Guest Editor: Martin Vacek

Contents · Obsah

Modal Metaphysics: Issues on the (Im)Possible II

The five articles in this volume were presented at the Issues on the (Im)Possible II, a conference organized by the Institute of Philosophy of Slovak Academy of Sciences and Slovak Metaphysical Society. As usual, the primary goal of the conference was to address questions regarding modality (broadly construed). We are happy to say that the conference fulfilled the expectations. This editorial has two sections. The first section sets the papers in a contemporary context of modal metaphysics and epistemology, while the second one provides an overview of the papers themselves.

<div align="center">1</div>

The problems of modality penetrate both our everyday reasoning and theoretical investigations about it. Reasoning about things that could, could not or must happen, beside things that do happen, belongs to our everyday practice. Similarly, modal notions play important – or even indispensable – roles in philosophical theorizing. To mention just a few: an occurrence of one event necessitates an occurrence of the other; a set of properties A supervenes upon a set of properties B just in case no two things can differ with respect to A-properties without also differing with respect to B-properties; a set of sentences implies so-and-so just in case those sentences could not all be true together without so-and-so being true.

In order to approach the problem of modality, realistically oriented philosophers proposed to postulate various kinds of entities. Among them, possible worlds were thought to bring the desired theoretical benefits. For, possibility, impossibility and necessity conceived as going-on in some, none and every possible world respectively enabled quite elegant analyses. The burden of proof has then moved from conceptual analyses to metaphysical and epistemological justification of such a commitment. In the former, philosophers have been struggled with the question: "What possible worlds, if they exist, are?" In the latter, given we accept the existence of possible (and impossible) worlds: "How can we know them?" Both questions are essential to contemporary metaphysics and, as it goes in philosophy in general, receive still new answers.

The old problem in metaphysics of modality is the ontological status of possible worlds. The traditional literature on the topic balances between two stances: possible worlds are either concrete or abstract in nature. Despite the fact that the distinction is far from being

clear, there is a general agreement that worlds are one of those categories. In opposition to realistic positions stand positions according to which possible worlds do not exist. However, such positions are still underdeveloped, or at least less developed than their realistic counterparts. It is therefore not a surprise that a part of this issue concentrates on the antirealist positions.

Given the (existent or non-existent) modal facts are about what is possible, impossible or necessary, the natural question arise: "How do we acquire beliefs, or even knowledge of such facts?" The answers vary from reliability of our intuitions; our ability to conceive, imagine or suppose the (im)possible; an appeal to counterfactual analysis of modal knowledge; the focus on the need of causal, empirical information to gain modal knowledge; to modal skepticism all the way down. Each of these positions has its own advantages but to find the best one is always a matter of distinct, and very often incompatible, criteria. We are happy that this issue contains contributions that aim at some positive results in modal epistemology.

2

Each of the five papers in this issue addresses some aspect of modal metaphysics, epistemology and logic.

Anthony Dardis discusses modal fictionalism, the Brock-Rosen argument and its consequence that modal fictionalists are committed to possible worlds. Dardis then suggests that we should consider alternative ways of analysing modal discourse. Namely, his Modal Fictionalism and Modal Instrumentalism *avoids the damaging conclusion of the Brock-Rosen argument by treating possible worlds as an instrument, an acceptable but false theory.*

A related paper by Zsófia Zvolenszky entitled An Argument for Authorial Creation *takes the thesis of artifactualism seriously. The paper builds on Kripke's initial idea that incorporating authorial creation is an intuitive and natural move. Beside the mere statement though, Zvolenszky constructs an argument according to which artifactualism, understood on the background of the causal-historical chain account of reference determination, plays well in comparison to both Meinongianism and nonactualism.*

Epistemologically oriented papers are Dusko Prelević's Modal Empiricism and Knowledge of De Re Possibilities: A Critique of Roca-Royes' Account *and* Why You Can't Actually Imagine the Impossible (But Think that You Can) *by Luke Malik.*

In the former, Prelević puts in contrast two representative accounts of modal epistemology: modal empiricism and modal rationalism. He states the difference between them as the difference between a priori and a posteriori access to modality. While the former is the view that there is an a priori access to metaphysical modality, the latter requires

a posteriori justification for at least some modal beliefs. Prelević then argues that the modal empiricists' epistemology of *de re* modality as presented in Roca-Royes' account faces serious problems and that modal rationalism is still an open alternative.

