what the case is with *Aristotle*, and evaluate the counterfactual claim by determining whether *he* is part of a world where he doesn't teach Alexander. On direct reference views, reference precedes the identification of relevant worlds. Accordingly, Aristotle must exist in multiple worlds, and the relevant worlds must contain the numerically same individual Aristotle. And since we want the semantic content of names in fiction to be of the same type as ordinary names, 'Holmes' should be directly referential as well. As a consequence, we reject counterpart theories of modality in our

model of fiction. Possibilism is not wedded to a counterpart theory, however. If we need robust modal realism, then transworld identity of individuals may be ensured by overlapping worlds or by adopting some form of modal dimensionalism.

3) For possibilism model fiction, there must be possible worlds containing magic, ghosts, dragons and people waking up as beetles or cockroaches. Widespread *essentialism* or proliferation of metaphysical necessities would be a problem for such accounts of fiction by limiting the kinds of stories authors could coherently tell even among stories that appear *prima facie* to be coherent. I will assume that if a story can be apparently coherently told in which an object of type *A* is described as being *not-B*, this is at least evidence that there are worlds where *A*s are *not-B*s.

The extent to which even widespread essentialism is an obstacle may be a function of what commitments our account makes to the metaphysics of modality. A constructivist or ersatzist approach *might* allow us to sidestep essentialist worries even consistently with accepting a wide range of metaphysical necessities: to provide the truth-conditions for sentences in fiction, we need the space of worlds to contain the logically (and analytically) possible worlds, but just as fictions are free to change laws of nature, fictions are free to suspend *metaphysical* necessities that are not also logical or analytical necessities and still be able to model the semantic contents of sentences correctly. If *x* is essentially human, *x* is human in all metaphysically possible worlds, but '*x* is not human' is not *logically* or analytically impossible; we can construct or describe logically and analytically consistent worlds in which *x* is not human, and stipulate that these are as relevant for our purposes as the metaphysically possible ones.

4) We may, as mentioned above, legitimately worry about how we can *refer* to merely possible objects. To defend possibilism as a model of fictional discourse, I need to say something about reference, and in the rest of this section I sketch a response to this worry (a complete answer will have to be postponed).

Whenever I utter a sentence containing a name I succeed in referring presumably in virtue of standing in the right sort of relation – perhaps a causal-historical relation – to the introduction of the name, and this relation is plausibly the same for my use of 'Holmes' as for my use of 'Caligula'. But whereas we can tell an obvious story about how 'Caligula' got attached to Caligula in the first place, it is less clear how Doyle attached 'Holmes'

to the merely possible object Holmes when introducing the name. Doyle cannot have established the reference of 'Holmes' by a baptismal act where the name was bestowed upon a demonstratively available object, since merely possible objects are not demonstratively accessible. Now, if demonstration were necessary for fixing reference, then a number of names for objects also outside of fiction would fail to refer, including names for abstract objects, individuals hypothesized to exist, or unobserved astronomical objects whose existence is inferred by calculation. In these cases, however, we can individuate referents as the individuals that uniquely satisfy certain descriptions. But descriptive individuation also seems unavailable in the Holmes case. Many merely possible objects – perhaps even some (or all) actual ones – satisfy the descriptions associated with 'Holmes' in the stories in different worlds. But insofar as 'Holmes' is directly referential – and thus *rigid*, picking out the same object in every world in which it exists – it must pick out *one* of them on pain of contradiction.[2]

Note that for descriptivists about the semantic contribution of proper names this kind of worry would not arise; for them, the contribution of 'Holmes' is a descriptively individuated *role* that can be satisfied by numerically different objects in different worlds without affecting the semantic content of sentences containing the name. But since we reject descriptivism, we need to know i) which object, among all suitable candidates, is Holmes; and ii) how Doyle managed to refer to it. I think the correct answer to the first question is that *it doesn't matter*. To be sure, we need 'Holmes' to refer to a single one among the merely possible referents, but it makes no difference *which one*. As for the second question, I suggest that when Doyle started writing about Holmes (or dreamt up the character), his implicit decision to describe a fictional character sufficed to ensure that some merely possible object was assigned, arbitrarily selected from among those that satisfy whatever properties – perhaps none apart from the name – Doyle associated with it at the time. On the surface, the stipulative introduction of 'Holmes' resembles a baptism, but instead of having a referent

---

[2] Suppose Holmes = α in world w_1 and Holmes = β in w_2 and $\alpha \neq \beta$. Since 'Holmes', by assumption, is rigid, 'Holmes' refers to Holmes in both w_1 and w_2, and the transitivity of identity gives us a contradiction.

demonstratively available, Doyle assigned 'Holmes' to *some (possible) individual or other* – the selection domain being restricted perhaps (only) by the condition that it be non-actual – that he could later develop to serve whatever purpose he wanted in the stories.

It is sometimes thought to be required for reference that there is some property that *o* must have for a name *n* to refer to *o* that can, in a non-circular manner, be invoked to distinguish *o* from other potential referents – either through a universally applicable condition such as *acquaintance* or a *causal relation* between name tokening and referent, or that *some* such condition must be satisfied in any particular instance, even if different relations – satisfying descriptive conditions or causal links – may do the trick in different cases. Of course, according to my suggestion there *is* a property Holmes has – *being arbitrarily selected as the referent of 'Holmes'* – in virtue of which the relation holds, but this property cannot be specified without invoking 'reference' (or a notion depending on *reference*, such as 'pick out' or 'select'). So, there is no non-circularly specifiable property in virtue of which 'Holmes' refers to a particular individual. Does it matter? There are at least two (compatible) lines of response available to the worry that it does.

First, one may deny that such a (non-circular) property is necessary for reference. A no-criterion theory of reference has been defended for names in general (cf. Breckenridge & Magidor 2012), and albeit controversial, I think that approach is more obviously acceptable in the special case of names for fictional characters. After all, given that the properties Doyle associated with Holmes at the introduction weren't contradictory, he was guaranteed that there is an appropriate individual, that the selected referent wasn't the *wrong* one, and that he would not go on to describe it incorrectly. There was no *unique* candidate for 'Holmes', but it is not entirely obvious that *unique satisfier* (across all worlds) is a better criterion for individuating referents than *arbitrary selected individual among numerous satisfiers* in the context of fiction. Second, what ultimately matters is not whether the relationship between 'Holmes' and the arbitrarily selected object is *a reference relation* or something else – perhaps *schmeference*, a relation between name and referent characteristic for individuating characters in fiction. What matters is that we can associate singular propositions with sentences containing 'Holmes' that have the truth-conditions they are supposed to have and play the role in our

theories of meaning and intentionality[3] they are supposed to play. This is something singular propositions containing arbitrarily selected, possible individuals seem able to do even if the relation between name and object is schmeference rather than reference.

The idea, then, is that when introducing the Holmes, Doyle selected a merely possible object to serve as referent for 'Holmes' among numerous potential candidates. Given transworld identity, the arbitrarily selected referent exists in multiple worlds where he instantiates different properties. Presumably, he is a barrel maker in some of these worlds, a detective in others, and a future space-traveler in yet others. When making the initial selection, Doyle didn't select any particular world; if he had not yet even decided that Holmes was going to be a detective when he started writing about him, Doyle had not yet ruled out the barrel-maker worlds. As long as the selected individual satisfies the properties Doyle ultimately came to associate with it in the stories in *some* world in which it exists, Doyle selected the right one.

To *guarantee* that Doyle selected the right one, then, we need to ensure that the selected referent *could* play the role of Holmes. I suspect that *any* merely possible object can do so, in which case we get the guarantee for free. This claim may be challenged. In particular, if many properties are essential properties of any individual that instantiates them, the guarantee would be harder to obtain. If *being human* is an essential property of any human, then Doyle must have decided to select a referent among possible individuals that are human. Having decided on species membership is perhaps not an unreasonable condition for making a correct selection, but any further properties taken to be essential makes the guarantee consecutively harder to obtain; if, say, *height* were an essential property, my account would be in trouble if Doyle initially selected someone shorter or taller than he eventually described Holmes as being. As mentioned above possibilism

[3] Questions of reference have, of course, both a semantic side and a cognitive side. I am here concerned with the former and have little to say about the latter. I assume, however, that *singular thoughts* are singular in virtue of their *form*, and hence that one can entertain a singular thought even if there is no referent the thought is, in fact, about. If this is correct, then there should be no in-principle obstacle to having a singular thought about an object to which the relation is schmeference rather than reference.

sits awkwardly with a proliferation of metaphysical necessities, and the account becomes less elegant if many properties are essential properties. We can, however, account – at some cost – for moderate degrees of essentialism; I return to this below. It is worth mentioning that full-blown *haecceitism* would *not* be a challenge to my account, however. If fictional characters possess individual essences, like *being Holmes*, then the problem of securing reference would not arise – in that case Holmes just is the unique individual in modal space that satisfies the description 'the individual that *just is* Holmes'.

3. The Model

According to a common model of information exchange (especially Stalnaker 1978) the content of a discourse contribution removes worlds incompatible with the content of that contribution from a *context set* of worlds reflecting the information presupposed or shared by the discourse participants (that is, the content intersects with the context set). This model can be adapted to fiction. (I leave it open whether we understand it as a model of the unfolding narrative or of Doyle's construction of the story – these may look rather different.) Now, Doyle, having decided to write about Holmes, can be taken to have individuated a context set containing all worlds in which the individual selected to be Holmes exists. Assuming that Doyle had not yet decided that Holmes was a detective or determined his height, the set contains worlds in which Holmes is not a detective or has different heights. Presumably, by Doyle's intention to engage in fiction and introducing Holmes as a fictional character, the *actual* world is not a member of the context set.

Having determined a context set, Doyle proceeded, over the course of the stories, as in ordinary discourse, to eliminate worlds from that set, starting with worlds where Holmes is not a detective, didn't live in London, and so on.[4] When introducing new characters, new reference selections are needed. Eligible candidates are those that populate scenarios in which Holmes could have taken part, and must thus inhabit worlds in the initial

[4] My presentation, to make the exposition easier to follow, does not reflect the actual Holmes storyline.

context set. So, for any eligible Moriarty candidate there must, at the outset, be a world in which Holmes interacts with that candidate as he satisfies the descriptions associated with Moriarty. The candidates are potentially any possible individual (at least among those not yet selected for other roles), and there is accordingly, in the initial context set, for any such individual *x* a world in which Holmes interacts with *x* playing the role of Moriarty. As for Holmes, when Doyle started writing Holmes may have been maximally incomplete, and every (contingent) property such that Holmes instantiates that property in some world. The initial selection, however, ensures that it is the numerically same referent that has these different properties in different worlds. Moreover, Holmes was never completely described; Doyle never specified Holmes's blood type, for instance. Incompleteness means that we are left with several worlds in which the same individual has different blood types, not different worlds with different referents.

As in ordinary discourse, certain updates of the context set are more difficult to accommodate. For instance, assume for the sake of argument that when starting to write *Fight Club*, Chuck Palahniuk had not decided that Tyler Durden and the narrator (the referent of 'I') were the same. At the outset, then, he selected different referents for the expressions, and, given the necessity of identity, there is no world in the context set in which they are the same. When subsequently deciding that Durden and I are, indeed, identical, Palahniuk needed to make a new selection. He would accordingly have to reinterpret everything in the narrative thus far relative to a new context set defined by the new referent. This consequence does not strike me as a failure of the model but rather as reflecting what actually happens when authors make momentous plot decisions (deciding that names for lesser characters in fact co-refer requires less backtracking). However, as the case suggests, widespread essentialism would make the model somewhat less elegant by making backtracking and reselection far more frequent and often required in cases where the author would be unaware that it is needed. As mentioned above, the plausibility of a possibilist analysis of fiction does to some extent hinge on apparently coherent scenarios described in fiction being genuinely *possible*.

In the next sub-sections I describe, first, how the model facilitates *creativity* and metafictional talk, and second how it captures the distinction between fictional and counterfactual talk. In Section 4 I provide an updated definition of *truth-in-fiction*.

3.1. Creating and describing characters

A common complaint raised against possibilist analyses is that they fail to capture the creative aspect of fiction making, turning it rather into a matter of *discovery*. Of course, when discovery is *discovery of possibilities*, the difference between discovery and creation is less than obvious. Consider the claim attributed – probably apocryphally – to Michelangelo that the sculpture was already complete within the marble block, and that he just had to chisel away the superfluous material. The description seems correct, if idiosyncratic, insofar as the resulting statue was, indeed, an already possible result of the process. Analogously, developing a story can be modeled as shrinking the context set, and developing a character as removing worlds by deciding which properties the character should have among those it could have had, to individuate a scenario that was, after all, already a possible outcome of the creative process.

Of course, Michelangelo's description neglects the crucial point that he *chose* which of many possible statues he was going to carve out, which is presumably essential to the process being *a creative process*. Similarly, freedom of choice is a distinctive feature of fictional storytelling, and Doyle was crucially free to narrow down the context set any way he wanted: As opposed to ordinary factual discourse, where the target is the actual world, Doyle freely chose his target world(s) among those available at any given point, and freely chose which steps to take to individuate it/them. Of course, possibilists will have to grant that in some sense Holmes was already *there*, and thus contradict the creationist intuition that Doyle really *brought Holmes into being*. But our intuitions that Holmes really is *brought into being* are shaky at best, and it is reasonable to wonder whether intuitions here are driven by prior metaphysical commitments rather than vice versa. At least I think the onus is on the critic to explain precisely what's missing from the possibilist account – keeping in mind it seems hard to deny that the Holmes stories are stories that could possibly have been told also prior to Doyle telling them.

Now, Doyle literally brought things like the narrative, name and features of the *representation* of Holmes into (actual) being, and thereby affected the causal structure of the actual world. The distinction between representations and scenarios represented must be managed with some care, however. The *stories* – at least the words and sentences – exist in the actual

world. Claims like 'in *Wozzeck* the eponymous protagonist is described as driven to murder by the oppressive mechanisms in his society' or (arguably) 'the plot of *L'Étoile* is very complicated' are claims at least partially about the actual world and aspects of our representations rather than the possible objects represented.

Keeping this distinction in mind, our account does not only account for in-fiction claims, but also common meta-fictional or critical claims. Consider the following (adapted from Brock 2002, 4-5):

(2) Holmes does not exist.
(3) Holmes is a fictional character.
(4) Anna Karenina is less neurotic than is Katerina Ivanovna (Howell 1979).
(5) Holmes would not have needed tapes to get the goods on Nixon (Lewis 1978).
(6) The character Odysseus who occurs in the *Odyssey* is identical to the character who occurs in *Inferno*, Canto 26, under the name 'Ulysses' (Howell 1979).
(7) There are characters in some nineteenth-century novels that are presented with a greater wealth of physical detail than is any character in any eighteenth-century novels (van Inwagen 1977).

(2) is straightforwardly true, as is (3) – *being a fictional character* entails *being a merely possible (non-actual) individual*, so (3) truthfully says (at least) that Holmes is a merely possible individual. (4), too, is true assuming that we measure their neuroticism based on the characters' appearances in the relevant stories. For (5), start with the worlds that serve as truth-makers for the stories. In the closest possible worlds – or at least some close worlds – relative to those (which are not necessarily worlds that define the *context set* for the stories, but counterfactual scenarios *accessible* from worlds in that set) in which Holmes got the goods on Nixon, he didn't use tapes. (6) and (7) require some care, since they talk both about representations and scenarios represented. (6) is true iff 'Odysseus' and 'Ulysses' co-refer, which is certainly likely on our account (though it would, of course, depend on factors related to the introductions of the characters); the two stories mentioned may, of course, not describe the same worlds of fiction, but that is – given transworld identity – not necessary for the *referent*

to be the same. For (7), it is perhaps best to interpret 'characters' as denoting possible objects and the sentence as making a claim about how *those* are being presented in the novels.

Other cases may be trickier. 'Holmes is famous' arguably seems to predicate a property Holmes cannot have in a world in which he doesn't exist, yet is ostensibly a claim about the actual world. However, the sentence probably requires paraphrase on any view; on abstract artifact theories, for instance, Holmes is an existing abstract object, and abstract objects are not detectives and thus intuitively the wrong targets for ordinary ascriptions of *fame*. I think a case can be made even at an intuitive level that the ascription requires paraphrase – Holmes does not *really* have the property of fame in the same way, say, Caligula does. A possibilist could at least say that he is much talked and thought about, so a suitable paraphrase doesn't seem unachievable.

3.2. Fiction and counterfactuals

Since fictional characters are possible individuals, and fictional stories represent series of events in possible worlds, the difference between fictional and counterfactual discourse is not a matter of semantic content or the ontological status of the truth-makers for the discourse. There is, of course, intuitive support for this. Although conversational markers like 'consider the following possibility' and 'once upon a time' conventionally serve to signal whether what follows is a counterfactual claim or a piece of fiction, respectively, it seems surprising – to say the least – that this would affect the nature of the proposition expressed by or metaphysical commitments of a subsequent utterance of 'there used to be life on Mars'.

That doesn't mean that there is *no* difference. Counterfactual talk and fictional storytelling are distinguished by felicity constraints on utterances and rules for updating or managing one's beliefs relative to new information. Counterfactual talk should be constrained by the requirement that the worlds of the context set be potential candidates for being the *actual* world; that is, counterfactual talk is ultimately about the actual world, and utterances of counterfactuals are felicitous only to the extent that they inform agents of *actual* states of affairs. Fiction is not so constrained.

The following suggestion makes the constraint more precise: Let the context set be a set *W* of worlds that contains (or is supposed to contain) the actual world. Contributions to a discourse are felicitous only if they narrow down the information consistent with the initial presuppositions, modeled as removing worlds from *W*. A counterfactual claim is accordingly felicitous only if it removes at least one world from *W*. How? Suppose *p* and *q* are true in all worlds $w \in W$. A counterfactual claim *u* to the effect that 'if *p* weren't the case, *q* would not have been' asks us to consider what the case might be in the closest worlds *w** accessible from the various worlds *w* where *p* is not the case, and *u* removes from *W* all worlds *w* from which the closest accessible worlds *w** in which although *p* does not hold, *q* still does. For fictional discourse, on the other hand, contributions proceed in the manner of normal, non-counterfactual information updates (though it is of course possible to incorporate counterfactual talk in fiction, as when fictional detectives speculate, counterfactually, about possible explanations for observations at a crime scene).

Insofar as the difference is a matter of pragmatic felicity constraints defined by discourse goals I predict that the distinction between counterfactual and fictional discourse might sometimes be hard to draw – especially in the absence of explicit conversational markers. This seems correct. For a relatively mundane illustration, consider *examples* in (say) a law textbook. Such examples may involve relatively elaborate stage setting and even names for characters. Insofar as their purpose is to illustrate facts about the actual world they are reasonably interpreted as describing counterfactual scenarios, yet they often look and behave like brief fictions. According to my proposal the correct classification depends on what role the story in the example is supposed to play, not on its truth-conditional content.

4. Truth in fiction

In light of the model we can offer a definition of truth-in-fiction. Now, the proposition expressed by a sentence in a fiction (or its truth-conditions) is of course determined by the (unrestricted) complete set of possible worlds. What counts as *true-in-fiction*, however, is the subset that reflects the information the fiction conveys, and which is relevant to interpreting

it, for instance with regard to understanding (non-explicit) character motivations. In 4.1 I introduce a definition of *truth-in-fiction*, and in the following sections I consider some upshots.

4.1. Truth in fiction

Above we assumed that *p* is *true-in-fiction-f* iff *p* is true in the closest possible worlds in which *f* holds. I return to *closeness* below, and will for the moment just assume it. Two other points are worth making. First, the formulation differs from Lewis (1978); according to Lewis, *p* is true in worlds where *f* is *told as known fact*. Lewis introduced the 'told as known fact' qualifier to rule out the possibility that, by cosmic coincidence, the Holmes stories report actual events, thus making the actual world a potential truth-maker. The qualifier would complicate out reference fixing story by relativizing reference to a merely possible storyteller. Since we already rule out the possibility of the actual world being a truth-maker by Doyle's intention to refer to a merely fictional character, however, we do not need the qualifier.

Second, fictional characters are incomplete. Many properties, such as a Holmes's blood type, are never settled in the stories in which those characters feature. Possible individuals are complete. In every world where Holmes exists he has a determinate blood type. It should, however, not be *true-in-the-stories* that Holmes has, say, blood type A rather than B, even if he does in some worlds consistent with the stories. One option we may dismiss is that Holmes is literally an incomplete individual inhabiting an incomplete world. Although fictional characters are under-described, they are often also attributed traits that entail that they are complete. A character described as 'a normal person' has a particular blood type since normal persons have determinate blood types. Incompleteness is not *metaphysical incompleteness*. Accommodating incompleteness is straightforward on our account. If we follow Lewis and deny that there must be a uniquely closest possible world, then *incompleteness* just means that different worlds consistent with the story are equally good truth-maker candidates, including worlds that differ only with respect to Holmes's blood type. These worlds also agree on everything explicitly stated in or entailed by the story, including Holmes being a detective and not eight feet tall, and also on a lot of stuff not explicitly asserted by (but not incompatible with) the

stories, such as the usual laws of nature and that Edinburgh is to the north of London.

Let W be the set of equidistant worlds representing fiction *f*. We define *truth-in-f* by quantifying over the members of W as follows:

> p is true-in-*f* iff p is true in all worlds $w \in W$
>
> p is false-in-*f* iff p is true in *none* of the worlds $w \in W$.
>
> p is neither-true-nor-false-in-*f* iff p is true in some worlds $w \in W$ but not others.

'Holmes is a detective' is thus true-in-the-fiction, 'Holmes is five feet tall' is false, and 'Holmes has blood type A' is neither, unless of course it is, unknown to us at present, entailed by some property actually ascribed to Holmes in worlds where the usual facts about biology hold.

The possibilist definition of truth-in-fiction has sometimes been criticized (cf. Sainsbury 2010; Wright 2014) for helping itself to controversial resources, such as *supervaluationism* (see Stone 2010).[5] Supervaluationism would be needed only if we let every world in which some possible individual – a different one in different worlds – satisfied the descriptions associated with the Holmes stories be a truth-maker for those stories; in that case, sentences of the form 'Holmes is *F*' would be associated with

---

[5] Stone and Wright both notice similarities between the problem of selecting a referent and *the problem of the many*, which can be illustrated as follows (cf. Lewis 1993): A cloud is an aggregate of droplets. At the outskirts of the cloud the density of droplets gradually falls off, but since the boundaries are vague it is impossible to tell where the boundaries of the cloud actually are. As a consequence, many different aggregates are equally good candidates to be the cloud, and we seem to have no means of identifying the cloud with any one of these rather than another. But all of the aggregates are clouds. And then we have many clouds, even though there is, of course, only one. The suggestive parallel to the selection problem breaks down quickly, however. Even if we developed a strategy for picking out a particular aggregate – perhaps an arbitrary one – as the value of 'that cloud' it would not solve the problem that the other aggregates are, by definition, clouds and that the world ends up containing vastly many more clouds than it should. The *problem of the many* is the ontological problem that the number of clouds is too large, not a problem of how to pick out particular aggregates. Though we need, as discussed in Section 2, to select a single referent for 'Holmes' among multiple equally good candidates, the existence of multiple candidates is not itself a problem for us as long as we can select one of them, as I argued we can.

multiple singular propositions containing different referents, and supervaluationism or similar means would be needed to evaluate these sentences. On our account, although there are multiple worlds among the truth-makers for sentences involving 'Holmes' where Holmes has somewhat different properties, it is *Holmes* who has these properties in the relevant worlds, not his surrogates, counterparts or doppelgangers. Worlds where surrogates, counterparts or doppelgangers have these properties are not truth-makers for the Holmes stories. We do not need supervaluationism.

Indeed, *truth-in-fiction* is, formally, just a standard box operator from modal logic ranging over the restricted set of closest worlds where the events described in the fiction take place, and '*p* is *false-in-fiction-f*' is equivalent to 'it is true-in-fiction-*f* that not-*p*.' Since 'Holmes has blood type A' is true in some but not all of these worlds it is not true-in-the-fiction, but it is *consistent with the fiction*: We could define an operator 'consistent-with-the-fiction' from 'true-in-the-fiction' the way we define the possibility operator from the necessity operator and negation.[6] The tautology 'Holmes either has blood type A or he doesn't,' however, is true-in-the-fiction since it is true in all the truth-making worlds. The consequence that all tautologies are true in any fiction has been criticized (cf. Proudfoot 2006, 11). However, given that the connectives mean the same in fiction as in ordinary speech all tautologies *should* be true in all fictions. We would have a problem if *true-in-fiction-f* distributed over the disjunction – that is, if *true-in-fiction* $(p \vee \neg p)$ entailed *true-in-fiction-p* $\vee$ *true-in-fiction-*$\neg p$ – but, insofar as *true-in-fiction* is treated as a restricted box operator, it doesn't. $\Box p \vee \Box q$ is not a consequence of $\Box(p \vee q)$ in any normal modal logic.

An advantage of our account is that what is *true-in-fiction* evolves over the course of creating a story. When the author adds information to the stories the set of worlds that reflect the stories – the *context set* of the stories – shrinks, and information previously consistent-with-the-fiction becomes

[6] Alternatively, assuming *closeness*, we could treat truth-in-fiction as a version of counterfactual reasoning. Let Γ be the set of claims asserted in the fiction. p is true according to the fiction iff $\Gamma \Box\!\!\rightarrow p$. Given the incompleteness of fictional characters we would then have to deny Conditional Excluded Middle, $(\Gamma \Box\!\!\rightarrow p) \vee (\Gamma \Box\!\!\rightarrow \neg p)$; claims about Holmes's blood type, for instance, will not have a determinate truth-value.

true-in-the-fiction. By the same token, our account has no problem explaining or predicting continuity across books involving the same universe, or other authors picking up where Doyle left off, inventing new stories involving the (numerically) same characters and settings. The recent TV series locating Holmes in modern-day London, for instance, might be thought of as counterfactual Holmes stories involving the (numerically) same characters. Similarly, Alban Berg's Wozzeck, Manfred Gurlitt's Wozzeck, and Georg Büchner's Woyzeck inhabit slightly different (sets of) worlds that are truth-makers for slightly different stories, but may still concern the (numerically) same individual located in those different worlds. (How to determine whether a character portrayed in one story is the same one portrayed in another will presumably depend either on certain psychological facts about the author, or causal links between an author's portrayal and previous introductions of characters.)

Of course, discussions may be had about what counts as indeterminate according to the stories, and the distance between the truth-making worlds for a fiction and the actual world may be different than the author intended if the author relies on false beliefs about, say, biology or geography, or if the story contains unintended plot holes or anachronisms. This is not by itself a problem for our account. We should, however, note that when thinking about fictions containing discontinuities or anachronisms, we are often inclined to engage in repair and, if possible, disregard such errors. I have no detailed story to tell about what (pragmatic) principles we rely on in such cases, but our model is able to accommodate various suggestions.

4.2. Closeness and indeterminacy

The *context set* for the stories, as discussed above, contains all worlds consistent with everything stated in a fiction, and is not limited by *closeness* – it should be entirely up to an author to suddenly and coherently decide that Greenland doesn't exist in her fiction, for instance, or that different laws of nature hold. Yet when interpreting a story we often assume that – unless contradicted by the story – ordinary facts hold, such as basic geography or laws of nature. I take it that such assumptions play an important role in making sense of and enjoying fiction – at least speculations about what characters could or should have done rely on importing unstated assumptions about the real world such as standard laws or

generalizations about psychology. The worlds that determine *truth-in-fiction* should be a subset of the context set rather than the whole, identified by the proper importation of facts about the real world, according to some parameter.

Closeness is an obvious candidate for adding the appropriate restrictions; that is, the worlds that determine *truth-in-fiction-f* are those consistent with what is stated in or entailed by *f* and are, on balance, *closest* to the actual world (or at least do not gratuitously depart from it). Assuming *closeness* would mean that speculations about details of relevant events not explicitly detailed in the stories may have determinate correctness conditions. And intuitively, we treat them as if they do; even speculations for instance about Holmes's sexual orientation or whether he had a neurodevelopmental disorder often seem to give rise to genuinely factual disputes, and insofar as what is *true-in-the-stories* is determined by *closeness*, claims about these matters may indeed (but may of course not) have a determinate truth value, depending on whether enough information is given in the stories, in combination with facts imported from the actual world, to settle the issue.

It is also worth pointing out that closeness may be incorporated in different ways. Lewis (1978) suggests relativizing *truth-in-fiction* to *the beliefs prevalent in an author's community* rather than to the actual world.[7] One possible upshot of doing so is expansion of the set of truth-making worlds. If the author's community does not have a prevalent belief about whether some fact *p* holds or not of the actual world, the truth-making set would have to include both *p* and ¬*p* worlds, thus introducing more indeterminacy.

Now, possibilism is not committed to *closeness*. In principle, a possibilist might define truth-in-*F* over the whole context set for *F* – what's explicitly stated (and entailed) in the fiction and nothing else – or use any other means to restrict the set of truth-making worlds. One could for instance imagine restricting the context set to the worlds that are most *aesthetically pleasing*, although it may in that case of course be difficult to formulate conditions for whether a particular fact ought to be imported.

---

[7] The question of whether the set of truth-making worlds should be relativized this way reflects prominent and – I take it – open debates among literary theorists. It is not particularly difficult to adjust the definition of *truth-in-fiction* to either option.

Indeed, given that i) providing truth-conditions for individual sentences does not depend on issues related to interpreting the fiction as a whole, and ii) our model of fiction as a type of discourse does not essentially rely on identifying the relevant truth-making subset, formulating the conditions for which facts to import is not obviously a greater or more pressing challenge for me than for other theories that appeal to *truth-in-fiction*. That said, I think many objections to *closeness* are unpersuasive. In 4.2.1 and 4.2.2 I sketch how we may answer some common complaints.

4.2.1. Accidental imports

It has been argued that *closeness* automatically imports facts we do not want to import (see Proudfoot 2006). One may for instance wonder whether it ought to be *true-in-Hamlet* that Sneferu built pyramids, which it probably is if *truth-in-Hamlet* is defined in terms of the closest worlds in which the events described in *Hamlet* take place. But although that consequence may be surprising, it is even in principle not potentially *problematic* – if there were good reasons to think it shouldn't be true-in-*Hamlet* that Sneferu built pyramids, then those reasons would also block the importation of that fact. The mere intuition stemming from the observation that Egyptian pharaohs are *irrelevant* to *Hamlet* is, I think, moot; it *is* irrelevant (otherwise it wouldn't get imported by default), but I see no good reason why irrelevance should imply truth-value gaps. Indeed, if pyramids ever became a question in discussions of *Hamlet* – for whatever reason – it seems legitimate to point out that *Hamlet* is supposed to take place in the actual world (that is, the worlds closest to the actual), and in the actual world Sneferu built pyramids.

Proudfoot (2006) sharpens the worry with examples of *bounded fictions*. Suppose we have a fiction *G* depicting a fictional war in Europe. Though the fiction does not concern England per se, it includes a map of part of downtown Cambridge, with the marked location of a number *n* of enemy tanks. It is accordingly *true-in-G* that there are *n* enemy tanks in the depicted part of Cambridge. Proudfoot's worry, though, is that it is *also* true-in-*G*, if truth-in-*G* is defined as the closest worlds consistent with the fiction, that there are no tanks in downtown Cambridge *outside* of the depicted area, ostensibly since worlds where as much as possible is kept the same as in the actual world except for the *n* depicted tanks are worlds where

there are no tanks outside of the bounds of the given map ('no-tanks worlds').

This result seems wrong.[8] But it is not clearly an objection to using *closeness* as a measure for truth-in-fiction rather than to using a particular *measure* for closeness. I suspect we have the intuition that it shouldn't be true-in-*G* that there are no tanks outside the depicted area because, intuitively and independently of fiction, we *don't* think that the *closest* worlds where there are *n* tanks in the depicted area are no-tanks worlds; hence, a measure of closeness that has the no-tanks result is simply the wrong measure.[9] We cannot just *keep any truth about the actual world we can* when introducing the tanks; we also need to supplement that change for instance with plausible stories of how the tanks got there. That 'there are no tanks outside the depicted area' is true of the actual world does not mean that no-tanks worlds are closer when we also make the background changes needed to get them there – in the closest worlds in which England is invaded, for instance, the invaders presumably employ more than *n* tanks. Since there seems to be many equally plausible background stories, it is indeterminate what is the case outside of the depicted area.

One upshot is that it will be difficult to provide a formal measure of closeness; a proper definition that takes into account *plausible background histories* (or consequences) would presumably require a formalization of *relevance* and probably a solution to some version of the *frame problem*. Fortunately, for the purposes for which we need *truth-in-fiction* – to interpret and fill in blanks in a reasonably systematic way – we don't need a formal criterion, but can be rather pragmatic about the matter. The less precise the measure for closeness is, the more claims will end up neither true nor false in the story, but it is not clear why that would be a *bad* thing.

---

[8] Proudfoot's primary point, though, is that possibilism cannot respect the *boundedness* of such fictions; i.e. exclude from being true-in-the-fiction any fact beyond the boundaries. She is right that possibilists cannot (easily) recognize such bounds, but as argued above this is not something that should worry them.

[9] The discussion of similarity relations is vast, and I cannot go into details (see Lewis 1979 for some considerations). Using *closeness* for truth-in-fiction does of course assume that some ordering can be achieved, but unlike other issues for which closeness matters (counterfactuals, perhaps verisimilitude) we don't need it to be very fine-grained.

Closeness still yields precision enough in the vast majority of cases to use it as a tool for interpreting fictions.

4.2.2. Competing models

The challenges to realism raised by Everett (2005) require further refinements. Consider fictions where it is unclear whether certain characters actually occur, such as Tatyana Tolstaya's *The Slynx*, in which "in the end, I think, it is pretty much left open whether or not there really is a Slynx" (Everett 2005, 630). Now, the set of possible worlds identified by Tolstaya does indeed contain both worlds where there is and worlds where there isn't a Slynx,[10] but I do not think we should conclude that it is neither *true-in-The-Slynx* nor *false-in-The-Slynx* that there is one; at least that is not the complete story. Holmes's blood type is indeterminate because it doesn't matter to the story; the existence or not of the Slynx *does* matter to Tolstaya's story and will matter significantly to how we understand other parts of it. Moreover, whereas worlds where Holmes have different blood types are equally close to the actual world, the non-Slynx worlds seem *prima facie* closer than the Slynx worlds on the (crucial) assumption that they are equally consistent with the rest of the story.

I think the correct thing to say is rather that the set of truth-making worlds are either the set of Slynx-containing worlds, or the set of those that do not contain a Slynx, but not both. Rather, the story leaves it unclear or indeterminate which of these sets contains worlds that are overall i) closer, and ii) cohere better with what is stated in the fiction. Hence, we get *competing models* of the fiction, and perhaps no clear means for deciding that one of them is the right one. Such indeterminacy is relatively common. In many cases, one may perhaps say that a text can give us multiple different fictions with very different answers to what is true-in-the-fiction. Another example is Patrick Ness's *A Monster Calls*, which may (legitimately) be read as a fantasy story and as a metaphor-loaded description of the main character's psychological processing of his life situation.

Competing models are also needed for Everett's *Frackworld* example, in which "[n]o one was absolutely sure whether Frick and Frack were really

---

[10] Thus, for the question of whether The Slynx is a fictional character the answer is 'yes'; the question is whether this character in fact appears in the story(-worlds).

the same person or not" (Everett 2005, 629). Now, insofar as Frick = Frack is either necessarily true or necessarily false, the competing interpretations require disjoint context sets. The interesting wrinkle in this case is that on the interpretation on which 'Frick' and 'Frack' refer to a single individual, the names refer to a different individual than they do on the interpretation on which they refer to different ones. Accordingly, relative to each interpretation a sentence 'Frick is *F*" will express different propositions, and the *meaning* of the sentences in the story will depend on what interpretation we assume. But this upshot is as expected if we assume that names are directly referential; an author of a story like this would indeed write sentences that are genuinely ambiguous, without having a determinate intention with regard to which interpretation is correct. Accordingly, I do not find the example to be a particularly worrisome.

4.3. Contradictions and surrealism

Sometimes storytellers make mistakes and locate the truth-making worlds for a fiction further away from the actual world than intended. But sometimes fictions also contain outright contradictions and impossibilities, true in *no* world. Inconsistent fictions and inconsistent character developments are commonly raised objections to possibilism. Now, the problem is at least somewhat less pressing on the account developed here. If the reference of a name for a fictional character were fixed by description, then ascriptions of contradictory properties would entail that no referent is picked out. However, insofar as reference is determined by stipulation, independently of the properties eventually ascribed, ascription of inconsistent properties is no obstacle to *referring* to that character. Instead, inconsistent ascriptions become a problem for defining truth-in-the-fiction and identifying an appropriate context set.

Though contradictory fictions are tricky, the problem should not be overstated. First, the problem arises from *genuine* contradictions. Watson's infamous war wound, which appears to move around over the course of the stories, does not generate a genuine contradiction – there *are* worlds where war wounds move around; they're just further away than we generally assume the Holmes worlds to be. We often respond to such cases with local repairs, but that is because we (for pragmatic reasons) choose to – the repairs are not *forced* upon us by the fiction. Certain genuine contradictions

may also be solved – as Lewis (1983) suggests – by compartmentalization, defining truth for *parts* of the fiction rather than the whole, for instance. (A further resource is appeal to competing models, as suggested above.) Intuitively, compartmentalization and local repairs reflect ordinary practice when facing continuity problems or contradictions, and having to resort to such measures is at least no objection to the possibilist account of how claims made in fiction can be meaningful and have truth-conditions.

Nor is it an obvious objection to possibilism that it has nothing to say about how to understand surreal stories or events. In the case of apparent violations of the laws of logic, for instance, we *should* be inclined to say that it becomes unclear what should be considered true-in-the-fiction and even what propositions the text is expressing. *Through the Looking Glass*, for instance, contains passages where it is less than clear that we have a firm grasp of the truth-conditional content. The unsurprising lesson to draw from such examples is that fiction serves purposes and exhibits qualities not tied to assertions (or pretense assertions) of matters of fact – its propositional, truth-conditional content – that may even *depend on* us being unable to pinpoint what claim it in fact makes. Similarly, that possible worlds can be used to define truth-conditions for claims in fiction and model important features of narrative fictions, does not mean that they are particularly useful for capturing the value of, say, poetry or humor. Our model is intended as a basic model of fictional storytelling as a type of discourse, not a template for analyzing *literary appreciation* or explaining all literary tools at an author's disposal.

One source of ostensibly problematic examples come from *embedded fictions*. Now, embedded fictions (modeled using worlds accessible from the worlds of the first-order fiction) are themselves not particularly problematic for possibilism, but sometimes fictions blur the distinction and let characters that are real according to the fiction interact with characters that are fictional according to the fiction, or enter the fiction-within-the-fiction. Insofar as 'fictional' entails 'non-actual', such scenarios are contradictory; in fact, they cannot even be meaningfully described, any more than a fiction using the phrase 'married bachelor' in a fiction describes a married bachelor. However, nothing in possibilism rules out worlds where individuals have different relationships with the scenarios they describe in their stories; the stories-within-the-fictions would not be *fictions* (for the participants in the first-order fiction) in *our* sense, and their characters not strictly

speaking *fictional characters* (insofar as they are available for causal interaction, which genuinely fictional characters, by definition, are not).[11] Fictions that describe such encounters or interactions do, by the meaning of 'fiction' for *us*, accordingly not describe interactions with fictions, but perhaps with fictions* – contents of stories that are not separated from life in the world of the fiction by being non-actual relative to that world, but by some other barrier.

That leaves us with examples of blatantly contradictory fictions such as *Sylvan's Box* (cf. Priest 2005), which purports to describe an object that is both there and isn't there at the same time, or Everett's (2005) examples where e.g. the symmetry of identity does not hold. In these examples the contradictions are explicit and central elements of the story, and compartmentalization or local repairs do not help. In such rare cases I am inclined to bite the bullet and deny that the stories are, in fact, best interpreted as completely meaningfully describing a contradictory state of affairs (see Hanley 2004).[12] *Truth-in-fiction* is a species of *truth*, and unless we are dialetheists 'it is both there and not there the same time' isn't in any sense *true*, not even true-in-fiction, even if the author claims otherwise.

5. Conclusion

I have endeavored to provide a partial defense of a possibilist approach to fiction against some common objections, and to (cursorily) elaborate on some potential advantages. Major obstacles remain. I have not discussed possibilism's potential commitment to modal realism. Nor have I developed my response to the selection problem in detail. But insofar as these questions can be answered, possibilism has some attractive features. First,

---

[11] That is: There is nothing contradictory about me interacting with Holmes, and there are worlds where I do. But it *is* a contradiction in terms to say that I might (causally) interact with a fictional character as such, and should be on any acceptable analysis of 'fiction'.

[12] Indeed, Everett's examples rather call for repair. A non-symmetric relation is by definition not an identity relation, so an author purporting to describe non-symmetric identity is using the word 'identity' wrong; what he or she describes is in fact a different relation, for instance a supervenience relation.

it provides a straightforward account of the *meaning* of claims made in fiction, which are assigned the same kind of semantic content as claims outside of fiction but, of course, supposed to be interpreted under assumptions that the interpreter ordinarily ought to be aware do not *actually* hold. Second, possibilism underpins a fruitful model for fictional discourse. Of course, further challenges remain, especially since fictional narratives are often far more complex than the toy examples discussed here. We have had nothing to say for instance about the role of the narrator (which presumably requires at least an extra level in the model), how to understand, say, fourth wall violations, or importation of assumptions related to genre conventions.

Finally, I have made no attempt to compare the possibilist analysis with its competitors. My goal has accordingly been rather modest, and given the outstanding challenges to possibilism – in particular concerning metaphysical commitments – this article is perhaps best understood as an attempt to help clarify where the real battle should take place.

Acknowledgments

The research behind this article was made possible in part through a grant from the SASPRO/FP7-Marie Curie Actions-COFUND scheme [grant number 0086/01/03/-b] as well as VEGA grant 2/0049/16: Fictionalism in philosophy and science (primary investigator Marián Zouhar). I am grateful to participants at the Philosophy of Language (1): Semantics of Fictional Discourse conference in October, 2016, as well as anonymous referees, for comments on earlier versions of this paper.