Luke Malik opens the old problem of imagination and its relevance to possibility (contra impossibility). His (Kripkean) position is that it is not actually possible to imagine metaphysically impossible things although he is happy to admit that it appears to be so. Malik's claim is backed up by a unique theory of imagination according to which epistemic frailty and misunderstanding the relation between objects and phenomenal properties give rise to the confusion between what is denoted and what it is represented.

Finally, Adam Tamas Tuboly's paper Quine and Quantified Modal Logic – Against the Received View *scrutinizes the debate about modality in general, quantified modal logic in particular. Tracing the debate back to Quine-Kripke paradigms, Tuboly thinks that although Kripke has provided some important responses to Quine's attack against modal logic, the gap between formal semantics and metaphysics leaves a lot of issues unaddressed. As he concludes, the debate about the modalities is still on the table and the ideas of Quine deserve a fresh start.*

<p style="text-align:center">* * *</p>

Martin Vacek
martinvacekphilosophy@gmail.com

Modal Fictionalism and Modal Instrumentalism

ANTHONY DARDIS

Department of Philosophy. Hofstra University
Hempstead. NY – 11549-1150. USA
Anthony.Dardis@hofstra.edu

ABSTRACT: Fictionalism is a strategy for retaining a theory without committing to its truth. This paper considers two kinds of fictionalism about possible worlds: modal fictionalism or "story operator" fictionalism, and modal instrumentalism. Difficulties for modal fictionalism are used to motivate endorsing modal instrumentalism.

KEYWORDS: Brock-Rosen – fictionalism – instrumentalism – modal fictionalism – possible worlds.

1. Introduction

Fictionalism is a strategy for accepting and using a theory we think is false or seriously dubitable (cf. Rosen 2005). Osiander couldn't believe that the earth moved but he wanted to use Copernicus' account of the revolution of the spheres. Hume argued that there is no such thing as causation but saw that we would not leave off talking about causes. We may be dubious that there are numbers or possible worlds, but we have overwhelmingly good reason to keep on talking about them. Fictionalism in each case tells us to count the theory as a fiction, and keep on using it.

There are many ways to do this. This paper discusses two recent versions of fictionalism about possible worlds, and uses a set of difficulties for one as a reason to endorse the other. The first is modal fictionalism or

"story operator fictionalism" (Rosen 1990).[1] We count the relevant theory (for example Lewis 1986) as a fiction, and then say true things by reporting what the fiction says, by way of equivalences:

> There could have been talking donkeys iff According to PW, there is a possible world that has talking donkeys as parts

(where "According to PW" is the story operator and PW is the fiction of possible worlds).

The other sort of fictionalism is instrumentalism. Instrumentalism treats PW as an instrument, an object for reasoning. It pairs sentences in natural language with sentences about possible worlds: for example, "there could have been talking donkeys" with "there is a possible world that has talking donkeys as parts". But it does not count the pairing as any kind of meaning equivalence.

This paper will describe the "Brock-Rosen" argument that modal fictionalism is committed to possible worlds anyway, and a variety of ways to respond to it.[2] The Brock-Rosen argument is remarkably resilient. The problem, I will argue, is the biconditional equivalence schema. Since modal instrumentalism does not use that schema it is not exposed to the Brock-Rosen argument. The paper ends by comparing modal instrumentalism with some other nearby fictionalist proposals about possible worlds. There are several further fictionalisms with which modal instrumentalism ought to be compared but will not for reasons of space. The conclusion is therefore limited: modal instrumentalism is a better view than modal fictionalism.

2. The Brock-Rosen argument

Modal fictionalism is designed for anyone who wishes to talk about possible worlds but doesn't believe there are any, and has no other strategy

---

[1]   Since Rosen (1990), "modal fictionalism" is the standard expression for story operator fictionalism, even though there are other ways to be a fictionalist about modality.