References

BRECKENRIDGE, W. & MAGIDOR, O. (2012): Arbitrary Reference. *Philosophical Studies* 158, No. 3, 377-400.

BROCK, S. (2002): Fictionalism about Fictional Characters. *Noûs* 36, 1-21.

EVERETT, A. (2005). Against Fictional Realism. *The Journal of Philosophy* 102, No. 12, 624-649.

HANLEY, R. (2004): As Good As It Gets: Lewis on Truth in Fiction. *Australasian Journal of Philosophy* 82, 112-128.

HOWELL, R. (1979): Fictional Objects: How They Are and How They Aren't. *Poetics* 8, 129-177.

KÖLBEL, M. (2002): *Truth without Objectivity*. London & New York: Routledge.

KRIPKE, S. (2011): Vacuous Names and Fictional Entities. In: *Philosophical Troubles: Collected Papers Volume 1*. Oxford: Oxford University Press, 52-75.

LEWIS, D. (1978): Truth in Fiction. *American Philosophical Quarterly* 15, 37-46.

LEWIS, D. (1979): Counterfactual Dependence and Time's Arrow. *Noûs* 13, 455-476.

LEWIS, D. (1983): Postscripts to Truth in Fiction. In: *Philosophical Papers*. Oxford: Oxford University Press, 276-280.

LEWIS, D. (1993): Many, But Almost One. In: Bacon, J., Campbell, K. & Reinhardt, L. (eds.): *Ontology, Causality and Mind*. Cambridge: Cambridge University Press, 23-42.

MACFARLANE, J. (2005): Making Sense of Relative Truth. *Proceedings of the Aristotelian Society* 105, 321-339.

PRIEST, G. (2005): *Towards Non-Being. The Logic and Metaphysics of Intentionality*. Oxford: Clarendon Press.

PROUDFOOT, D. (2006): Possible Worlds Semantics and Fiction. *Journal of Philosophical Logic* 35, 9-40.

SAINSBURY, M. (2010): *Fiction and Fictionalism*. London: Routledge.

STALNAKER, R. (1978): Assertion. In: Cole, P. (ed.): *Syntax and Semantics, Vol 9*. New York: Academic Press, 315-332.

STONE, J. (2010): Harry Potter and the Spectre of Imprecision. *Analysis* 70, 4, 638-644.

SULLIVAN, A. (2005): Rigid Designation, Direct Reference, and Modal Metaphysics. *Pacific Philosophical Quarterly* 86, 577-599.

THOMASSON, A. (1999): *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

VAN INWAGEN, P. (1977): Creatures of Fiction. *American Philosophical Quarterly* 4, 299-308.

WRIGHT, B. (2014): Many, But Almost Holmes? In: García-Carpintero, M. & Martí, G. (eds.): *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 290-306.

Organon F 24 (3) 2017: 366-383

A Revisionary View of Texts, Textual Meaning, and Fictional Characters

ANDERS PETTERSSON[1]

ABSTRACT: Using ideas from John Searle, Roy Harris, Michael Reddy, and Nelson Goodman, I argue that texts, such as they are commonly conceived, lack brute existence. The common idea of texts is a conceptual construction which is useful in practical everyday contexts but not in serious theorizing, where it creates illusions and contradictions. One of these illusions is the idea of an objective textual meaning, a meaning which is "in the text": what we actually have in the way of textual meaning are the ideas of various persons – authors, readers, and commentators -- about the meaning of the text. When applied to fictional characters, this way of viewing things explains why it makes sense to regard fictional characters as being created and as lacking brute existence.

KEYWORDS: The ontology of texts – textual meaning – fictional characters – John Searle – Roy Harris.

This paper will introduce and explain a partly new perspective on texts, textual meaning, and fictional characters.[2] The discussion will

[1] Received: 17 March 2017 / Accepted: 7 June 2017

✉ Anders Pettersson
Department of Culture and Media Studies
Umeå University
Humanisthuset, Biblioteksgränd 3, 901 87 Umeå, Sweden
e-mail: anders.pettersson@umu.se

[2] The general perspective is presented much more comprehensively in my book – Pettersson (2017). However, fictional characters are not discussed in the monograph,

finally lead up to a reflection on two alternative conceptions of verbal communication.

The word "text" refers, here, to any whole piece of verbal utterance or discourse – short or long, oral or written, literary or non-literary. Flaubert's *Madame Bovary* (1857) will be used as an example of a text and Madame Bovary's husband, Charles Bovary, as the main example of a fictional character.

Charles Bovary is first introduced in the opening sentence of Flaubert's text, a sentence which reads, in Flaubert's original French,

> Nous étions à l'Étude, quand le Proviseur entra, suivi d'un nouveau habillé en bourgeois et d'un garçon de classe qui portait un grand pupitre.[3]

and in Eleanor Marx-Aveling's English translation,

> We were in class when the head-master came in, followed by a "new fellow", not wearing the school uniform, and a school servant carrying a large desk.[4]

The new fellow is Charles Bovary.

1. Brute existence and mentally constructed existence

Ontological considerations play an important role for the understanding of texts, textual meaning, and fictional characters. In particular, there is a distinction worth drawing between two kinds of existence, which could be called "brute existence" and "mentally constructed existence". A planet is a good example of an entity enjoying brute existence: a planet is simply

---

and the ontology of what I call "commentator's meaning" is treated differently here than in the book.

[3] See Project Gutenberg. Gustave Flaubert, *Madame Bovary*, accessed 16 March 2017, http://www.gutenberg.org/cache/epub/14155/pg14155.html.

[4] See Project Gutenberg. Gustave Flaubert, *Madame Bovary*, trans. Eleanor Marx-Aveling, accessed 16 March 2017, http://www.gutenberg.org/files/2413/2413-h/2413-h.html.

there, irrespective of what humans think or say. The contents of a daydream, on the other hand, do not possess that mode of being. They form part of the imaginings of the daydreamer, who can also change the contents at will. Still, the contents of a daydream exist, in a sense. They enjoy a mentally constructed existence. [5]

The distinction between brute and mentally constructed existence is a modification of John Searle's distinction between brute and institutional facts (see, especially, Searle 1995, 1-2). The main difference between the two dichotomies is that Searle only considers mentally constructed existence which involves social agreement and thereby results in institutional facts, while the mentally constructed also includes purely individual mental products. Traffic rules can serve as examples of institutional facts in Searle's sense, while daydreams can not. Traffic rules are not features of brute reality, not parts of the structure of the universe: they depend on institutionalized human agreement for their very being. Still, we would certainly like to say that traffic rules exist.

2. The mode of existence of languages

What are the brute realities in connection with a language, for example, English? If by "English" we mean English as used, verbal communication in English, what are brutely real must be the physical sounds or physical marks produced and received and the mental processes in senders and receivers which underlie the ideas associated with these sounds or marks.

English as a language system, a "grammar", is something different and something which lacks brute existence. A grammar of English is a linguistic description of the knowledge presumed to be at the command of a fully competent speaker of English. A grammar is made up of a phonetics, a lexicon, a syntax, and so forth: it consists of phonemes, words, sentences,

[5] I will take the existence of an outer world for granted and suppose that the outer world contains, at least, physical objects and biological organisms, including humans. While I will think of human mental processes as brutely existing, I will regard the contents of thoughts, such as, for example the contents of a daydream, as mental constructs.

and suchlike.[6] The phonemes, words, sentences, etc. are obviously mental constructs. They belong to the linguistic description of a system supposed to be employed when communication in English takes place. They are, so to speak, elements of the map constructed by linguists, not elements of the linguistic terrain supposed to be mapped.

Many theorists seem to deny this. Many like to say, for example, that people utter words and write down sentences (see, e.g., Lycan 2008, 72; and Chaudhuri 2010, 11). However, while we are certainly used to saying, casually, that people do such things, this cannot be literally true. An English sentence is not a physical phenomenon, and one cannot very well utter or write down something which lacks physical existence.[7]

3. Textual meaning

Senders mean something by the texts they issue, and receivers achieve some kind of understanding of the texts. When writing and publishing *Madame Bovary*, Flaubert meant something by the physical marks he produced and expected to be reproduced and disseminated. He will no doubt have intended the marks to be interpreted as standing for meaningful French words and sentences, and he must also have entertained some wider, albeit vaguer, expectations concerning the overall import of his novel. Readers of *Madame Bovary*, for their part, attempt to understand Flaubert's text. Deciphering the physical marks in their respective copy of the novel, they construe a verbal understanding of its words and sentences, and they also seek a point or points in Flaubert's story, something that can make the novel meaningful to them. I will call such things as Flaubert's

---

6 For the concept of a grammar see, e.g., Fromkin (2000, 7), or Baker & Hengeveld (2012, 18-19).

7 Many philosophers would object that *tokens* of words and sentences are in fact concrete and can be uttered, while word- and sentence-*types* are abstract. I do not share that idea. The view requires us to think of physical sound and physical marks as being linguistic entities, but a soundwave or a configuration of ink cannot very well, in itself, belong to a language and be a word or a sentence, although philosophers often presuppose that it can – see, e.g., Bach & Harnish (1979, 285, note 1), and Wetzel (2014, sec. 1.1).

intended meaning "sender's meaning" and receivers' constructions of meaning "receiver's meaning".

There are many possible ways of making the concepts of sender's meaning and receiver's meaning more precise if one wishes to do so with some specific research objective in mind. However, since my purpose here is not of any narrowly circumscribed kind I will leave the two concepts open and intuitive. It should be emphasized, though, that sender's meaning and receiver's meaning are mental entities, and that each receiver's meaning will be his or her own. While it is possible to speak of the one and only sender's meaning of *Madame Bovary*, it can never make sense to speak of the one and only receiver's meaning.

One may believe that there exist not only the sender's meaning and the various receivers' meanings, but also the true meaning of the text. Most theorists take it for granted that a text has some definite body of meaning associated with it, whether the theorist conceives of that meaning itself as being indeterminate (as poststructuralists typically do) or as being more or less determinate (like most adherents of other schools). Yet it is difficult to see how objectively true textual meaning could come into existence. Where *Madame Bovary* is concerned, a third party – a critic, say, or a schoolteacher – can certainly present an interpretation of the text. But that interpretation will inevitably just represent one more idea about the textual meaning of the novel – not a sender's meaning or a receiver's meaning, since the critic or teacher will be placed outside the situation of actual literary communication, but something which can be called a "commentator's meaning".

The belief in true textual meaning, the belief that every text has some definite body of meaning associated with it, seems to be very nearly universal. One of the rather few people who has challenged such a view in a conscious and reflective manner is the British linguist Roy Harris. For Harris, there is the sender's understanding of the meaning and the receiver's understanding, and if these do not coincide there is no supreme authority to appeal to. "The signs that occur in first-order communication are those that the participants construe as occurring", he writes, "and what is signified is what the participants construe as having been signified. *There is no higher court of appeal*" (Harris 1998, 145). Harris holds that "where two or more participants are involved a message must be open to two or more interpretations. And these cannot be guaranteed to coincide. Furthermore,

where they conflict, no one interpretation holds a privileged position *vis-à-vis* another" (Harris 1998, 84).

I fully agree with Harris, and I will return to these matters later in the essay in order to defend the perspective on textual meaning sketched here. But let us first look at the implications of the way of thinking about communication presented thus far for the understanding of a special kind of element of textual meaning: fictional characters. I will add a revisionary view of fictional characters to the revisionary view of textual meaning just presented.

4. The mode of existence of fictional characters

It should be clear that a fictional character, for example, Charles Bovary, lacks brute existence. That is what most fundamentally distinguishes fictional characters from genuine human beings. In the most simple and straightforward sense of "exist", then, fictional characters do not exist. Yet fictional characters obviously enjoy a mentally constructed existence. Flaubert mentally constructed Charles Bovary. He had ideas, imaginings, in which Charles Bovary figured, and eventually Flaubert made ideas about Charles Bovary part of the textual meaning of his novel *Madame Bovary*. Ideas about Charles Bovary indubitably form part of the sender's meaning of the text, and such ideas will also have to form part of any defensible receiver's meaning and commentator's meaning. These brief remarks seem to me to answer all basic questions about the ontology of fictional characters. Not to put too fine a point on it: Charles Bovary does not exist. What exist are ideas about Charles Bovary.

Ideas about Charles Bovary should not be understood as being ideas about some extra entity existing apart from the ideas. As a reader of *Madame Bovary*, I form an idea of the fictive situation described in the first sentence of Flaubert's novel. My idea of the fictive situation features, among other things, a classroom, a class servant, and a new boy who will later prove to be called Charles Bovary. The classroom, the class servant, Charles Bovary, et cetera, are some of the constituents of my idea. There is no need to reckon with any extra, somehow independent, entities forming the referents of my idea – a fictive classroom, a fictive class servant, the fictional character Charles Bovary, and suchlike. My idea is just such-and-such an idea, an idea featuring such-and-such elements.

Once again I build on previous thinkers. This time Nelson Goodman is the key reference. Goodman maintained that there is nothing that a picture of a unicorn or a picture of Pickwick is a picture *of;* rather, a picture of a unicorn is a special kind of picture: a unicorn-picture. Similarly, a picture of Pickwick is a special kind of picture: a Pickwick-picture. The existence of unicorn-pictures and Pickwick-pictures does not, according to Goodman – with whom I entirely agree –, force us to suppose that unicorns, or fictional characters like Pickwick, enjoy some kind of separate, independent existence (cf. Goodman 1968, 21-22).

Two additional comments. First, note that I do not claim that Charles Bovary exists in the minds of individual people. It is more to the point to describe me as maintaining that Charles Bovary does not exist, not anywhere, but that there can be ideas of Charles Bovary in the minds of many individuals.

Second: it can seem as if we often refer to Charles Bovary in a way which cannot be understood as involving reference to any specific, individual mind. For example, in the Wikipedia article about *Madame Bovary* we read: "Charles Bovary is a shy, oddly dressed teenager arriving at a new school where his new classmates ridicule him."[8] But such an utterance should not be understood as referring to a non-mental Charles Bovary. The anonymous writer may well have thought of the utterance as genuinely referring, but in reality the writer has produced a comment on the novel, conveying commentator's meaning, which is just as mental as sender's meaning and receiver's meaning.[9] Nobody would want to challenge the substance of what the writer says. Still, we are faced with just another person-bound idea about Charles Bovary.[10]

---

[8] See entry "Madame Bovary" in Wikipedia. Available at: https://en.wikipedia.org/wiki/Madame_Bovary.

[9] The Wikipedia writer's utterance has a sender's meaning, of course, but part of this sender's meaning is a commentator's idea about Charles Bovary.

[10] Commentator's meanings can be of different kinds. I understand the utterance in question as meant to convey something more than a purely personal perception: as meant to point to features expected to be experienced by all competent readers of Flaubert's novel.

5. More about fictional characters: a defence of antirealism and creationism

The way of thinking about fictional characters introduced in the previous section differs from the approaches current in the philosophical discussion. It is not possible to enter really deep into these differences here, but I will relate my own standpoint on fictional characters to the current philosophical debate on a couple of points.

Distinctions like the one between brute and mentally constructed existence tend to play a marginal role at best in standard ontology, and they seem to be largely neglected in the philosophy of fictional entities.[11] This may seem surprising, since the question of whether or not fictional characters exist – the question of realism or antirealism about fictional characters – is one of the main issues in the field (see Kroon & Voltolini 2016, secs. 2.1, 2.2). True, everybody will probably be ready to accredit fictional characters with at least a mentally constructed existence, so that the question of realism or antirealism will concern brute existence. Yet it is my impression that not drawing the distinction between brute and mentally constructed existence gives rise to a certain amount of confusion in the debate.

The standpoint advocated here, according to which fictional characters possesss mentally constructed existence but not brute existence, is probably best characterized as a variety of antirealism about fictional characters. Arguments have been raised against antirealism. A particularly favoured argument rests on the observation that there are truths about fictional characters. It is commonly thought that if you can make a true statement about some singular entity, that singular entity must exist: if the statement "Bratislava is the capital of the Slovak Republic" is true, then Bratislava must exist (cf. e.g., Hale 1987, 11; and Effingham 2013, 172). If this is a valid principle, it seems that Charles Bovary must exist, for it appears hard to deny the truth of the statement "Charles Bovary is a fictional character".[12]

---

[11] The philosophy of fictional entities is not a specialty of mine. In the rest of this section I draw heavily on the overview of the field presented by Fred Kroon and Alberto Voltolini in their article Kroon & Voltolini (2016).

[12] See the discussion of assertions of this kind in Kroon & Voltolini (2016, secs. 2.1.2, 2.1.3).

However, the principle in question obviously presupposes that brute reality is in itself, independently of any human representational schemes, divided into separate objects fit to function as referents of linguistic statements, and it also presupposes that language can reflect the structure of brute reality. Both suppositions are controversial[13] and in my view mistaken. This is not the place to pursue the matter in depth, but the very simplicity of the principle should raise suspicion. If this line of thought were tenable, one could prove the brute existence of such abstract objects as natural numbers just by pointing out that it is true that two and two make four.[14]

Now to another point. Some theorists believe that fictional characters just exist, timelessly as it were. Other theorists, sometimes called "creationists", hold that fictional characters are created – a creationist will maintain, for example, that Charles Bovary was created by Gustave Flaubert.[15] I certainly believe that Charles Bovary was mentally constructed by Flaubert: Flaubert had imaginings about Charles Bovary and he made such imaginings part of his sender's meaning of *Madame Bovary*, actually causing imaginings of this kind to also form part of every defensible reader's meaning and commentator's meaning of his novel. I suppose this way of viewing the matter makes me a creationist of sorts.

---

[13] Thus, e.g., Searle comments, on what he calls conceptual relativity: "Systems of representation, such as vocabularies and conceptual schemes generally, are human creations, and to that extent arbitrary. It is possible to have any number of different systems of representations for representing the same reality" (Searle 1995, 151; and cf. Searle 1995, 163-165). What is or is not a single object will then also depend on the chosen system of representation and not just on brute reality. Regarding language, N. J. Enfield remarks that "language is not a means for *reflecting* how things are, but rather a means for *portraying* it in certain ways" (Enfield 2015, 2).

[14] I should perhaps add that I have no problem at all with accepting the proposition that Charles Bovary is a fictional character. Indeed, when I read the first sentence of *Madame Bovary* the new boy figures in my mental representation of the scene as a boy but also as a fictional character, for I am fully aware of the fact that I am reading a piece of fictional discourse. However, in taking the new boy as a fictional character I take him eo ipso as *lacking* brute existence. Thus the fact that Charles Bovary is a fictional character does not appear, for me, to furnish material for an argument for his brute existence. Quite the opposite.

[15] On creationism, see Kroon & Voltolini (2016, sec. 1.3).

An objection that has been raised against creationists is that they are bad at specifying the identity criteria of fictional characters.[16] It is true that I have not specified any criteria for being Charles Bovary, but I do not believe that any such criteria exist. Charles Bovary does not brutely exist, and therefore he does not actually have any quite specific identity. What exist are ideas about Charles Bovary.[17] To repeat: Flaubert had imaginings about Charles Bovary and he made such imaginings part of his sender's meaning of *Madame Bovary*, actually causing imaginings of this kind to also form part of every defensible reader's meaning and commentator's meaning of his novel. For me, that is the whole story, or at least the central part of the story. (People who have merely heard about Charles Bovary quite vaguely and in second or third hand may also entertain ideas about Charles Bovary.) Identity criteria play no role in this account. [18]

---

[16] Thus Kroon and Voltolini comment, in sec. 1.3, that creationism "makes it hard to see how to individuate a fictional entity".

[17] Kroon and Voltolini also write, still in sec. 1.3, that creationism "fails to account for the idea … that there must be a sense in which fictional objects *actually* have the properties that characterize them in the relevant stories". In my view, there are properties that Charles Bovary has according to the various representations of various people – Flaubert, his readers, et cetera – but these may not all coincide. The perceived properties will probably be in accordance to a large extent, but they may also differ on many points, and there are no properties that Charles Bovary *actually* has in any absolute sense. He lacks brute existence, and people's ideas of Charles Bovary may differ.

[18] At this point it lies near at hand to object that a traffic rule, which is also a mental construct, seems to have a specific content, so that a mental construct can very well have a specific identity. I would say, however, that the fundamental situation is the same in both cases. The existence of the traffic rule is an institutional fact backed up by a formal authority (while the existence of Charles Bovary is not), so there will be an authoritative verbal formulation of the traffic rule, and one will have attempted to make the meaning of the formulation as univocal as possible, i.e., one will have tried to ensure that the sender's meaning and all receivers' meanings will coincide as much as possible (something which was hardly a main concern in connection with *Madame Bovary*). Despite this, people may of course understand the rule differently, and rules are in fact open to interpretation. What is the "true" way of understanding the traffic rule may, ultimately, have to be decided in court. (Concerning the "true" characteristics of Charles Bovary, there is no court to turn to.)

I foresee the objection that I should at least be able to point to criteria for being an idea *about Charles Bovary* as opposed to an idea not pertaining to Charles Bovary. But why should I? I believe I can explain in what sense Charles Bovary exists and in what sense he does not. I have no qualms about speaking of ideas about Charles Bovary, since there are ideas which it appears natural and uncontroversial to call ideas about Charles Bovary, for example, some of the ideas making up Flaubert's sender's meaning of *Madame Bovary* and defensible readers' meanings and commentators' meaning in connection with the novel. But I do not believe that there are any pre-given criteria to unveil which will effect a distinction between ideas that are and are not about Charles Bovary. Where could such criteria conceivably come from? One can certainly impose criteria, but that will be an arbitrary thing to do unless one does so for some quite specific theoretical or practical purpose.

6. Conventions and textual meaning

Let us now leave fictional characters aside and come back to textual meaning. Many will no doubt want to dispute the idea that there is no true meaning associated with a text but "only" a sender's meaning and, possibly, various receivers' meanings and commentators' meanings. Some might even suspect that establishing the true, necessarily non-mental, meaning of *Madame Bovary* will give access to a Charles Bovary very different from the elusive figure whom I have been speaking of.

In the rest of this paper I will discuss three important arguments for the existence of true textual meaning and, naturally, attempt to refute the three arguments. The discussion will eventually lead over into questions about the nature and ontology of texts, and a revisionary view of texts will be added to the revisionary views of textual meaning and fictional characters.

According to the first of the three arguments for the existence of true textual meaning that I will consider, language and context – semantic conventions, pragmatic conventions, cultural context, and so forth – determine the meaning of texts; consequently, a text has a true meaning. Richard Gaskin has recently formulated such an idea by contending, referring to literature, that "the meaning of a work of literature is its original meaning" and

that "the original meaning of a work of literature is a function of the meanings that its component words have in the language at the time of that work's promulgation, of the contemporary significance of the syntactic constructions into which those words are fitted, and of the work's historical and literary context" (Gaskin 2013, 219).

I find that far from credible. Think of *Madame Bovary*. On an account like Gaskin's the true (original) meaning of the novel is supposed to be a function of the meanings of its words and syntactic constructions at the time of publication and of the novel's historical and literary context. But can one really specify all the relevant features of the historical context, all the relevant features of the literary context, all the contemporary meanings of the words forming part of *Madame Bovary*, and the significance of all the syntactic constructions used there? And can one, having done all that, also demonstrate how all these factors *function together* to fix the textual meaning of Flaubert's novel? To my mind, already because of their very enormity none of the five tasks can be actually performed. Nor, in my view, can any of the tasks be performed with any plausible claim to objectivity. But those who assert, like Gaskin, that there is true textual meaning arising through the mechanism just mentioned invariably content themselves with making the assertion. They never make any attempt to prove their point by specifying the concrete linguistic and cultural facts supposed to be relevant and by demonstrating how a textual meaning becomes defined as a function of those facts.

Nobody denies that linguistic and cultural knowledge plays a role for the understanding of *Madame Bovary*. But the idea that language and culture provide the novel with a definite meaning is an entirely different proposition. I find such standpoints unrealistic. I also find them empty, lacking in substance, as long as they remain naked assumptions.

7. Semantics and textual meaning

A second argument for the existence of true, non-personal textual meaning could take linguistic semantics as its starting-point. Linguists tell us that words have word-meanings and that sentences have sentence-meanings. (A sentence-meaning is supposed to be a function of the meanings of the constituent words and the syntactic structure of the sentence – cf., e.g.,

Birner 2013, 24.) There is no reason to dispute the (mentally constructed) existence of word-meanings and sentence-meanings, and this can foster the impression that texts must indeed have definite, impersonal bodies of meaning associated with them, since texts consist of words and sentences possessing word-meanings and sentence-meanings.

I do not believe that such an impression is correct. In my view, when linguists describe word-meanings they try to capture what members of the linguistic community mean by these words when using them. (As Kent Bach and Robert M. Harnish once pointed out, "what words mean is what we mutually believe them to mean" (Bach & Harnish 1979, 133).) Likewise, when linguists describe sentence-meanings, what they actually describe is some central aspects of what senders would typically mean by utterances of these sentences and how receivers would typically understand them. The linguists' meaning ascriptions are thus to be seen as a kind of commentator's meanings: as idealized or generalized characterizations of certain elements of what actual senders and receivers would (supposedly) mean or understand by real utterances of the words or sentences.

In my view, it is consequently not the case that the sentence-meanings described by linguists determine what is meant by real utterances of those sentences. Things are the other way round: the real senders' and receivers' meanings of actual utterances of the sentences determine what the sentences mean. The linguist's map of the language does not determine the makeup of the real-world terrain of discourse in the language. On the contrary: the linguist's map of the language should try to picture, in a useful fashion, linguistically relevant aspects of what is meant, and by what means, in actual communication in the language.

Linguists and philosophers of language sometimes seem to me to turn things upside down, as if they believed that the tail is actually wagging the dog. Linguists like to say that language is governed by rules, meaning the rules formulated by linguists, seemingly oblivious of the circumstance that these rules (or, rather, these observations of linguistic regularities) cannot have come down from some semantic heaven but will have had to derive from their own ideas about the prevalent or correct use of the language. And philosophers of language sometimes seem to take sentence-meanings as the prime facts of language, letting sentence-meanings determine the meaning of actual utterances. Thus Searle has argued that an utterance of

the sentence "Snow is white", if it is an utterance of the sentence worth to be taken seriously, amounts to an assertion that snow is white *because that is what the sentence means*. For him, "the meaning of the sentence 'Snow is white' by itself determines that its appropriate utterance *counts as* a statement to the effect that snow is white" (Searle 2010, 10). Searle does not seem to ask himself what makes the sentence "Snow is white" itself mean what it means.

In this section, I wanted to point to the idea that linguistic semantics shows that there must be some objective meaning associated with real-world texts and utterances, and I wanted to indicate some of my reasons for not sharing the idea. Very much more could certainly be said about semantics, its nature, and its scope, but once again I have touched upon issues that I cannot discuss in depth within the confines of this essay.

8. Texts and textual meaning

A third ostensible reason to believe in the existence of "true" textual meaning comes from our standard way of conceptualizing human commuication. A simple communication model features a sender, a text, and a receiver. Applied to *Madame Bovary*, the picture will look like this:

Flaubert ⟹ *Madame Bovary* ⟹ Reader

But – so this line of thinking goes – a text contains words and meanings. Who would want to deny that *Madame Bovary* contains words and meanings? There must consequently be words and meanings in the text in the middle of the figure, not only in Flaubert and his reader, and it seems that the meaning in the text itself must be non-personal and objective.

The problem with this line of thinking is that our ordinary way of conceptualizing communication is not fit to be taken literally. As the American linguist Michael Reddy has shown, the simple communication model rests on a complex metaphor according to which senders insert their thoughts or feelings into physical objects (texts) from which receivers can then retrieve them. (We are used to supposing that a sender can "put his thoughts into words" and that the receiver can then "get something out of" those words

– see Reddy 1979/1993, 164-201.) But thoughts cannot be taken out of somebody's head, and a physical book does not contain any inner cavities into which such thoughts can be inserted. The ordinary picture of communication is straightforwardly metaphorical and the text, as characterized above, is a contradiction in terms. The *Madame Bovary* figuring in the model will have to be a physical object, since it is supposed to exist on its own outside sender and receiver. The text is also supposed to be possible to read, and in order to be possible to read *Madame Bovary* will have to be a physical entity: one cannot read something which lacks material existence. However, according to the model, the physical object which is *Madame Bovary* also contains non-physical elements: words and meanings are not material entities. *Madame Bovary*, as conceived according to our everyday conceptualization of communication, is thus an ontological monstrosity, physical and non-physical at the same time.

It is actually very easy to transform the ordinary conceptualization of communication into something more intellectually tidy. What exists between sender and receiver is not a text as ordinarily conceived but a physical something, in this case, a physical copy of *Madame Bovary*:

Flaubert ⟹ Physical copy of *Madame Bovary* ⟹ Reader

The physical copy is just physical. There are no words in the physical copy and no meanings, but there are word-ideas and meaning-ideas in the sender and in the receivers. So we are back with only person-bound meanings. The idea of an objective text of such a character that it can encapsulate objective meaning proved illusory.

But what, then, is *Madame Bovary*, if Flaubert's novel is not a physical object supplied with words and meanings? The best alternative way of thinking of *Madame Bovary* may be to put aside entirely the idea that there is ever any such thing as a unitary object that is *the text itself*. There are, instead, a cluster of interrelated entities: there are the physical copies, and there are the various word-ideas and meaning-ideas entertained by the author, the readers, and the commentators. This way of thinking, this alternative perspective, eliminates reference to unitary texts-themselves, but it still gives us everything we need to think or talk about when thinking or talking about texts – for what else is there to refer to, apart from

physical copies, sequences of signs, and textual meaning? The cluster conception of what a text is removes the ontological contradictions surrounding texts, for the physical copies are of course physical through and through, while the sequence of signs and the textual meaning are non-physical through and through. Consequently, the cluster conception does not give rise to the same kind of theoretical illusions as the ordinary conception of a text, for example, the illusion that a physical copy can contain an immaterial meaning. But the cluster conception is certainly a revisionary notion.[19]

It is worth emphasizing that the ordinary conception of texts and the cluster conception are both mental constructs. Both are human ways of conceptualizing certain aspects of the brute facts of human communication. I would also like to add that both conceptions have their pros and cons. Like so many everyday conceptions, the ordinary conception is profoundly illogical but also, because of its very lack of intellectual precision, eminently practical and easy to handle. One would not want to be without the ordinary conception of texts. Yet, if taken seriously, the ordinary conception does not make sense, and it gives rise to aporias and illusions. It is therefore good to have the cluster conception to fall back on whenever theoretical clarity is more important than conversational ease. On the other hand, the amount of precision required by talk in cluster conception terms – talk not about presumedly unitary texts but about more or less specific copies, sequences of signs, and meanings – makes cluster-conception formulations too cumbersome to use in less demanding contexts.

9. Conclusion

Despite the use of *Madame Bovary* as an example, my real focus in this essay has been on the general understanding of verbal communication, literary or non-literary. In the previous section I suggested, following Reddy,

[19] While the idea of a text suggested here is, in this form, original, it is associated with a special family of theories about what a text is, theories that are usually called "eliminativist". About "eliminativist" thinking about texts, see Livingston (2016, sec. 3.2).

that our standard way of conceptualizing communication is in need of radical reform. I have pointed to an alternative way of understanding what is going on when people communicate in speech or writing, and I have used reflections on texts, textual meaning, and fictional characters as means of introducing that way of thinking and making it more concrete.

References

BACH, K. & HARNISH, R. M. (1979): *Linguistic Communication and Speech Acts*. Cambridge (Mass.): MIT Press.

BAKER, A. E. & HENGEVELD, K. (eds.) (2012): *Linguistics*. Malden: Wiley-Blackwell.

BIRNER, B. (2013): *Introduction to Pragmatics*. Malden: Wiley-Blackwell.

CHAUDHURI, S. (2010): *The Metaphysics of Text*. Cambridge: Cambridge University Press.

EFFINGHAM, N. (2013): *An Introduction to Ontology*. Cambridge: Polity.

ENFIELD, N. J. (2015): *The Utility of Meaning: What Words Mean and Why*. Oxford: Oxford University Press.

FROMKIN, V. A. (ed.) (2000): *Linguistics: An Introduction to Linguistic Theory*. Malden: Blackwell Publishing.

GASKIN, R. (2013): *Language, Truth, and Literature: A Defence of Literary Humanism*. Oxford: Oxford University Press.

GOODMAN, N. (1968): *Languages of Art*. London: Oxford University Press.

HALE, B. (1987): *Abstract Objects*. Oxford: Basil Blackwell.

HARRIS, R. (1998): *Introduction to Integrational Linguistics*. Oxford: Pergamon, Elsevier Science.

KROON, F. & VOLTOLINI, A. (2016): Fiction. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* (Winter 2016 Edition), available at: https://plato.stanford.edu/archives/win2016/entries/fiction/.

LIVINGSTON, P. (2016): History of the Ontology of Art. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* (Summer 2016 Edition), available at: http://plato.stanford.edu/archives/sum2016/entries/art-ontology-history/.

LYCAN, W. G. (2008): *Philosophy of Language: A Contemporary Introduction*. 2nd ed. London & New York: Routledge.

PETTERSSON, A. (2017): *The Idea of a Text and the Nature of Textual Meaning*. Amsterdam: John Benjamins.

REDDY, M. (1979/1993): The Conduit Metaphor: A Case of Frame Conflict in Our Language about Language. In: Ortony, A. (ed.): *Metaphor and Thought*. 2nd ed. Cambridge: Cambridge University Press, 164-201.

SEARLE, J. R. (1995): *The Construction of Social Reality*. New York: The Free Press.

Searle, J. R. (2010): *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.

Wetzel, L. (2014): Types and Tokens. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* (Spring 2014 Edition), available at: http://plato.stanford.edu/archives/spr2014/entries/types-tokens.

Organon F 24 (3) 2017: 384-404

Tichý and Fictional Names

DANIELA GLAVANIČOVÁ[1]

ABSTRACT: The paper examines two possible analyses of fictional names within Pavel Tichý's *Transparent Intensional Logic*. The first of them is the analysis actually proposed by Tichý in his (1988) book *The Foundations of Frege's Logic*. He analysed fictional names in terms of free variables. I will introduce, explain, and assess this analysis. Subsequently, I will explain Tichý's notion of *individual role* (office, thing-to-be). On the basis of this notion, I will outline and defend the second analysis of fictional names. This analysis is close to the approach known in the literature as *role realism* (the most prominent advocates of this position are Nicholas Wolterstorff, Gregory Currie, and Peter Lamarque).

KEYWORDS: fictional characters, fictional names, roles, Transparent Intensional Logic, variables.

1. Introduction

The semantic analysis of fictional names is one of the central topics in the (analytic) philosophy of fiction. This issue is closely related to the ontology and metaphysics of fictional characters, or more broadly, of fictional entities. By fictional names I will understand expressions that seem to be introduced in order to speak about fictional characters. Fictional names will

[1] Received: 3 April 2017 / Accepted: 27 June 2017

✉ Daniela Glavaničová
Department of Logic and Methodology of Sciences
Comenius University in Bratislava
Faculty of Arts, Gondova 2, 814 99 Bratislava, Slovak Republic
e-mail: dada.baudelaire@gmail.com

thus be expressions such as *Sherlock Holmes*, *Toru Watanabe*, *Thérèse Raquin*, *Jean-Sol Partre* etc. I will confine my attention to such *individual* fictional names, leaving aside fictional names of other sorts (e.g., names of fictional cities, events, bridges, schools and so on). Moreover, the focus will be on the literature, leaving thus aside movies, music, fine arts and so on.

Since Pavel Tichý and his *Transparent intensional logic* (TIL) will be central for this paper, let me briefly introduce him and his work. Pavel Tichý was a Czech logician and philosopher who developed TIL, a framework for the analysis of natural languages (but also for the analysis of languages in general). Tichý has laid the foundations of TIL in his *The Foundations of Frege's Logic* (see Tichý 1988). The goal of TIL was ambitious from the very beginning: Tichý claimed his system not only to correct the shortcomings of Frege's theory (hence the title) and of Russell's theory, but also to be the right medium for modelling our whole conceptual scheme (see Tichý 1988, ix). The proposed system is *hyperintensional*.[2] Tichý models meanings as structured procedures (*constructions*; see Tichý 1988, 56-65). Any well formed expression of TIL represents a construction (definitions of constructions can be understood as determining the syntax of the language in passing). Moreover, the system is fundamentally *functional* (even sets are understood primarily as characteristic functions) and these functions are *partial* (so a function can be undefined for some arguments). Every entity of TIL has its logical *type* (cf. Tichý 1988, 65-66). In sum, TIL is a *hyperintensional partial lambda calculus with types*. Most technical details of TIL are unimportant for this paper, so I will stop here for the time being.

Of course, one can ask: Why Tichý? Why this complicated system of TIL? Let me briefly motivate the project: First, Tichý sketched an interesting suggestion for an analysis of fictional names worth of further exploration. Indeed, this will be one of the aims of this paper. Second, Tichý extensively discussed so-called *individual roles* (offices, things-to-be) that have been also extensively discussed in the current TIL. One of the positions occurring in the literature concerned with the semantics of

[2] See Jespersen & Duží (2015) on the notion of hyperintensionality.

fictional names is the so-called *role realism*.[3] Though it is not an aim of this paper, I think that a comparison of the roles as understood by Tichý and the roles as understood by role realists might be fruitful. Third, a new cousin of the role realism can be formed using roles as understood by Tichý. I will outline and motivate such account, though, it shall be noted that I will not offer a comprehensive exposition and defence here. Fourth, fictional names are often thought to be non-referring (empty) expressions (see Braun 2005). TIL is partial, so emptiness is something that is not a problem for an advocate of TIL. Fifth, there are various issues concerning fictional names that call for a more sophisticated theory of meaning, or to say it frankly, that call for hyperintensionality. Again, I will stop here. I hope this suffices for the reader as an initial motivation.

A short note on methodology: I will treat the theories of fictional names as providing us with *semantic models*. For instance, when I suggest to model fictional characters in terms of individual roles, I am not thereby suggesting that we should *identify* fictional characters with individual roles.[4]

The structure of the paper is as follows: First, I recapitulate Tichý's reasons for denying certain sort of descriptivism of fictional names. I further introduce and explain his analysis of fictional names in terms of *free variables* (section 2). Subsequently, I assess his analysis (section 3). I then explain his account of individual roles (section 4). I use Tichý's roles to sketch and motivate an alternative account of fictional names (section 5). The account is close to descriptivism, but it does not amount to the sort of descriptivism that Tichý criticised. The final section 6 anticipates and discusses some objections that can be directed to the approach suggested in the section 5.

---

[3] The most important role realists are Nicholas Wolterstorff, Gregory Curie and Peter Lamarque; see Wolterstorff (1980), Curie (1990) and Lamarque (2008; 2010).

[4] Note that this approach allows me to abstract from some properties of the "real thing"; cf. Bielik, Kosterec & Zouhar (2014, 112) and Bielik (2012).

2. The first analysis: fictional names as free variables

The main aim of this section is to introduce and explain the analysis of fictional discourse suggested by Pavel Tichý. The analysis of fictional names was proposed in Tichý (1988); more precisely, it was merely sketched there because the whole account is explained on four pages (keep in mind, however, that the account presupposes the whole framework of TIL, as developed and defended in the book). Though some points may seem outright suspicious to the reader, I will postpone the assessment of the theory to the next section.

When we use fictional names, Tichý claims, we "seem to name particular entities and yet are unable to say which entities they are" (Tichý 1988, 261). Because of this, fictional reference is for him a case of so-called *unspecified reference*. How to account for fictional reference, then?

Tichý starts with showing us how an analysis should not look like: the section begins with the criticism of Fregean descriptivism of fictional names, thus paving the way for Tichý's own account.[5] He mentions several problems faced by a descriptivist:

1. *The choice problem*: It is not clear how to extract essential features that should be incorporated in the sense of a fictional name such as *Sherlock Holmes*. If one suggests that we should incorporate every single detail mentioned in any of these stories that would mean that one could not understand the first story before finishing the last. If one suggests that each story gives the name its own sense that would mean that one could not understand the first sentence of the story before finishing the last sentence.

Even if we suppose that we have somehow managed to extract such essential features from the Holmes stories as a whole (e.g., *Sherlock Holmes* is said to have the same sense as *the pipe-smoking detective*), there will still be many difficulties (see Tichý 1988, 262).

---

[5] Tichý attributes the view directly to Frege, and discusses it only in the context of Frege's work. However, the analysis of fictional discourse was not central for Frege (he was repeatedly trying to put it aside). Moreover, the position goes under the caption *descriptivism of fictional names* in the current literature. I will thus speak directly of descriptivism.

2. *The problem of different truth-conditions*: The sentence *Sherlock Holmes played the violin* can be true even if it does not hold that the pipe-smoking detective played the violin (i.e. if it was not true that there was *exactly one* pipe-smoking detective and that he played the violin). The truth of the sentence with relevant description is not necessary for the truth of the corresponding sentence with a fictional name. Moreover, even if there was exactly one pipe-smoking detective, it would not follow that *Sherlock Holmes* is his name and that Doyle was writing *about* him. This is the famous Kripke's objection to descriptivism of proper names.[6] In general, this means that the truth of the sentence with relevant description is not sufficient for the truth of the sentence with the fictional name in question. These sentences have different truth-conditions, so the relevant description can be hardly understood as a good analysis of the given fictional name.

3. *The attitude problem*: Tichý considers as a most troublesome difficulty for descriptivism the attitude the reader is supposed to take with respect to the thought expressed by the sentence containing fictional name – what sort of attitude it should be? Tichý claims that readers neither believe nor pretend to believe that the thought in question is true. However, readers nevertheless make some inferences concerning fictional names.

The above criticism motivates the introduction of Tichý's own account. This account can be summed up as follows:

1. Fictional names are analysed in terms of *free variables*.
2. Tichý's *two-dimensional theory of inference* (inspired by Frege and Gentzen) is used for the analysis.
3. The logical analysis is provided in terms of *TIL*.