[2]   The Brock-Rosen argument is, of course, only one of the problems modal fictionalism faces: see Nolan (2011).

for avoiding commitment to their existence.[3] Possible worlds theory provides "translations" from sentences in natural language and quantified modal logic into sentences about possible worlds (cf. Lewis 1968, 117). For example:

> There could be talking donkeys iff $\Diamond\exists x(Tx \wedge Dx)$ iff there is a possible world that has talking donkeys as parts.
> Necessarily there are talking donkeys iff $\Box\exists x(Tx \wedge Dx)$ iff all possible worlds have talking donkeys as parts.

Modal fictionalism adds a layer of translation. It treats possible worlds theory as a fiction, and then reports what the fiction says the translation is:

> There could be talking donkeys iff According to PW, there is a possible world that has talking donkeys as parts.
> Necessarily there are talking donkeys iff According to PW, all worlds have talking donkeys as parts.

Brock and Rosen independently published an argument that modal fictionalism so understood does not do what it is supposed to do (see Brock 1993; Rosen 1993). The Brock-Rosen argument goes like this:

(1)   According to PW, at all possible worlds there is a plurality of worlds.
(2)   Necessarily, there is a plurality of worlds iff According to PW, at all possible worlds there is a plurality of worlds.
(3)   Therefore, necessarily, there is a plurality of worlds.
(4)   Therefore, there is a plurality of worlds.

(2) is the translation offered by modal fictionalism; (3) comes from (1) and (2) by biconditional elimination. (4) follows from (3) in some but not all modal logics; for example it does follow in S5.

About (1): It is clear that possible worlds theory does say that there are worlds. In Lewis (1968) this follows from Lewis's Postulates 7 and 8, plus

---

[3]   The usual foil for discussions of modal fictionalism is Lewis's account of possible worlds, on which they are concrete spatio-temporal objects just like the actual world. Modal fictionalism and modal instrumentalism work equally to defuse ontological commitment for other realist accounts of possible worlds as well.

the assumption that things could have been other than they are. (The argument of Rosen (1993) starts with the assumption that things could have been other than they are: "It is contingent whether kangaroos exist.") That there are worlds is not a world-bound truth, so, apparently, it is true at every world. Hence, apparently, it is true that according to PW, at all possible worlds there is a plurality of worlds.

2.1. Reflexive fictionalism

Despite the clarity of the objection, it is difficult to be confident about what's really going on. One index of the difficulty is the fairly wide variety of responses that have been made to the Brock-Rosen argument. I think what may be the clearest diagnosis comes from Nolan – O'Leary-Hawthorne (1996); the diagnosis does not, however, indicate what the right treatment should be.

They argue that the issue is what they call "reflexive fictionalism". Possible worlds theory models reasoning about possibility and necessity, that is, about modality. The theory itself has some modal status.[4] Applying the theory to itself to express that modal status, we get the Brock-Rosen argument. We get the same result for any fictional account that applies to itself. The parallel argument for the fiction of properties (PT) for example, goes like this:

(1)   According to PT the property of being red has the property of being monadic.
(2)   The property of being red is monadic iff According to PT the property of being red has the property of being monadic.
(3)   Therefore, the property of being red is monadic.

So the fictionalist about properties is committed to properties after all.

2.2. Denying (1): no worlds at worlds

Noonan, Divers and Kim in different ways argue that possible worlds theory doesn't underwrite (1).

---

[4]   See §4 below for a brief discussion of how amodalism (Cowling 2011), the view that some sentences lack modal status, interacts with modal fictionalism and modal instrumentalism.

2.2.1. The letter of fictionalism

Despite the intuitive plausibility of (1), Lewis's (1968) counterpart theory renders it as a falsehood. Rosen expressed the argument in the terms I used above: according to PW, at all worlds, there is a plurality of worlds. Noonan (1994) reminds us that the expressions "at all worlds" and "there is a world at which" are used in counterpart theory to restrict quantifiers. "There is a world at which there are talking donkeys" is translated as "There is a world w and there are things x such that the x's are in w and the x's are talking donkeys". Understood this way, the claim that at all worlds there is a plurality of worlds says that all worlds have worlds as parts. But there are not worlds *in* worlds. So any sentence of the form "at all worlds there are worlds" or "there is a world at which there is a world" is false.