Note that this is not the way how Tichý actually put forth his approach. Nevertheless, I think that the summary captures all crucial ingredients of his analysis.

---

[6] Kripke (1972, 157) expressed the objection as follows: "The mere discovery that there was indeed a detective with exploits like those of Sherlock Holmes would not show that Doyle was writing *about* this man."

Example: A simple example will provide an explanation of these three requirements. Consider the overused sentence *Sherlock Holmes is a detective*. In the first-order logic, the most standard analysis will be simply $D(s)$, where s is an individual constant and D is an individual predicate. I now gradually amend this analysis:[7]

1. $D(x)$
2. $D(x)$ / $D(x)$
3. $[^0D_w\, x]$ / $[^0D_w\, x]$[8]

Explanation:

1. The first amendment captures Tichý's requirement to analyse fictional names in terms of free variables. It can be understood as a case of *logical form* in the first order logic.

2. The second amendment incorporates Tichý's two-dimensional theory of inference. How to read this? One can find two readings in Tichý (1988, 263, 264), but it is not entirely clear which is the preferred one. First, one can read it as follows: for any individual, if the individual is a detective, then this individual is a detective (Tichý 1988, 263). Second, one can read it also in the following way: for any valuation, if the D(x) is true under this valuation, then D(x) is true under this valuation.

3. Finally, the third amendment comes with TIL analysis. I will not explain the whole foundations of TIL here (as mentioned above, one can find it in Tichý 1988, 56-66). However, I will explain this particular example. What is new in 3 is the construction $[^0D_w\, x]$; it is an abbreviation of $[[^0D\ w]\ x]$. This construction contains two free variables: w (for possible worlds) and x (for individuals). Not surprisingly, variables are also constructions. ^0D is a construction called *trivialisation*. It takes the object

---

[7] Again, proceeding in these 'steps' is a heuristic device that will help me to explain the analysis without explaining everything from the preceding more than 260 pages of the book.

[8] This construction is abbreviated. I explain it below. Also, Tichý would include variables for time-moments, so, more precisely, his analysis would be $[^0D_{wt}\, x]$ / $[^0D_{wt}\, x]$. However, since time-moments will be unimportant for the present discussion, I omit them to make the presentation simpler (and more comprehensible).

D and returns the very same object, the property of being a detective (a function from possible worlds to sets of individuals, where a set of individuals is understood as a function from individuals to truth-values). The brackets stand for the construction called *composition*. To put it simply, composition [$^0D\ w$] consists in executing the construction ^0D, thus obtaining the abovementioned mapping, executing *w* to obtain an argument of this mapping, and applying the mapping to the argument constructed by *w*, thus obtaining a function from individuals to truth-values (a set of individuals). Finally, the whole [[$^0D\ w$] x] consists in applying the latter mapping to the argument constructed by *x*, thus obtaining a truth-value.

However, up to this point, I omitted one important issue: "As a variable itself, a construction containing a variable may construct one entity relative to one valuation and another entity relative to another" (Tichý 1988, 62). The promiscuous nature of free variables makes it the case. Because of this, constructions containing free variables are, in some sense, not self-sufficient, not independent, incomplete – open. Open construction is a TIL-version of the logical form from the first order logic.[9]

Of course, it is also needed to say *which variables are free* according to Tichý; (the definition is to be found in Tichý 1988, 73-74). In general, *bound* variables are of two sorts: variables bound by trivialisation and variables bound by λ operator (again, a new notion: to put it simply, λ helps us to generate functions by abstracting from particular values of arguments). Free variables are, quite straightforwardly, variables that are not bound.

3. Assessing the first analysis

Let me now assess the analysis explained in the previous section. I, again, proceed along the three 'steps' introduced above. To begin with, I sum up some of its advantages. After that, I summarize some of its disadvantages. The list is not be exhaustive, but I mention the points that seem to be of the uttermost importance.

---

9 For more on the notion of logical form, see Duží & Materna (2005).

The first step safeguards the unspecified reference of fictional names (i.e., that a fictional name does not refer to a particular individual). Importantly, this requirement also assures that though fictional names do not refer to particular individuals, if the fiction contains (syntactically) different fictional names, they can be differentiated also at the level of semantics. This is so because they are analysed in terms of different free variables. For example, the sentence *Watson is a friend of Sherlock Holmes* can be (preliminarily) analysed as $F(y, x)$.

The importance of the above point stems from the fact that TIL involves so-called *objectual* theory of variables. This theory of variables assures that two variables are simply *two different objects*, they are not undifferentiated gaps.[10]

Because of this feature, the problem of co-identification, formulated by Stacie Friend, is avoided as well (see Friend 2014). This problem consists in the fact that when different people use the same fictional name, they seem to be talking about the same fictional character. However, antirealists about fictional characters maintain that there are no fictional characters – how to explain the seeming co-identification, if nothing is identified? Yet it was suggested that it is a problem for realists, too. Be it as it may, Tichý's theory does not lead to this problem: If one has two fictional names, one has two free variables.

As Raclavský repeatedly indicated, the analysis also nicely captures how we read or write fiction in the initial stages, when we have neither a complete list of properties ascribed to a fictional character nor a complete list of relations between various fictional characters (see Raclavský 2009; 2015). Note also that this feature makes the theory a good candidate for a realist position resistant to some of the Everett's worries concerning the initial stages of creating a work of fiction (cf. Everett 2013, sec. 7.4).

---

[10] Compare this with the Gappy Proposition Theory, for instance, the one formulated and defended in Braun (2005). David Braun struggled with the problem of differentiating between 'gaps'. The sentence *Holmes is a detective* seems to mean something different from what the sentence *Watson is a detective* means. Braun was trying to avoid descriptivism of fictional names, so these sentences are bound to express the same gappy proposition. His suggestion was to speak about different ways of believing. In this case, we can differentiate between 'Holmes-ish' and 'Watson-ish' ways. But how can one understood or formally model these different ways (while still avoiding descriptivism)?

Tichý motivates the second step as follows: "the reader of a fictional story is occasionally expected to draw inferences from what the text explicitly says" (Tichý 1988, 264). However, anything of the form A / A is a logical truth. Hence, it will be probably useful to invoke an analogy with a logical or mathematical system.[11] Tautologies of the system are logically true in one system, but may be untrue in another. Similarly, sentences of the fiction are true in one fiction, but may be untrue in another.[12]

The third step allows us to take on board all the advantages of TIL – hyperintensionality, partiality, types, and so on. I do not recapitulate all the advantages of TIL here, since the reader can easily see it just by scanning the work done in the current TIL.[13]

Let me now turn to disadvantages. First, it is not clear how the choice problem (which Tichý attributes to descriptivism) is avoided: Which constructions should be taken as inputs for reader's inferences?[14] And which logical, conceptual, or factual background assumptions are presupposed? Tichý does not seem to be counting with any factual assumptions, though these are needed for some obvious inferences (see Lewis 1978).

As regards the second step, a sentence with a fictional name amount under this analysis to a *trivial logical fact*. Tichý anticipates complaints (since works of art do not seem to be composed of logical facts) and defends his account in the following way: "But this is hardly an objection to it as a construal of Doyle's sentence. One does not turn to fiction to *learn* anything new. If one wants to learn, one had better reach for a book on a non-fiction shelf" (Tichý 1988, 263).

---

[11] Tichý seems to be approaching towards such an analogy (cf. Tichý 1988, 262-263), and it would facilitate the defence of his approach. But he does not venture far enough: maybe his objectual, realist stance on logic is behind this reticence (see the Preface in Tichý 1988, or the first chapter thereof for evidence). Yet by venturing too far I am *not* inviting the reader to abduce my views on logic.

[12] Recall the 'In fiction F' operator from Lewis' (1978) crucial paper.

[13] E.g., see the contents of Duží, Jespersen & Materna (2010) or Raclavský (2009).

[14] One of the reviewers did not agree that the choice problem is a problem for Tichý's analysis, suggesting that we should take as an input simply those that can be seen as meanings of the sentences of the respective stories. But a role realist can employ the very same strategy. The choice problem cannot help us in deciding between Tichý's original approach and a version of role realism based on TIL (see Section 5).

I disagree. We learn something even when we read novels – we learn about the content of the relevant novel (e.g., a literature student can use this knowledge afterwards – during tests, exams, while writing a paper, for a research in the field of literary criticism, etc.).[15]

While the sentence *Sherlock Holmes was a pipe-smoking detective entails that Sherlock Holmes was a pipe-smoking detective* is not informative (e.g., for a reader, for a literary critic), the sentence *Sherlock Holmes was a pipe-smoking detective* is informative. When we find such sentence in the Holmes stories, we indeed learn something about the content of the work, about the character of Sherlock Holmes, etc. Therefore, while I agree that *Sherlock Holmes was a pipe-smoking detective entails that Sherlock Holmes was a pipe-smoking detective* "does not represent an interesting piece of knowledge", I do not agree that it is not an objection to Tichý's account. Indeed, it is an objection to his account as a construal of Doyle's sentence *Sherlock Holmes was a pipe-smoking detective*, since this sentence is in some sense informative – in a sense in which Tichý's analysis of this sentence is not.

Furthermore, novels often contain some factual claims, personal view and attitudes of the author, some moral or aesthetic judgements etc. (consider any deeply personal novel, for instance, Bukowski's *Ham on Rye*; if the movies were included, Woody Allen would give us a plenitude of good examples). Many novels are also partly based on historical events and are describing real people (for example, consider the novel about Gödel and his wife, *The Goddess of Small Victories*, written by Y. Grannec).

Another serious trouble was suggested by one of the anonymous reviewers. Variables are always proper (a variable is evaluated under every valuation). Therefore, Tichý's analysis counts as a realistic one. How to read nonexistence claims then? It is not easy to see how this account could possibly make sense of sentences such as "Sherlock Holmes does not exist".

Also, whilst free variables may be plausible candidates for explaining the initial stages of writing / engaging with some fiction, they do not seem

---

[15] Sometimes we even learn new factual information, since writers often use real cities, real events, and write true things about them. Yet the sort of information I had in mind is more closely related to the notion of *analytic information* as employed in Duží (2010).

as plausible candidates for things we love, hate, admire, etc., when we take some emotional attitudes towards works of fiction.

Jiří Raclavský has mentioned another worry in one of his papers (see Raclavský 2007), namely that it is not always clear whether the given name is fictional or not. For instance, how can one know whether some name occurring in a certain novel is a proper name of a real person (not described in the novel with a historic accuracy) or a fictional names of a fictional character (based or modelled on some real person)? Yet this problem is probably not specific to Tichý's account.

These were the most serious advantages and disadvantages of Tichý's analysis. Let me now proceed towards an alternative account, an analysis in terms of Tichý's notion of *individual roles*.

4. Tichý's individual roles

In section 2, I have sketched Tichý's reasons for denying certain sort of descriptivism of fictional names. However, definite descriptions that a descriptivist uses to analyse fictional names are, indeed, very similar to unspecified reference. Consider a case when one uses a definite description, e.g., *the president of France*. This description does not explicitly mention any specific individual who occupies the French presidential office. Furthermore, there are cases when one uses a definite description and is not, for good reasons, able to tell the reference: e.g., when seeking who the murderer is. Moreover, there are cases when a definite description necessarily does not have a referent – consider the famous Quine's example of *the round square cupola on Berkeley College*.

Tichý devoted much attention to the analysis of definite descriptions. He analysed them in terms of his notion of an *individual role* (office, thing-to-be). The plan of this section is to explain his account of these individual roles. Tichý's most extensive exposition of this notion is to be found in his *Individuals and Their Roles* (see Tichý 2004a).[16] Intuitively,

[16] The first version of the text appeared in his unpublished manuscript *The Introduction to Intensional Logic* from 1976 (in English). It was reworked and published in 1987 (in German), its Slovak translation was published in 1994 and finally, the English translation was published as Tichý (2004a). I quote from the (2004a) version.

an individual office is something an individual can be: for instance, the description *the president of France* changes its extension as time goes by. *The president of France* denotes something that an individual can be – an office, a role.

There are several important features of individual offices that need to be mentioned here:

1. Roles have *requisites*: A requisite is "what it takes for a material object to hold that office" (Tichý 2004a, 717). For instance, "the property of being a king is a requisite of the office of King of France, such that every occupant must have the relevant property" (Duží, Jespersen & Materna 2010, 128).[17]

2. Roles have *properties*: the President of France is an eligible office; the President of France exists (is occupied), the King of France does not exist (is not occupied).

3. Roles are *abstract* entities: they are not to be confused with individuals actually occupying them (if there are any).

4. Roles can be *occupied* by different individuals at different possible worlds and times.

5. Roles may be *indeterminate*: "A thing-to-be … may be largely indeterminate as to physical properties … A thing-to-be, i.e., an office occupiable by material objects is not the sort of thing which might conceivably have a weight, precise or vague" (Tichý 2004a, 716-717).

6. Roles can be *unoccupied*: "Given a thing-to-be, it may happen that nothing is it; it may happen that there is no such thing… nothing even remotely similar can be said of a material body" (Tichý 2004a, 716).

7. Roles are objects on its own right, so one can *quantify* over them.

8. Roles are given, or determined by certain *constructions*. If one invokes the Fregean *sense* and *reference* distinction,[18] one can say that these

---

[17] Surely, this is just an intuitive explanation, not a precise definition. See section 6 where I discuss this notion more extensively.

[18] Fregean, not Frege's. Frege used these terms differently. He did not employ constructions in the first place. Also, the reference was for him usually an extensional entity. Here the constructions represent sense and intensions reference. What is borrowed

constructions are senses of definite descriptions and roles are their reference.

9. One can also have various *attitudes* to these roles, or to constructions determining these roles; moreover, one can *talk* about these roles (or constructions denoting them) and *compare* various roles, talk about their properties and requisites, etc.

Interestingly enough, Tichý mentions the fictional name *Sherlock Holmes* in the discussed paper. He considers the sentence *Sherlock Holmes does not exist* and asks: "whom do I deny existence of? Sherlock Holmes, i.e., a person? Hardly. As is well known, none of the persons in the world is Sherlock Holmes. It thus appears that … I deny existence of nothing at all" (Tichý 2004a, 720). Afterwards, he writes that the same goes for his new Rolls-Royce when saying *My new Rolls-Royce does not exist* (the latter was analysed in terms of roles). However, his new Rolls-Royce is a role – what about Sherlock Holmes? Moreover, though Duží, Jespersen & Materna (2010) agree with the first analysis proposed by Tichý (fictional names as free variables), they analyse the name *Santa Claus* in terms of roles (see Duží, Jespersen & Materna 2010, 90-92). Yet the name *Santa Claus* seems to be similar to the name *Sherlock Holmes*. In this paragraph, I was just trying to show that the possibility to analyse fictional names in terms of roles is not entirely extraneous to TIL. Let me now explore this possibility.

5. The second analysis: fictional names as individual roles

A closer look at the characteristics of roles 1-9 from the preceding section reveals that it may be promising to use this account of roles for the analysis of fictional names. Indeed, this is something what I will do in this section. Let me start with the following proposal:

> *Fictional names should be analysed in terms of individual roles (as characterised in the preceding section). However, these individual*

---

from Frege is just the metaphorical understanding of the sense as a way of giving the reference.

roles are necessarily non-occupied (empty). As it is with the other expressions, we can pronounce a fictional name to speak about its sense (a construction), about its reference (a role), or about its extension (which, as it should be clear, does not exist).

I follow the characteristics 1-9 and explain their usefulness for the analysis:

1. Requisites are helpful in analysing the properties ascribed to fictional characters by the authors, such as *Sherlock Holmes is a pipe-smoking detective*: If there was someone fulfilling the role of Sherlock Holmes, this entity would be a pipe-smoking detective. It is literally true that a role of Sherlock Holmes has this requisite: it does not matter that the office of Sherlock Holmes is not occupied. A similar strategy can be used for ascribing many other properties as requisites for being that fictional character (weight, height, age…).[19]

2. Properties are useful for the analysis of sentences such as *Sherlock Holmes is a fictional entity* or (the character) *Sherlock Holmes is famous*. A quite tough sentence *Sherlock Holmes was created by Conan Doyle* can be analysed thanks to properties of roles as well. Intuitively, the creation of Sherlock Holmes consists simply in the fact that Doyle picked an expression or introduced a new one for his character and chose some initial properties. This can be captured as a role property of being firstly described by Doyle. This is, of course, a sort of creation in the spirit of role realism (cf. Wolterstorff 1980; see also the distinction between *characters per se*

---

[19] I won't go into details here, but this is not the only reading of such sentences that we can get. If we employ *de dicto* and *de re* distinction, as I surely would do in a more comprehensive exposition of my views, we can distinguish:

(i) Sherlock Holmes (de dicto) has a requisite (= first order property)…

(ii) Sherlock Holmes (de dicto) has a (second order) property…

(iii) Sherlock Holmes (de re) has a (first order) property…

The sentences of the form (i) and (ii) can be literally true, if the role of Sherlock Holmes has the respective requisite (e.g., being a pipe-smoking detective) or the respective property (e.g., being a famous fictional character or being non-occupied), whilst no sentence of the form (iii) will be true, since there is no "res", no occupant of the Sherlock Holmes office.

and *fictional characters* in Lamarque 2010, 201), which would not satisfy all theoreticians.[20]

3. The fact that roles are abstract entities makes them intersubjective. From the ontological point of view, it is important that this approach does not postulate any *new* entities (roles are needed also for the analysis of non-fictional discourse). In addition, roles are no *queer* entities; they are baked from the very same ingredients as properties or relations. It can be seen that this approach adopts a middle way between realism and irrealism (as understood in Sainsbury 2009). It is not a genuine realist approach, since it does not claim that there is a 'robust' Sherlock Holmes (occupying the individual office of Sherlock Holmes); and it is not a genuine irrealist approach, since individual offices are abstract entities (but are no more 'exotic' than numbers and sets).

4. The fact that roles can be occupied by different individuals is not important here, since these roles are necessarily empty.[21]

5. Indeterminacy is crucial, since fictional characters are not described completely.

6. The fact that roles can be empty is crucial too: since there are no entities named by fictional names, there is no *real* Sherlock Holmes. This is important for the non-existence claims, such as the claim that Holmes does not exist.[22]

7. The possibility of quantification over fictional characters is important for the analysis of sentences such as *Many characters occurring in Murakami's novels are sad.* The same goes for sentences comparing different fictions.

---

[20] For instance, Amie Thomasson, an abstract artifactualist, criticised this way of approaching the authorial creation; see Thomasson (2009).

[21] Currie (1990) would not agree with this point (see section 6).

[22] One of the reviewers claimed that there could be only one trivialization of necessary empty role in TIL, so every fictional name would have the same meaning. It depends on the identity criteria of constructions in general and constructions of necessarily empty roles in particular. I do not see any obstacle in differentiating between different (constructions of) empty roles on the basis of their requisites. E.g., the meaning of *Holmes* is different from the meaning of *Watson*, because the requisites for being Holmes differ from the requisites for being Watson.

8. Constructions are important here for many reasons, but the most crucial one is this: Since the reference of a fictional name is a necessarily empty role, one needs something to differentiate between various empty roles. This falls within the competence of constructions. They can embody the 'Holmes-ish' and 'Watson-ish' ways of believing, in an exact way.

9. We sometimes hate, love, pity, envy or simply think about fictional characters. All these are attitudes. It is thus desirable that this approach can acknowledge this issue. Who do girls love, when they love, for instance, Mr. Darcy? Surely not some real, full-blooded entity, since (unfortunately) there is no such entity. They love a fictional character. All they have is a fictional name, a bunch of words describing this fictional character, the respective bunch of meanings, and probably also corresponding imagination of the fictional character. Yet my imagination of Mr. Darcy is surely not identical to him, since it is subjective, and not intersubjective. An intersubjective entity is needed if we want to make sense of sentences such as "Many girls love Mr. Darcy" or "There is a fictional character that many girls love, namely, Mr. Darcy."

6. Possible objections

The first difficulty, and perhaps the most serious one, is the already explained Kripke's objection (see footnote 5 of this paper). I am not willing to bite the bullet (as Currie partly does in his 1990, 180) and say that I *can* be Sherlock Holmes. Nor I am willing to say that there can be Sherlock Holmes, since as I understand the expression "Sherlock Holmes", if it stands for anything, it stands for *that* fictional character from Doyle stories. As Kripke maintained, even if someone very similar happened to exist, it would still not suffice for saying that Doyle was writing *about* this man (similarly, if a winged horse-like creature appeared, it would not suffice for saying that Pegasus is real after all). If someone asked us where the grave of Sherlock Holmes is, we would see it either as a joke or as a terrible misunderstanding. Under the supposition that Sherlock Holmes cannot exist Kripke's objection is easily avoided as soon as there is something *blocking the occupation* in the case of the actual existence of an individual satisfying all the properties explicitly ascribed to the fictional character by the author of the fiction in question. As I formulated the analysis, fictional

characters are necessarily empty roles. However, a precise formulation of the role property blocking the occupation is still needed.

Whether we accept the supposition that the role of Sherlock Holmes is necessarily empty or not, I maintain that this issue needs to be resolved before the actual analysis can start. We need to know whether some term is a fictional name (a name of a purely fictional character) or a genuine individual proper name (a name of a real person). Since fictional characters are not real individuals, this matter is of semantic significance (under the supposition of Millianism). And the resolution may be sensitive to the purpose of the subsequent analysis (see below the answer to the choice problem).

A related objection pertains to the notion of requisites. As Miloš Kosterec has reminded me several times, if fictional characters are modelled in terms of necessarily empty roles, requisites cannot be defined as Duží, Jespersen and Materna suggest in their (2010, 361-362), for every property would be a requisite of such roles. This is a serious worry, but there are some options how to deal with it. One option is to take Currie's route (mentioned above) and admit that it is possible that there was real Sherlock Holmes (indeed, one of the reviewers was suggesting precisely this). Another option is to use definitions of requisites from the above book, but change the material implication for some other sort of implication. A further option is to treat the notion of requisites as primitive. Finally, my preferred option is to define the requisites in terms of the content of the respective work of fiction. Note, however, that there are principal reasons why the essential properties cannot be defined once and for all: identity of fictional characters is interest relative, and so is the extent of their essential properties (again, see below). Also, note that the issue of determining requisites would then be closely connected to another task, that of determining truth in fiction (see Lewis 1978).

Another worry is the choice problem. I am sympathetic to the reply along the lines of Lamarque (2010, 200) who argues that "fictional characters are initiated types, grounded in acts of story-telling, i.e. fictional narratives, although not essentially bound to anyone, even if tied to a reasonably determinate historico-cultural context. Their identity is interest-relative depending on demands placed on their identity conditions, which in turn determine the extent of their essential properties." If one aims to compare Sherlock Holmes from Doyle stories to his counterpart from modern

Holmes series, it may be useful to treat them as different fictional characters. If one aims to compare Holmes to other fictional detectives, it may be useful to treat old and modern Sherlock Holmes as one character. If one judges Grannec's novel as a biography about Kurt Gödel, no differentiation between the real and the fictional Gödel is made, and if something false is said about him, it may be properly said that Grannec made a mistake. On the other hand, if we read the piece just as a novel, the real Gödel is distinguished from the fictional one, and it makes no sense to criticize Grannec for making the fictional Kurt different from the real one. I think that this sort of sensitivity is widespread, it exists in reality and needs to be acknowledged by theories of fictional characters.

Furthermore, by opting for a role-based analysis of fictional names, one is either denying the claim that all *proper names* are *directly referential* expressions, or one is denying that all *fictional names are proper names*.

I prefer the latter. For most, however, it is not even an option to question the status of fictional names as proper names (cf. Friend 2014). On the other hand, some would opt precisely for this suggestion: "We should not start by assuming that fictional names are genuine proper names. We need to know more about fictional names and proper names before we can decide whether they are" (Currie 1990, 128).

Let me now consider these two options. Direct reference theory of proper names says that proper names refer to individuals irrespective of the properties exemplified by these individuals. It is a plausible, intuitively appealing view, enjoying the great popularity. Woody Allen would have been Woody Allen even if he had not directed *Annie Hall* and even if he had not been a director at all. His name is due to naming conventions, not the facts of his life. On the other hand, it seems that fictional names are not directly referential. Follow my simple line of reasoning: to begin with, take an arbitrary fictional name (say, let us pick again the overused *Sherlock Holmes*). First, if *Sherlock Holmes* fails to refer, it is not a directly referential expression (since it is not a *referential* expression). Second, if this expression refers (in the sense that it has an extension), this extension must be a tall man, a detective, must live at the Baker Street, and must have properties explicitly ascribed to him. So, in this case, it is again not a directly referential expression (it refers, but not *directly*). Be it as it may, this expression is not a directly referential expression.

The above thoughts motivate the following: no fictional name is directly referential, but all genuine proper names are directly referential. This entails that fictional names are not proper names, so the unity is lost.

Is that a problem? I don't think so. Yes, the syntactic form of fictional names suggests that they belong to the family of proper names. But there is also a considerable dissimilarity between the way fictional names function in the language and the way genuine proper names function in the language (even ignoring the above claim that fictional names are not directly referential). For instance, if one wants to name a child, one has an individual *given in advance*, and one wants to name this given individual. This is not the case in fiction. The author chooses an expression that will serve as a fictional name and some initial properties that will be ascribed to the character – though there may be some mental idea of this fictional character given in advance, but this idea is not identical to this character – the character is not yet born. Moreover, there are considerable differences in the identity conditions. While no individual is identical to another individual (numerical identity), the case of fictional characters is not that strict. Recall the quote from Lamarque (2010, 200) saying that identity of fictional characters is interest relative. Nothing remotely similar can be said of individuals and their proper names.

Acknowledgments

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0149-12. The paper is based on my talk presented at the conference *Philosophy of Language (I): Semantics of Fictional Discourse*: I am indebted to all who participated in the fruitful discussion. I am also grateful to the research group of the Institute of Philosophy of the Slovak Academy of Sciences for letting me join their seminars. Without them, this paper would not come into existence. I am grateful to Miloš Kosterec, Róbert Maco, Marián Zouhar and Zsófia Zvolenszky for discussions. Thanks go also to reviewers for numerous useful suggestions and objections.

References

BIELIK, L. (2012): How to Assess Theories of Meaning? Some Notes on the Methodology of Semantics. *Organon F* 19, No. 3, 325-337.

BIELIK, L., KOSTEREC, M. & ZOUHAR, M. (2014): Model metódy (1): Metóda a problém [A Model of Method (1): Method and Problem]. *Filozofia* 69, No. 2, 105-118.

BRAUN, D. (2005): Empty Names, Fictional Names, Mythical Names. *Noûs* 39, No. 4, 596-631.

CURRIE, G. (1990): *The Nature of Fiction*. Cambridge: Cambridge University Press.

DUŽÍ, M. & MATERNA, P. (2005): Logical Form. In: Sica, G. (ed.): *Essays on the Foundations of Mathematics and Logic 1. Advanced Studies in Mathematics and Logic*. Monza: Polimetrica International Scientific Publisher, 115-154.

DUŽÍ, M. (2010): The Paradox of Inference and the Non-Triviality of Analytic Information. *Journal of Philosophical Logic* 39, No. 5, 473-510.

DUŽÍ, M., JESPERSEN, B. & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic*. Berlin: Springer.

EVERETT, A. (2013): *The Nonexistent*. Oxford: Oxford University Press.

FRIEND, S. (2014): Notions of Nothing. In: García-Carpintero, M. & Martí, G. (eds.): *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 307-332.

JESPERSEN, B. & DUŽÍ, M. (2015): Introduction to the Special Issue on Hyperintensionality. *Synthese* 192, No. 3, 525-534.

KRIPKE, S. A. (1972): *Naming and Necessity*. Springer.

LAMARQUE, P. (2008): *The Philosophy of Literature*. Blackwell.

LAMARQUE, P. (2010): *Work and Object: Explorations in the Metaphysics of Art*. Oxford University Press.

LEWIS, D. (1978): Truth in Fiction. *American Philosophical Quarterly* 15, No. 1, 37-46.

RACLAVSKÝ, J. (2007): Individua a individuové úřady Pavla Tichého [Individuals and Individual Roles of Pavel Tichý]. *ProFil* 8, No. 1., Available at: http://www.phil.muni.cz/~raclavsky/texty/ticheho_urady.pdf

RACLAVSKÝ, J. (2009): *Jména a deskripce: logicko-sémantická zkoumání* [*Names and Descriptions: Logico-Semantic Investigations*]. Olomouc: Nakladatelství Olomouc.

RACLAVSKÝ, J. (2011): Sémantika jmen ve fikci: obhajoba a rozvinutí Tichého koncepce [The Semantics of Names in Fiction: A Defense and Elaboration of Tichý's Approach]. *Organon F* 18, No. 1, 72-83.

SAINSBURY, R. M. (2009): *Fiction and Fictionalism.* London & New York: Routledge.

THOMASSON, A. (2009): Fictional Entities. In: Kim, J., Sosa, E. & Rosenkrantz, G. (eds.): *A Companion to Metaphysics*. Oxford: Blackwell, 10-18.

TICHÝ, P. (1988): *The Foundations of Frege's Logic*. Berlin & New York: de Gruyter.

TICHÝ, P. (2004): *Collected Papers in Logic and Philosophy*. Svoboda V., Jespersen, B. & Cheyne, C. (eds.), Prague: Filosofia, Czech Academy of Sciences; Dunedin: University of Otago Press.

TICHÝ, P. (2004a): Individuals and Their Roles. In: Svoboda V., Jespersen, B. & Cheyne, C. (eds.), Prague: Filosofia, Czech Academy of Sciences; Dunedin: University of Otago Press, 711-748.

WOLTERSTORFF, N. (1980): *Works and Worlds of Art*. Oxford: Clarendon Press.

Organon F 24 (3) 2017: 405-426

Using "not tasty" at the Dinner Table

ALEX DAVIES[1]

ABSTRACT: John MacFarlane argues against objectivism about "tasty"/"not tasty" in the following way. If objectivism were true then, given that speakers use "tasty"/"not tasty" in accordance with a rule, TP, speakers would be using an evidently unreliable method to form judgements and make claims about what is tasty. Since this is implausible, objectivism must be false. In this paper, I describe a context in which speakers deviate from TP. I argue that MacFarlane's argument against objectivism fails when applied to uses of "not tasty" within this context. So objectivism about "not tasty" is still a viable position within this context.

KEYWORDS: MacFarlane – objectivism – predicates of personal taste – relativism – sociology.

1. MacFarlane's TP

Although he acknowledges some exceptions (which we'll turn to shortly), MacFarlane (2014, 4) claims that we employ the following method to decide when to call things "tasty"/"not tasty":

[1] Received: 9 June 2016 / Accepted: 10 November 2016

✉ Alex Davies
Department of Philosophy
University of Tartu
Jakobi 2, 51014 Tartu, Estonia
e-mail: alex.stewart.davies@gmail.com

TP If you know first-hand how something tastes, call it "tasty" just in case its flavour is pleasing to you, and "not tasty" just in case its flavour is not pleasing to you.

Though TP takes the form of a prescription, because it is intended as a description of our method for using "tasty"/"not tasty" it should be understood as a descriptive proposal.[2] According to MacFarlane (2014, 141), TP delineates under what conditions we take the use of "tasty"/"not tasty" to be *warranted*. MacFarlane (2015, 21, 141) takes TP – so understood – to be a datum that any satisfactory theory of the meaning of "tasty"/"not tasty" must respect.

It is on this basis that MacFarlane (2014, 2-7) justifies rejection of objectivism about "tasty"/"not tasty". MacFarlane defines objectivism as the conjunction of (a) and (b):

Objectivism

(a) "tasty" is true of some things, false of others, and

(b) whether "tasty" is true or false of a thing, on a particular occasion of use, does not depend on the idiosyncratic tastes of the speaker, assessor, or anyone else.

As MacFarlane (2014, 2-3) intends the thesis, contextual variation in the extension of "tasty"/"not tasty" is consistent with objectivism. He allows, for example, that the extension of "tasty"/"not tasty" shifts in the same way that the extensions of other gradable adjectives (e.g. "red", "tall", "flat" etc.) shift with context. What objectivism disallows is variation in extension with the 'idiosyncratic tastes of the speaker, assessor, or anyone else.' According to objectivism, just as the fact that A ('or anyone else') has a visual experience of redness when looking at an object doesn't make A's claim that the object is red true, so too, the fact that A ('or anyone else')

[2] In this paper when I speak of "tasty"/"not tasty" I am referring to both positive and negative forms of the predicate. When I speak of "tasty" I am referring to only the positive form of the predicate. When I speak of "not tasty" I am referring to only the negative form of the predicate. Although talk of "tasty"/"not tasty" may be cumbersome, it helps me to be unambiguous about which forms of the predicate I intend to speak of at different parts of this paper.

has a pleasing gustatory experience when eating a particular food doesn't make A's claim that the food is tasty true.

MacFarlane rejects objectivism because he thinks it has an implausible implication when it is combined with two further assumptions. These are: firstly, that there is a lot of divergence across speakers in what foods they find pleasing; and secondly, that TP accurately describes the conditions under which we think the application of "tasty"/"not tasty" to a food is warranted. Given these two assumptions, speakers will make different judgements and claims about the tastiness of the same food. But given objectivism, a large number of such speakers must be making a mistake. Therefore using what one finds pleasing and displeasing as a way of reaching a verdict on whether some food is tasty must be a highly unreliable way of making a true judgement or claim about whether some food is tasty. MacFarlane thinks it implausible that we would use an evidently unreliable method to make taste judgements and claims and for this reason he rejects objectivism.

MacFarlane acknowledges that there are uses of "tasty"/"not tasty" that diverge from TP.[3] Following Lasersohn (2005, 670), MacFarlane (2014, 155-156) distinguishes between exo-centric uses of "tasty"/"not tasty" and auto-centric uses of "tasty"/"not tasty". With auto-centric uses of "tasty"/"not tasty", one uses the predicate in accordance with what one oneself finds pleasing and not pleasing. With exo-centric uses one uses "tasty"/"not tasty" in accordance with what someone else finds pleasing and not pleasing. Exo-centric uses of "tasty"/"not tasty" obviously do not conform to TP: they are exceptions to that rule. A dog-owner who says, "That's tasty", speaking of dog food, need not be going wrong in saying this, even though she herself finds the taste of the dog food displeasing or even if she does not know how dog food tastes: she can speak in accordance with what her dog finds pleasing. However, MacFarlane's objection against objectivism is that use of "tasty"/"not tasty" in accordance with *anyone's* idiosyncratic gustatory likes and dislikes is not a reliable way to form true judgements and claims about what is tasty, if objectivism is true. So, given objectivism, one is just as likely to be unreliable in

[3] Note that when someone doesn't use "tasty"/"not tasty" in accordance with what tastes she finds pleasing and displeasing because she has lost track of how something tastes, this is not an exception to TP.

one's pursuit of truth if one uses "tasty"/"not tasty" exo-centrically as one is if one uses it auto-centrically. This point applies just as much to exo-centric uses of "tasty"/"not tasty" that defer to what pleases or displeases a contextually salient group as it does to exo-centric uses that defer to a contextually salient individual.

In this paper, I will provide a partial defence of objectivism by identifying a context in which we consider application of "not tasty" to an item unwarranted even though we know first-hand how it tastes and that its flavour is displeasing to us, unless what we find displeasing tracks some feature of the item besides our own displeasure at its taste.[4] In this context, our use of "not tasty" deviates from TP in a way that makes trouble for MacFarlane's argument against objectivism. I describe this context in section 3. In section 4, I describe a more accurate alternative to TP. In section 5, I show how the alternative can accommodate MacFarlane's reasons for accepting TP. In section 6, I argue that the more accurate alternative to TP shows that MacFarlane's argument against objectivism fails when it is applied to the use of "not tasty" in the context described in section 3. I begin in section 2 by registering MacFarlane's reasons for accepting TP.

2. MacFarlane's defence of TP

MacFarlane provides three reasons to believe that TP describes the conditions under which we count application of "tasty"/"not tasty" as warranted. Firstly, MacFarlane takes this to be self-evident. He asserts:

[4] For an alternative response to MacFarlane's argument against objectivism see Hirvonen (2016). Hirvonen argues that "tasty"/"not tasty" is used in accordance with objectivism in most contexts. But, against MacFarlane's assumption to the contrary, she argues that it is not implausible that we are systematically mistaken in our taste judgements and claims. I will not be defending such an error theory. Rather, I will be defending an objectivism that is restricted to contexts in which we appear to be making an earnest attempt to track certain objective properties of food which food plausibly does possess some of the time and which we plausibly are able to detect some of the time. This restricted objectivism is in no need of an error theory.

> To a pretty good first approximation, we call a food "tasty" when we find its taste pleasing, and "not tasty" when we do not. (MacFarlane 2014, 3)

He offers no evidence in support of this assertion. Secondly, MacFarlane (2014, 4) claims that the following sentences seem 'odd':

(1) I'm not sure whether espresso is tasty, but I hate how it tastes.

(2) I've never been able to stand the taste of durian. Might it be tasty?

(3) I love orange juice and hate tomato juice. But who knows? Perhaps tomato juice is tastier.

In each case, there's a dis-connect between whether the speaker likes the taste of an item and whether the speaker applies the predicate "tasty"/"not tasty" to that item. This dis-connect seems responsible for the oddness of (1)–(3). This suggests that ordinarily there should be no such dis-connect: you should use "tasty" when something is pleasing to you and you should use "not tasty" when it is not pleasing to you.

Thirdly, MacFarlane thinks that our use of "tasty"/"not tasty" has a certain characteristic purpose. If we did use "tasty"/"not tasty" in such a way that (1)-(3) were not odd then, he thinks, the predicate would lose this purpose:

> We classify things as tasty or not tasty in order to help guide our gustatory deliberations. We eat things we regard as tasty because we expect them to taste good to us. Conversely, we may avoid eating things we don't know are tasty, because they might taste bad to us. But these explanations presuppose something like TP. (MacFarlane 2014, 4)

MacFarlane offers no evidence in favour of this third claim about the purposes to which people tend to put "tasty"/"not tasty". As with the first, he takes it as self-evident.

The datum underlying MacFarlane's second reason for accepting TP is compatible with the alternative that I am going to propose: the alternative doesn't imply that (1)–(3) don't sound odd. However, I will be disputing MacFarlane's first and third reasons for accepting TP. I will discuss MacFarlane's reasons for accepting TP again in section 5.

In light of the possibility of exo-centric uses, one should already be suspicious of MacFarlane's first and third reasons for accepting TP. Simple recognition of exo-centric uses of "tasty"/"not tasty" should be enough to cast doubt on the first reason. Examination of other purposes to which predicates of personal taste can be put – which are listed by Lasersohn (2005, 671-673) – should be enough to cast doubt on the third reason. But as has already been mentioned, exo-centric uses nonetheless do not cast doubt on MacFarlane's argument against objectivism. The exception to TP that I describe in the next section, however, will do so.

3. Negative taste assessments at the dinner table

Let's call an utterance of the sentence "this is not tasty" as used to speak of some food, *a negative taste assessment*. In this section I describe a context in which, even if a speaker knows first-hand how something tastes and that the taste displeases her, we consider a negative taste assessment unwarranted insofar as the assessment doesn't track a feature of the food besides its being not liked by the speaker or anyone else.

Around some dinner tables, there is an expectation that you eat the food on your plate, even if the taste of the food is displeasing to you.[5] In her study of Australian family dinner times, Greishaber (1997, 658) notes that children were subject to the following rules (amongst others):

- All food is to be eaten.
- Some vegetables are to be eaten for the evening meal.
- A portion of vegetables is to be eaten at the evening meal.

[5] As should be familiar to anyone who remembers discovering, as a child, that not everyone does dinner time in quite the same way as one's own family does dinner time, different dinner tables foster different kinds of context. For example, Ochs et al. (1996) note that in the Italian households they observed, children were actively encouraged to perform and develop their personalities by choosing what to eat and what not to eat, whereas their American counterparts were chastised for failing to eat anything if their only reason for doing so was that they just didn't like it. In this paper I focus upon a particular kind of dinner table context. I do not mean to suggest that all dinner table contexts are like the one we discuss.

- What is on the plate must be eaten.
- What is requested must be eaten. (Also noted by Ochs, Pontecorvo & Fasulo 1996, 17)

These rules are not contingent upon the child's finding the food pleasing. Often these rules apply as much to adults as they do to children. Enforcement of the rules is no doubt different for children as compared with adults. A child may be forced to eat any leftover food the following day or she may be forced to forfeit dessert if she violates the rules, whereas an adult probably will not. Despite differences in the means of enforcement, often the same rule is nonetheless in place for adults: eat what's on your plate, even if you find it displeasing. If you go to dinner at a friend's house or at your mother's house, and labour has been spent in financing and cooking the food, then, even if you don't find the food pleasing, you will feel an intense pressure to eat the food because you know you're breaking a rule and you know that you will likely face emotional and social sanctions for doing so.

Now, if you are in a context where you are under an obligation to eat food even if you find it not pleasing, then you may seek to excuse yourself from eating the food so that you don't have to do something you find not pleasing. What kinds of excuse are available? You can excuse yourself from eating the food in the following two ways. Firstly, you can try to excuse your non-eating by drawing attention to your idiosyncratic food preferences. The food just happens not to appeal to you but this doesn't reflect any defect in the food. Secondly, you can try to excuse your non-eating by drawing attention to some respect in which the food or its preparation deviates from how it ought to be or normally is and where the measure of how it ought to be, or normally is, is not simply a question of how pleasing the food is to you or anyone else. Here are some examples of such deviancy. A pastry is soggy and moist but pastries are supposed to be dry. The beer taps in the student bar haven't been cleaned recently but they should be cleaned regularly. Dad used Aunt Bessie's gravy but mum usually uses Paxo when she makes the roast. In each case, there is a way the food or its preparation ought to be or normally is but is not in fact. In each case, the deviation from a norm is not identifiable with whether anyone finds the food pleasing or not. Whether dad *did* use a different brand of gravy from the normal one, whether this pastry *is* soggy while standard pastries are

dry, and whether the beer taps *have* been cleaned recently and should have been, are states of affairs that are what they are regardless of who finds the gravy, pastry or beer tasty. To excuse oneself from consuming something by drawing attention to some such deviance in the food or drink from some such norm is to provide a different kind of explanation of one's non-eating than to excuse oneself by simply drawing attention to whether the food is gustatorily pleasing.

The contrast between the two possible sites of explanation for not eating food is clearly drawn in the following remark from the Picky Eating Adult Support Group:

> If we go visit relatives or friends we have to tell them ahead of time that we are bad eaters. We don't want them to think that they are bad cooks or insult them so we have to swallow our pride and admit to a major fault and nobody likes to admit and tell people what their faults are. (Joyce 2015)

The writer contrasts the two kinds of explanation for non-eating: that she and her husband are 'bad eaters' and that their relatives or friends are 'bad cooks'.

Given the nature of the context under consideration, there are advantages and disadvantages to these two ways of excusing one's non-eating. Consider the first kind of excuse. It's easy to know whether one finds food displeasing, and (with a few caveats) it's hard for others to show otherwise.[6] So this kind of excuse is easy to provide and hard to challenge. However, in the context under consideration, the fact that one finds food displeasing will not allow one to escape the consequences of violating the rule about eating even what you find displeasing. In that context, it doesn't matter whether one dislikes the taste of the food: that's not a reason for not eating it. Consider the second kind of excuse, viz. that the food is deviant. Even if the first kind of excuse is not acceptable in the context, this second kind of excuse *can be* an acceptable excuse in the context. However, it's

[6] It's possible to challenge someone's claim that she does not like something by citing the fact that she ate it last week: there are coherence constraints on our food preferences which others can use to challenge our own claims about what we like and dislike (cf. Barker 2013).

also easier to challenge. If the food is not identifiably other than it should be, then this kind of excuse will appear unwarranted.

To see this, suppose that one puts forward a negative taste assessment in order to excuse one's non-eating in a context where one is under an obligation to eat even what one finds not pleasing. In the contexts we're envisaging, that assessment is accountable to the following question: "What's wrong with it?" (cf. Wiggins 2004, 34-35). This question presupposes that there is something wrong with the food. One can respond to this question by rejecting the presupposition: "Nothing. I just don't like it." Then one's excuse becomes of the first kind: one is explaining one's non-eating with what one finds gustatorily displeasing. But if one does this when the rule is that you eat even what you find displeasing, this excuse won't work. It won't ease the violation of the rule. Alternatively, one can accept the presupposition of the question. Then one will be pursuing the second kind of excuse: one is explaining one's non-eating with the food itself and not with what one finds gustatorily displeasing. However, in that case, one needs to provide an answer to the question and an answer to the question will have to identify something about the food that explains one's non-eating but which does not reduce to the fact that one finds the food displeasing. Insofar as one cannot do that, there will seem to be something amiss with the negative taste assessment: one's inability to answer the question will seem to betray one's negative taste assessment as unwarranted. This is so even though one knows first-hand that one finds the food displeasing.

So we have a kind of context wherein there's an obligation to eat food even if one finds it displeasing. The fact that one finds the food displeasing is therefore no excuse for not eating the food. But a second kind of excuse that focuses upon a way in which the food deviates from how it ought to be *is* an acceptable excuse for not eating the food. Let's say that when someone is putting forward a negative taste assessment in these circumstances as the second kind of excuse (and not the first), she is in *the dinner table context*. The dinner table context seems to make trouble for TP. Negative taste assessments are of the form, "This is not tasty." They involve application of "not tasty" to food. Suppose the food is not pleasing to you and you know this first hand. Then, according to TP, you are warranted in applying "not tasty" to the food and thus in making a negative taste assessment. However, if you make this negative taste assessment in a context wherein you are under an obligation to eat even what you find not pleasing,

and if you proffer the negative taste assessment as the second kind of excuse for not eating the food, then your assessment will seem unwarranted insofar as you are unable to provide an answer to the question, "what's wrong with it?" which grants the presupposition of the question. Just like Joyce, the adult picky eater, if you were to make the negative taste assessment in the dinner table context when nothing is awry with the food (as distinct from one's disliking it), you'll seem to be mis-explaining your non-eating. TP therefore predicts that we should find the negative taste assessment warranted when it does not seem to be.

One might try to defend TP by suggesting that all we see here is an aversion to being rude. One might say, on MacFarlane's behalf, TP is not meant to describe variations in the use of "not tasty" that derive from attempts to avoid being rude. So the observations made here pose no problem for TP, as MacFarlane intends it: once we screen out uses of "not tasty" that are shaped by a sensitivity to politeness, we'll find no exception to TP. Let's grant that uses of "not tasty" which are guided by attempts to avoid being rude are irrelevant to the correctness of TP. Nonetheless, there will still be some instances of the dinner table context in which we systematically avoid negative taste assessments, even when we're not attempting to avoid being rude. There is a difference between the following two kinds of predicament. Firstly, you know that there is something deviant about the food – it does deviate from a relevant norm regarding its state or production – but you hold back from making negative taste assessments of the food because, even though you know this, you don't want to be rude. Secondly, you refrain from making negative taste assessments because you know that there is nothing deviant about the food, despite the fact that you don't like it. You know that if you made the assessment, your assessment could be held to account with the question "what's wrong with it?" and you know you wouldn't be able to point to anything besides the fact that you don't find the food pleasing. In these latter cases, you are refraining from making a negative taste assessment of the food but not because of a fear of being rude. It would be more accurate to say that you know you would be providing an incorrect explanation of your non-eating. You would be locating the reason for your non-eating in the food when really it lies in your quirky food preferences. Insofar as there is a difference between refraining from making a negative taste assessment simply because you don't want to be rude and refraining

from making a negative taste assessment because you know you would be providing an incorrect explanation of why you're not eating the food, the phenomenon exhibited in the dinner table context is not reducible to attempts at avoiding rudeness.

4. An alternative to TP

The problem with TP is not hard to fix. We can amend TP as follows:

TP' If you know first-hand how something tastes, call it "tasty" just in case its flavour is pleasing to you, and call it "not tasty" just in case its flavour is not pleasing to you *and* you are not in the dinner table context.

TP' requires that you use neither "tasty" nor "not tasty" when you are in the dinner table context. Note that this does not mean that in such a context you cannot make other kinds of assessment of the food e.g. "I don't like it." It's just that, if you do this, you will putting forward an assessment that doesn't count as an excuse for not eating that is acceptable in the context. But TP' does allow you to use "tasty" and "not tasty" in accordance with what tastes are pleasing (or not) to you (when you know this first hand) when you are not in the dinner table context.

TP' accounts for the fact that in (at least many) contexts besides the dinner table context, a negative taste assessment can be used in accordance with TP unproblematically. Consider, for instance, the context that takes up centre stage in MacFarlane's discussion of "tasty"/"not tasty":

> You bite into a fresh apple. It is the tart kind that you particularly like, and it is perfectly ripe. "Tasty," you say, without a moment's hesitation. (MacFarlane 2014, 1)

Unlike in the dinner table context, in this "solitary mutterer" context life is relatively simple: there's no obligation which one is excusing oneself from by making an assessment of the food. An assessment therefore cannot be an excuse from such an obligation. In such a context, we do seem to make negative taste assessments of the apple just in case we find its taste not pleasing and we know its taste first hand. MacFarlane's mistake isn't that

TP is never right but rather that he over-generalizes from his "solitary mutterer" context.

I doubt very much that TP' is the end of the story – but that's not its point. The conversation analyst Harvey Sacks once remarked:

> however rich our imaginations are, if we use hypothetical, or hypothetical-typical versions of the world we are constrained by reference to what an audience, an audience of professionals, can accept as reasonable. That might not appear to be a terrible constraint until we come to look at the kinds of things that actually occur. (Sacks 1984, 25)

I offer TP' not because I think it is correct but because it is a fair summary of what takes place both in the imagined context upon which MacFarlane focuses and in a context that has been documented in existing sociological fieldwork – fieldwork which has informed the content of this paper. TP' thus makes explicit the fact that contexts of use for "tasty"/"not tasty" can provide *varied* constraints on the warranted use of "tasty"/"not tasty". These constraints are not predictable from the imagined contexts of use that may first form in one's imagination – Sack's 'hypothetical-typical versions of the world'. Moreover, as we will see in section 6, these constraints can be relevant to our semantic and pragmatic theorizing about the predicate. So, despite its likely incompleteness, TP' makes salient the importance of treating contexts of use as unknown unless studied and not as objects so well known that we can rely solely on our imaginations to gather data about how language is used within them.

5. MacFarlane's reasons for accepting TP

Let's return to MacFarlane's reasons for accepting TP. His first reason for accepting TP is that it is self-evident that TP is true: never mind the context, this is just obviously how we use "tasty"/"not tasty". His second reason for accepting TP is that (1)–(3) sound odd and TP predicts this. His third reason for accepting TP is that "tasty"/"not tasty" serves a purpose it would not or could not serve if we didn't use it in accordance with TP.

We can now see that the first and third reasons are unfounded. The one context of use that MacFarlane seems to bear in mind when thinking about

the use and purpose of assessments made using "tasty"/"not tasty" is the solitary mutterer context. Sure enough, there the predicate is plausibly used in the way, and serves the purpose, that MacFarlane outlines. The same cannot be said of other contexts – such as the dinner table context we have been discussing. So the first and third reasons do not have the contextual generality MacFarlane supposes them to have.

MacFarlane's second reason for accepting TP was that (1)–(3) sound odd:

(1) I'm not sure whether espresso is tasty, but I hate how it tastes.

(2) I've never been able to stand the taste of durian. Might it be tasty?

(3) I love orange juice and hate tomato juice. But who knows? Perhaps tomato juice is tastier.

TP predicts this oddness – because to say (1)–(3) is to violate TP. What does TP' imply about the oddness of these sentences? It makes exactly the same predictions as TP except when speakers are in the dinner table context. In that context, TP' predicts that speakers will consider positive taste assessments unwarranted because the speaker dislikes the taste of the food and TP' predicts that speakers will consider negative taste assessments unwarranted even though the speaker does not like the taste of the food. None of this implies that (1)–(3) should not be odd. But, because (unlike TP) TP' does not imply that speakers in the dinner table context will say that something is not tasty when they do not like its taste, TP' also does not imply that (1)–(3) *are* odd. However, it is not hard to find a subsidiary rule which is compatible with TP' but which does predict the oddness of (1)–(3):

Subsidiary Rule

If you know first-hand that an item tastes bad to you (or that one item tastes worse to you than another) and you are in the dinner table context then don't express uncertainty about whether the item is tasty (or about whether one item is less tasty than the other).

When used in the dinner table context, (1)-(3) violate this rule. But then we can conclude that the oddness of (1)-(3) is compatible with both TP and TP' and that it can even be predicted by both – at least, when the latter is

supplemented with a suitable subsidiary rule. So MacFarlane's second reason for accepting TP does not provide us with any reason to favour TP over TP'.

6. TP and objectivism

Recall from section 1 that MacFarlane (2014, 2) defines objectivism about "tasty"/"not tasty" as follows:

Objectivism

(a) "tasty" is true of some things, false of others, and

(b) whether "tasty" is true or false of a thing, on a particular occasion of use, does not depend on the idiosyncratic tastes of the speaker, assessor, or anyone else.

As we noted earlier, MacFarlane rejects objectivism because he thinks that it has an implausible consequence, given two further assumptions: firstly, that for any food, some speakers find it pleasing and others find it displeasing; and secondly, that TP is the rule speakers follow in counting applications of "tasty"/"not tasty" as warranted. Given these two assumptions, for any food, some speakers will judge it tasty and others will judge it not tasty but one group of speakers will be mistaken. This implies that speakers are using an evidently unreliable method for forming judgements and claims about what is tasty. MacFarlane rejects objectivism because he thinks it is implausible that speakers would use an evidently unreliable method to form taste judgements and claims.

If, as we have found, TP is not our method for using "tasty"/"not tasty" then what should we say about objectivism? One thing we cannot say is that objectivism is a live option for "tasty"/"not tasty" in all contexts. We have already conceded that there are some contexts – like the solitary mutterer context – in which TP is a pretty good description of the method, in that context, for using "tasty"/"not tasty". So in such contexts, MacFarlane's argument against objectivism seems plausible. If objectivism were true, then in those contexts, we would be using a method for using "tasty"/"not tasty" that is evidently unreliable and perhaps it is unreasonable to suppose that we would use such a method. Similarly, nothing has

been said which suggests that MacFarlane's argument goes wrong when applied to uses of "tasty" – regardless of whether those uses appear inside or outside the dinner table context. But what about uses of "not tasty" within the dinner table context? In that context, our use of "not tasty" is more sophisticated than TP. In that context, we consider negative taste assessments to be warranted only if the fact that a speaker finds the food displeasing correlates with some respect in which the food is deviant: *did* dad use Auntie Bessie's instead of the usual Paxo? Gustatory displeasure has to be an indicator of something beyond itself.

Compare this use of "not tasty" in the dinner table context with the use of "red." Recall that "red" is an expression which MacFarlane takes to be objective, despite being context-sensitive. He thinks that although one is warranted in using "tasty"/"not tasty" in accordance with one's idiosyncratic pleasure reactions to food, one is not analogously warranted in using "red" in accordance with one's idiosyncratic visual reactions to objects. So, if you're colour blind (e.g. you see both red and green as brown), that doesn't mean that you are warranted in applying "brown" to red and green objects. But, MacFarlane thinks, you *are* warranted in applying "tasty" to foods just in case you like them and "not tasty" just in case you don't.

The dinner table context is a context in which this contrast between "red" and "not tasty" breaks down. We plausibly use each expression in accordance with a sense: be it our sense of taste (whether tastes please us) or our sense of colour (whether a thing visually seems red to us). But with "red" generally, and with "not tasty" in the dinner table context, we do this only if our subjective experience of taste or colour (systematically) corresponds to certain features of our environment. If it does not then we do not count the use of either predicate in accordance with the corresponding sense as warranted. Rather, we switch to more subjective ways of speaking. In the case of "red", we switch to speaking about how things look or seem to us e.g. "it seems brown to me." In the case of "not tasty", we switch to speaking about what we like, e.g. "I don't like this." In short, *pace* MacFarlane, in the dinner table context, the negative form of "tasty" (viz. "not tasty") is just as objective as "red". In the dinner table context, if our own taste preferences aren't tracking deviancy in the relevant food (whatever deviancy may amount to in the context), then we don't rely on our own tastes to guide our use of "not tasty" – just as when, if how things seem

visually to someone doesn't correspond to how they are (as when she is colour blind), then that someone won't use "red" in accordance with how things visually seem to her. So for uses of "not tasty" in the dinner table context, MacFarlane's reason for rejecting objectivism finds no application. In that context, we do not employ a method for using "not tasty" which is highly unreliable in getting us to make true judgements and claims, given the truth of objectivism in that context. So, as far as MacFarlane's argument against objectivism is concerned, objectivism about "not tasty" is still an option for the dinner table context.

I will now close the paper by addressing three *prima facie* routes out of this conclusion. The concerns raised in the following discussion will force the concession that only certain kinds of dinner table context are a problem for MacFarlane's argument against objectivism. But, I will argue, the concerns raised are not successful defences of the argument.

The first is as follows. We could identify the use of "not tasty" in the dinner table context with an exo-centric use. Even if, in such a context, we don't follow TP, perhaps we're using "not tasty" in accordance with what someone else (or some group) finds displeasing. If so, then, given objectivism, even in the dinner table context, we would be forming taste judgements and claims using an evidently unreliable method. Does this possibility undermine the sketched defence of objectivism for "not tasty"? There's no denying that in the dinner table context one *can* use "not tasty" exo-centrically. For example, a deferential husband may use "not tasty" in accordance with his wife's gustatory dislikes. Noticing that his wife doesn't like the food he's cooked, and wanting to be congenial, he meekishly excuses himself from eating it by saying, "this is not tasty," whilst using "not tasty" in line with his wife's gustatory dislikes. If the question were raised, "what's wrong with it?" the correct answer would be: my wife doesn't like it. Thus, despite this being the dinner table context, "not tasty" is used in accordance with someone's gustatory dislikes, i.e., in a way that is unreliable given the truth of objectivism. However, although one can do so, one does not have to use "not tasty" in this way in the dinner table context. One could also use "not tasty" in a way that is shaped by a sensitivity to whether the food deviates from some norm, where the norm cannot be identified with anyone's gustatory dislikes. In that case, the correct answer to the question "what's wrong with it?" will describe some way in which the food deviates from a norm other

than how some privileged individual or group reacts to the food in question. Insofar as there are some such uses of "not tasty," there are some instances of the dinner table context in which MacFarlane's argument against objectivism finds no application: in those instances, there's no use of an evidently unreliable method for reaching judgements about what's tasty, even given objectivism. On the contrary, speakers are exhibiting a careful sensitivity to whether the method described by the original TP rule really is tracking certain contextually relevant objective facts about the food: they aim to refrain from making a negative taste assessment if they suspect their displeasure isn't tracking anything deviant about the food. So even though, of course, one can use "not tasty" exo-centrically, whilst excusing oneself from eating food that one finds displeasing and which one is under an obligation to eat, that doesn't mean that the dinner table context provides no trouble for MacFarlane's argument against objectivism.

A second *prima facie* route out of this conclusion focuses on the matter of what settles which norm, in a given context, determines how food ought to be (or normally is). Let's assume that the norm is not identifiable with some privileged person's or group's gustatory reactions to the food. Thus, whether or not a food deviates from the norm is not simply a matter of whether the food is displeasing to some privileged person or group. For example, it might be that in the context, pastries are deviant if they are soggy and not dry. Whether a given pastry is soggy and not dry, is not a matter of whether the pastry is pleasing to some privileged person or group. However, what factors influence which norm is in place in a given context? Perhaps which norm is in place in a given context will be determined in some more or less intimate way by what certain privileged persons find gustatorily pleasing. For example, although *being soggy* is not the same as *being pleasing to X*, the fact that, in a given context, pastries are deviant if they are soggy, may be strongly influenced by what some privileged persons find gustatorily pleasing. If so, then isn't the second kind of excuse that we have been considering – the one that explains non-eating by appeal to deviancy in the food and not by appeal to someone's gustatory reactions to the food – just as dependent on the idiosyncratic tastes of a privileged person or persons? If it is, then MacFarlane's argument against objectivism would apply even to "not tasty" as used to put forward this second kind of excuse.

I agree that gustatory reactions may (though they need not) play a role in fixing the norm against which deviancy is measured in a given context, even if deviancy itself is not a matter of deviating from what someone finds pleasing. Again: although in some context, pastries *are* supposed to be dry and not soggy, this norm might be in place in the context because certain privileged persons find soggy pastries gustatorily displeasing. However, if one attempts to defend MacFarlane's argument against objectivism in this way, then it's hard to see how a contrast can be sustained between "not tasty" and other gradable adjectives like "red" or "tall." Subjective factors which vary between speakers are just as likely to play a role in settling the contextually relevant standards for being tall or being red as they are for settling a norm of deviancy for food to which "not tasty" might be applied: this is so, even if the standards themselves are not simply that something seems tall to some privileged person or that some object seems red to some privileged person. For example, in a given context, an apple might have to be red on its surface in order to qualify as a red apple. Whether an apple is red on its surface is not identifiable with how the apple seems to any privileged person. However, the fact that that is how an apple has to be in order to be red, in the context, could be something which is dependent in some more or less intimate way upon the subjective states of some privileged persons. So if this is reason enough to say that use of "not tasty" varies with the idiosyncrasies of the speaker, an assessor or somebody else, then MacFarlane's anti-objectivism argument will apply to *all* gradable adjectives. I take this to be a *reductio* of this defence of the anti-objectivism argument. MacFarlane aims to draw our attention to a behaviour of "tasty"/"not tasty" which distinguishes it from other gradable adjectives. One can deny (in the way just described) that the argument against objectivism falters at the use of "not tasty" which I have described in this paper, only by forfeiting this ambition. The sense in which "tasty"/"not tasty" is supposed to be subjective, and which distinguishes it from other gradable adjectives, is that, when using it in accordance with your gustatory likes and dislikes, you don't have to worry about whether your subjective assessments of something's being tasty are really detecting anything beyond whether you find the food gustatorily pleasing. But in this sense of subjective, the use of "not tasty" in the dinner table context that we have focused upon is not subjective.

Finally, a third *prima facie* route out of our conclusion begins with the thought that we should acknowledge that some answers that one can give in response to the challenge "what's wrong with it?" are exemplified by the following nuanced assessments: the cheese clashes with the gravy, or, the coriander is misplaced. One might think that such remarks can show that the challenged negative taste assessment (*qua* excuse for not eating the food) is warranted. Yet, such remarks might reasonably be called "subjective". Whether the cheese clashes with the gravy is surely a matter of what displeases a privileged X. If so, then surely the negative taste assessment itself may reasonably be called "subjective" because it is being used in accordance with what displeases a privileged X. But if so, then, it seems, even in the dinner table context, "not tasty" is still being used in accordance with what displeases a privileged X. So MacFarlane's argument still applies to such uses of "not tasty": even in those contexts, speakers are using "not tasty" in a way that, if objectivism were true, would imply that speakers are using a method to make taste judgements and claims which is evidently unreliable. Speakers wouldn't do that. So objectivism must be false in these contexts.

I have already conceded that there are some versions of the dinner table context in which the standard against which food is assessed is reducible to what displeases a privileged X (cf. the congenial husband). That concession doesn't ruin the objection laid out in this paper because there are other contexts in which the standard used does not reduce to such. A similar response can be given to the current concern. Let's begin by sorting answers to the question "what's wrong with it?" which seem to describe the food, and not what displeases a privileged X, into two groups. There are those that are not distinguishable from claims to the effect that the food is displeasing to a privileged X: there is no conceptual gap between whether the claim is true and whether the food is displeasing to X. This would happen if, for example, someone says, "the cheese clashes with the gravy" but the use made of this sentence is interchangeable with the speaker's use of "X doesn't like this". Then again there are those answers that *are* distinguishable in this way: there is a conceptual gap between whether the claim is true and whether the food is displeasing to X. This would happen if, for example, someone says, "the cheese clashes with the gravy" but the use made of this sentence is not interchangeable with the speaker's use of "X doesn't like this".

Now, in order for the objection that I raise against MacFarlane's argument against objectivism to fail for the reason laid out in this third *prima facie* response, there can be no context in which only reasons of the second kind – reasons that are not reducible to the dislikes of X (whoever X may be) – suffice to make a negative taste assessment warranted. Insofar as there are such contexts, there are uses of "not tasty" which are not guided by what is displeasing to a privileged X unless that correlates with something else. The objection indicates that there might be dinner table contexts in which answers are given to the question, "what's wrong with it?" which are not explicit statements of the form "X doesn't like this" but which are nonetheless used in a way that is interchangeable with such statements and which make a negative taste assessment warranted. However, this doesn't show that there are no dinner table contexts in which the answers given to "what's wrong with it?" are of the second kind; and that's all we need in order to make trouble for MacFarlane. Are there any such contexts? I think there are likely to be two kinds of such contexts.

Firstly, regardless of whether nuanced assessments are reducible to a matter of whether the food is displeasing to a privileged X, there are contexts in which nuanced assessments generally will not make a negative taste assessment (*qua* excuse for not eating one's food) warranted. There are some contexts where, if one tried to use a nuanced assessment in this role then one would come across as drawing spurious distinctions which don't correspond to anything real in the food: certainly, not real enough to excuse one's non-eating. Just suppose, for example, that a friend of yours, has cooked a nutritious lunch for you, despite her relative poverty, and you try to justify not eating it because, "the cheese clashes with the gravy." It does not seem to me that such a nuanced assessment will show the negative taste assessment to be warranted in this context: neither you, nor your friend, would think that this suffices, in the context, to make the negative taste assessment warranted. Even if the cheese clashes with the gravy, provided the food meets some other more down-to-earth standards, this won't show that your negative taste assessment is warranted—regardless of whether nuanced assessments are reducible to a matter of whether the food is displeasing to a privileged X. In this kind of context, only non-nuanced assessments can be given in response to the challenge "what's wrong with it?" and stand a chance of making the challenged negative taste assessment warranted. Insofar as the non-nuanced assessments draw attention to

uncontroversially objective features of the food or its production, the responses will not be reducible to whether the food is displeasing to a privileged X.

Secondly, it's not at all obvious that nuanced assessments themselves inevitably reduce to assessments of whether the food is displeasing to some privileged X: i.e. that inevitably there is no conceptual gap between whether the food is displeasing to X and whether there is, for example, a clash between the cheese and the gravy. This thesis needs an argument. MacFarlane provides none and, as far as I am aware, neither does anyone else. Insofar as there are *some* contexts in which there's no such reduction, even if nuanced assessments do show a negative taste assessment (*qua* excuse for not eating) to be warranted, those contexts will be just as problematic for MacFarlane's argument against objectivism as the contexts just considered, i.e. those in which nuanced assessments fail to show a negative taste assessment (*qua* excuse for not eating) to be warranted.

Acknowledgments

I would like to thank the anonymous reviewers of this journal for very helpful comments. These helped me to improve the paper significantly. The research of this paper was funded with Estonian Research Council grant PUT632 and IUT20-5.

References

BARKER, C. (2013): Negotiating Taste. *Inquiry: An Interdisciplinary Journal of Philosophy* 56, No. 2/3, 240-227.

GRIESHABER, S. (1997): Mealtime Rituals: Power and Resistance in the Construction of Mealtime Rules. *The British Journal of Sociology* 48, No. 4, 649-666.

HIRVONEN, S. (2016): Doing without Judge-Dependence. In: Meier, C. & van Wijnbergen-Huitink, J. (eds.): *Subjective Meaning. Alternatives to Relativism* Berlin: de Gruyter, 47-67.

JOYCE (2015): *Joyce's Story*. Available at: http://pickyeatingadults.com/index.php/2015/08/03/joyces-story/

LASERSOHN, P. (2005): Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy* 28, 643-686.

MACFARLANE, J. (2014): *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.

OCHS, E., PONTECORVO, C. & FASULO, A. (1996): Socializing taste. *Ethnos: Journal of Anthropology* 61, No. 1-2, 9-46.

SACKS, H. (1984): Notes on Methodology. In: Atkinson, J. & Heritage, J. (eds.): *Structures of Social Action: Studies in Conversation Analysis.* Cambridge: Cambridge University Press, 21-27.

WIGGINS, S. (2004): Talking about Taste: Using a Discursive Psychological Approach to Examine Challenges to Food Evaluations. *Appetite* 43, No. 1, 29-38.

Organon F 24 (3) 2017: 427-432

Peter Olen: *Wilfrid Sellars and the Foundations of Normativity*
Palgrave Macmillan, London, 2016, viii + 240 pages[1]

A few decades ago, Wilfrid Sellars – in contrast to Quine, or Davidson – was not a philosopher for the philosophical masses. He held sway in certain circles and his disciples were becoming more broadly influential, but he himself was not the kind of philosopher who would be the target of extensive interpretative and exegetic effort. However, times have changed and Sellars has become almost a fashionable figure, with a growing number of books and papers devoted to his philosophy.

Despite the fact that Sellars is beginning to level up with Quine or Davidson in respect to the attention he attracts, he is still (and will remain) quite different from them in respect to the nature of his work: his writings are neither so elegant and transparent as those of Quine, nor so neat and self-contained as those of Davidson. His teaching is very intricate, indeed sometimes to the point of unintelligibility. And what holds of his work in general, holds the more about the first period of his productive academic life, which spanned from the late forties to the early fifties. As a result, though we already have two insightful book-length expositions of Sellars' teaching (due to O'Shea 2007, and deVries 2005), a thorough study of this early period, which Peter Olen offers in his book, is certainly not superfluous.

Olen's book (like the works of Sellars himself) is not easy to read. I must say I wrestled with some passages of his book just like the biblical Jacob with the angel; but in the end most of them did give me their blessing. (The kind of obstacles I had with the text will become apparent in what follows.) Despite this, Olen's detailed discussions of the twists and turns of Sellars' thought are mostly illuminating, though I would not agree with everything he puts forward. In particular, I would not subscribe to some of his categorical rejections of Sellars' views as "misunder-

[1] ✉ Jaroslav Peregrin
Institute of Philosophy, Czech Academy of Sciences
Jilská 1, 110 00 Praha 1, Czech Republic
e-mail: peregrin@flu.cas.cz

standings" or "conflations". This, of course, is not to say that Sellars' labyrinthine early development is utterly free of needless detours and blind alleys, nor that Sellars' failures should not be criticized. Only I think that in some cases these apparent failures are not so clearly due to Sellars himself.

Many of the musings of the early Sellars can be ascribed to his coping with the legacy of Rudolf Carnap. Carnap's step from the syntactic phase of his *Logische Syntax der Sprache* (Carnap 1934) to the semantic phase of his *Introduction to Semantics* (Carnap 1942) was marked by his decision that he could go beyond *formal* theory of language without abandoning the theory of language that is *pure* (in the sense that the theory does not rest on empirical facts). Olen points out that this step was the subject matter of much dispute and criticism in the philosophy department at Iowa (Bergmann, Hall, Hinshaw, Feigl), where Sellars found himself in the forties; and he points out that it was also a good source of confusion if only because the term *formal* was construed differently by different philosophers (and it was quite ambiguous in Sellars' own early writings). Olen writes bluntly about the Iowa philosophers *misreading* Carnap. (Personally I would hesitate to call it a misreading. It seems more probable that the Iowa philosophers simply wanted Carnap to be more consistent with the earlier views of his syntax period than he himself was willing to be.)

In the *Logical Syntax of Language,* Carnap stresses that what a sentence about denotation, such as "The word 'luna' in the Latin language designates the moon" is really stating (and which becomes apparent when it is transformed from the misleading material mode of speech into the formal one) is that "There is an equipollent expressional translation of the Latin into the English language in which the word 'moon' is the correlate of the word 'luna'" (Carnap 1934, 215). I think that Sellars, like other readers of Carnap's *Syntax* understood that there are some deep reasons for the *prima facie* strange claim that the sentence does not simply report a relationship between an expression and its denotation – reasons which later led Sellars to his theory of "meaning as functional classification" (cf. Sellars 1974). Yet in his *Introduction to Semantics* Carnap not only comes to disregard these reasons and takes denotation at face value, but, moreover, he claims that we can have *pure* semantics which incorporates *contingent, extralinguistic* objects. But as a claim such as "'c' denotes Chicago" obviously presupposes some factual claims, such as 'There is Chicago', what is the sense of *pure* in *pure semantics*?

Regarding the acceptance of Carnap's "semantic turn" by the Iowa philosophers, especially Bergman and Hall, Olen writes:

> While it is still plausible that all three philosophers are talking past each other, it is difficult to deny Bergmann's and Hall's confusion rests on a misreading of Carnap's work. Hall's initial exploration of pure semantics, for example, starts by asking the question "How is a word or sentence about extra-linguistic matter of fact related to the matter of fact it is about?"… This badly misconstrues Carnap's project: pure semantics is not concerned with constructing a factual relationship between a language and extra-linguistic objects. (p. 27)

Trying hard as I may, I cannot understand why asking the question "How is a word or sentence about extra-linguistic matter of fact related to the matter of fact it is about?" would lead us to "constructing a factual relationship between a language and extra-linguistic objects" and thus to a misconstrual of Carnap's project. Understanding, on a general (or "pure", if you wish) level, what is the nature of the link which attaches something factual to an expression making the former into the meaning of the latter seems to me to be the most basic step in the explication of the concept of meaning, rather than "constructing a factual relationship".

Also, another difference between Carnap and Sellars, as articulated by Olen, seems to me to be problematic:

> While Sellars claims conformation rules fulfill essentially the same role as meaning postulates…, this claim misses a substantial difference between the two concepts. Specifically, the fact that meaning postulates, located in a semantic meta-language, are determined by a matter of decision is what makes them contingent in any syntactical or semantical system. Conformation rules, however, cannot be contingent – their inclusion in pragmatic accounts of empirically meaningful languages is presented as a necessary aspect of pure pragmatics. (p. 57)

As far as I can see, what Sellars is claiming is that an empirically meaningful language must contain some conformation rules (or, for that matter, meaning postulates). Carnap, on the other hand, does not ponder upon "empirical meaningfulness", because a substantial part of his investigations focuses on the languages of pure logic (and another part focuses on the non-empirical languages of mathematics). Thus the *general* necessity of including conformation rules (meaning postulates) arises only when we restrict our attention to "empirically meaningful languages". Moreover, both *specific* meaning postulates and *specific* conformation

rules are "determined by a matter of decision" if what we are dealing with are abstract ("pure") languages, whereas they are a factual, contingent matter when occurring within factual languages.

To avoid misunderstanding, I am not claiming that Sellars was right and Carnap wrong; I just think that the situation is more complicated than to be characterized as Sellars' misunderstanding of Carnap. The same, it seems to me, holds about the Sellars' alleged "conflation" of "two different senses of 'language' – calculi or semantical systems on the one hand (the languages investigated in pure studies of language), and natural language on the other" (p. 148). Though Sellars' views on the matter are admittedly somewhat convoluted, I do not think that "conflation" is an appropriate verdict.

Olen shows that the original writings (roughly up to the 1949 paper 'Language, rules and behavior', where he flies away from pure pragmatics) are – terminologically, conceptually and doctrinally – rather messy. (The papers were published as Sellars 1980.) He also shows how important it was for Sellars to find a suitable space for philosophy such that it would be safely isolated from science. (An attitude which a reader tending to see Sellars, perhaps under the influence of Rorty, as an ally of Quine may find perplexing.) And it was Carnap's idea of *pure* theory of language that served him as his point of departure – only he believed that if what is to be accounted for are "empirically meaningful languages", then this kind of theory has to be extended all the way to pragmatics. This went *contra* Carnap, for Carnap, though he thought that the pure theory can be extended beyond the formal theory, *viz*. from syntax to semantics, held that this is all, and that pragmatics must be left to the mercy of sciences such as psychology or sociology.

The importance that metaphilosophical considerations held for Sellars helps Olen to throw some new light on the transition of Sellars from his first philosophical phase, the phase of pure pragmatics, to his second one, where a more concrete notion of normativity moved to the fore:

> While Sellars' early conception of a linguistic rule conforms to the conception of a rule found in Carnap's and other analytic philosophers' works in the 1930s and 1940s, Sellars' later articulation of the norm-governed, behavioral conception of linguistic rules marks a drastic departure from his early meta-philosophy. Such rules, because of their behavioral nature, could not be formulated within a formalist meta-philosophy. It is only after Sellars abandons his early formalist position that he can develop the normative, sui generis conception of language and linguistic rules. (p. 130)

I think this is correct and revealing. However, I have some problems with Olen's distinction between what he calls the *external* and the *internal* notions of normativity, which he uses as a further explanatory tool:

> Carnap's description of rules as conforming to "the customary usage in logic" exhibits what can be called an 'internal conception of normativity'. This conception of normativity is defined by two distinct claims: (1) while rules still employ normative terminology (e.g., 'correct', 'incorrect', 'ought', 'ought not'), 'normative force' is only found relative to the voluntary adoption of a given language and, (2) the language used to characterize or explain linguistic rules does not require irreducible, sui generis terminology in order to explain their constitutive role in language. … The external conception of normativity – what I claim is found in Sellars' later conception of linguistic rules – is embodied in the idea that the proper characterization of rules requires the enlisting of normative terms with "surplus meaning over and above" descriptive terms (which generates an irreducible, sui generis conception of normative vocabulary) and these rules use – but are not exhausted by – behavioral and social science concepts… (pp. 132-133; 138)

I am not sure I understand this. As for the "internal conception of normativity", I do not understand it as a "conception of normativity" at all. Ad (1): It is hard to imagine a language the adoption of the rules of which would not be "relative to the voluntary adoption" of the language – it is hard to imagine a language the rules of which would be forced on me independently of my willingness to follow them. (To a certain extent, this might describe the situation of an infant linguistic novice who is taught the rules independently of her will?) Ad (2): what kind of "sui generis terminology" is relevant here? And concerning the "external conception of normativity": I think that Sellars' later conception of normativity indeed is characterized by the assumption that there is a "normative mode" of using expressions that is not wholly reducible to the declarative mode (where the difference could conceivably be characterized as a "surplus meaning"), but I do not understand why this conception of normativity is "external" and why it is opposed to the conception which Olen calls "internal".

As I read Sellars, he realized that the way in which natural languages are rule-governed is in that rules are not merely something that is appropriate for a theoretician to account for our linguistic practices, but rather that the practices do "incorporate" the rules, that they are instances of what Sellars later called "rule-governed

behavior". I think that it is for this reason that he saw Carnap's concept of rule as embodied in a definition, as a "snare and a delusion" (Sellars 1953, 329). I think that the distinction between "external" and "internal" concepts of normativity, invoked by Olen, is not of much help here: if we are to see as a rule anything to which a normative force can be added, as it were, from the outside, then the question is whether there can be anything that could not be seen as a rule.

All in all, I find the route of Sellars' early philosophy a fascinating philosophical "coming-of-age story", which Olen anatomizes with an unusual fervor. As I said, I do not find everything that Olen writes about Sellars' early development uncontroversial, but despite this, the book is certainly duly thought provoking; and it builds a new entering wedge into the fascinating world of Sellars' thought.

Jaroslav Peregrin

References

CARNAP, R. (1934): *Logische Syntax der Sprache*. Vienna: Springer.

CARNAP, R. (1942): *Introduction to Semantics*. Cambridge (Mass.): Harvard University Press.

DEVRIES, W. A. (2005): *Wilfrid Sellars*. Chesham: Acumen.

O'SHEA, J. (2007): *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge: Polity.

SELLARS, W. (1949): Language, Rules and Behavior. In: Hook, S. (ed.): *John Dewey: Philosopher of Science and Freedom*. New York: Dial Press, 289-315.

SELLARS, W. (1953): Inference and Meaning. *Mind* 62, 313-338.

SELLARS, W. (1974): Meaning as Functional Classification. *Synthèse* 27, No. 3-4, 417-437.

SELLARS, W. (1980): *Pure Pragmatism and Possible Worlds*. Atascadero: Ridgeview.

Organon F 24 (3) 2017: 433-434

Conceivability & Modality
June 19-20, 2017, Sapienza University in Rome[1]

Problems of modality have been employing metaphysicians for ages. Be it an attempt to provide the best logic of modal claims, truth conditions of such claims, an account of entities these claims are about, or the way we know whether such claims are true of false, the phenomenon is always present. Since the scope of problems the phenomenon reveals, it is not a surprise that conferences dedicated to them were, are and will, be organised. And, as it seems, *Issues on the (Im)Possible* conference found her 'epistemology' sibling: *Conceivability & Modality* conference.

This year, *Conceivability & Modality* was hosted by the Sapienza University in Rome. As the main topics of the conference adumbrated – conceivability & modal epistemology, logic of conceivability and the history of conceivability – together with the list of speakers – Francesco Berto (University of Amsterdam), Albert Casullo (Nebraska), Boris Kment (Princeton), Tito Magri (Sapienza), Antonella Mallozzi (The Graduate Center – CUNY), Daniel Nolan (Notre Dame), Jonathan Schaffer (Rutgers), Tom Schoonen (University of Amsterdam), Anand Vaidya (San Jose State University), Barbara Vetter (Freie Universität Berlin) – the event brought together the most recent debates in modal epistemology.

After a short and warm welcome from Antonella Mallozzi and Tito Magri, the conference commenced with Anand Vaidya's 'Re-Conceiving Conceivability in light of the History of 20th Century Theories of Conceivability'. In it, Vaidya wondered where conceivability theory can go in light of the vast amount of criticism it has received, and aimed to see what lessons we can learn from these criticisms. Daniel Nolan's 'Imaginative Resistance as Parochialism' argued that although using our ability to conceive plays some role in modal epistemology, the keys to modal knowledge need to be sought elsewhere.

[1] ✉ Martin Vacek
Institute of Philosophy
Slovak Academy of Sciences
Klemensova 19, 811 09 Bratislava, Slovakia
e-mail: martinvacekphilosophy@gmail.com

After lunch, Antonella Mallozzi presented such an account of modal epistemology according to which essential properties of individuals (and kinds) are characterised by their causal roles. This 'Putting Modal Metaphysics First' approach promises to fit nicely with Kripke's examples of a posteriori necessities as well as other cases. In her 'Potential Knowledge' Barbara Vetter considered different ways in which knowledge of our own abilities and powers, as well as knowledge of the dispositions and tendencies of the objects, can be acquired. The end of the first day belonged to Francesco Berto's 'Logic Will Get You from A to B. Imagination Will Take You Everywhere'. Berto proposed a view on imagination as a logically anarchic activity which combines a modal semantics with a mereology of contents.

The second day of the conference started with Tito Magri's 'True Humean Modalities' the aim of which was to address modal Humeanism via the essential commitments of modal Humeanism and, secondly, how these commitments shape the Humean conception of conceivability as a guide to possibility. In his 'Modal Scepticism and Kung's Epistemology', Tom Schoonen critically evaluated Kung's theory of imagination and suggested that it provides a very weak modal epistemology, unable to account for our knowledge of certain common modal claims.

Al Casullo's post-lunch talk entitled 'Modal Empiricism: What is the Problem?' overviewed Gordon Barnes's argument on behalf of Kant's contention that knowledge is a priori. Casullo's goals were to a) uncover several significant gaps in the argument and b) to show that it suffers from a common defect in rationalist arguments. 'The Conceivability Test for Possibility' given by Boris Kment explored the justificatory force of the conceivability test. For him, any application of the conceivability test to establish a modal claim must rest on pre-existing modal beliefs. Finally, Jonathan Schaffer's view as how 'To See the Worlds in a Grain of Sand' had the following form: possibly $p =_{df}$ if there is an intrinsic profile F such that the world's being F ground-entails p's being true.

Conceivability & Modality was an extraordinary event. It attracted influential philosophers working on the epistemology of modality as well as made the impression that the research programmes in modal epistemology play indispensable roles in theorising about modality.

Martin Vacek