Saving the letter of modal fictionalism leaves the spirit troubled. For surely it remains true that counterpart theory does say that there are worlds. Since counterpart theory talks about all worlds, the truth of the claim that there are worlds doesn't vary from world to world. So it is necessary if it is true. The modal fictionalist wants to say that there are no worlds other than the actual world. She may want to say that this claim is contingent, or necessary. So there are things we want to say about the modal status of possible worlds theory. That canonical possible worlds theory cannot say them seems more like a problem than a solution.

2.2.2. Advanced modalizing

Divers (1999a) calls modal claims that are about individuals in more than one world, and claims that are about more than one possible world, "advanced modalizing". Among the former are claims about properties (understood as sets of actual and possible individuals), sets, numbers, and cross-world modal comparisons ("my car could have been the same color as yours is"). Among the latter are the claims of interest to the modal fictionalist concerned about the Brock-Rosen argument: possibly there is a plurality of worlds, and necessarily there is a plurality of worlds. The problem for Lewis is that he is clearly committed to advanced modal claims, but counterpart theory has no means of expressing them.

Divers's solution is a "redundancy interpretation" of advanced modal claims. Counterpart theory adds a world quantifier to modal claims. Its

function is to restrict quantification inside the scope of the quantifier to worlds. The redundancy interpretation says that for the advanced modal claims, the possibility operator has no such effect:

> The simplest and best semantic story is this. Whenever the possibility operator expresses a *non-trivial* semantic function on quantificational sentences, it is, indeed, always that of altering the scope of formerly world-restricted quantifiers. So in cases where the quantifiers were not formerly world-restricted, the possibility operator has no semantic effect on the content of the sentences within its scope. The possibility operator is semantically redundant in such a context, a semantically vacuous expression on a par with "It is the case that". (Divers 1999a, 229)

We treat the necessity operator in the same way, either by defining it classically in terms of possibility and negation, or just by stipulating that, when it governs a sentence containing quantifiers that are not world-restricted, it is semantically redundant. Then the claim "Necessarily, there is a plurality of worlds" comes out saying only "There is a plurality of worlds."

The redundancy interpretation provides a solution to the Brock-Rosen problem. PW does say that there are possible worlds. That is an advanced modal claim, so it isn't handled as a quantification over worlds. Instead, it is a claim that can be directly translated into modal English, as "Possibly, there is a plurality of worlds". The fictionalist then says: and in fact there is not a plurality of worlds.

Divers – Hagen (2006) argue that this is not going to work. The problem, roughly, is that the redundancy interpretation treats all three of the following in exactly the same way (for A a sentence of advanced modalizing): It is possible that A; it is necessary that A; it is contingent that A. In all three cases, the modal operator drops off. So we have the consequence that

> it is possible that A iff it is necessary that A.

So we cannot simply endorse the contingency of PW and then say it is false; having endorsed the contingency of PW, we are also committed to its necessity.

2.2.3. Plural evaluation

Kim (2002) argues for a proposal about how any possible worlds theory should interpret "for all worlds" and "there is a world" claims. The proposal solves a variety of problems for counterpart theory, and so for modal fictionalism, including the Brock-Rosen problem. The core idea is to look at tuples (pairs, triples, etc.) of worlds in addition to looking at single worlds (see Bricker 2001). Kim's preferred version of the theory uses mereological sums of worlds – multiverses – rather than tuples. Then the modal fictionalist offers translations like these:

Possibly P iff Acc to PW   there is a one-world multiverse at which
                                   P or
                              there is 2-world multiverse at which P or
                              ..., or
                              P is true unrestrictedly
Necessarily P iff Acc to PW   at all one-world multiverses, P and
                              at all 2-world-multiverses, P and
                              ..., and
                              P is true unrestrictedly

Consider "Necessarily there is a plurality of worlds". This is translated as a conjunction of claims about multiverses. The first is: at all one-world multiverses there is a plurality of worlds. Kim holds that this conjunct fails, since there are no worlds at worlds. Hence (1) in the Brock-Rosen argument comes out false.

The structure of this response is the same as the structure of Noonan's response. "According to PW at all worlds there is a plurality of worlds" comes out false, and so "Necessarily there is a plurality of worlds" comes out false as well. But by the informal argument I gave above, it is *not* false that according to PW at all worlds there is a plurality of worlds. Hence I think Kim's system has the same difficulty that Noonan's defence of the letter of fictionalism has. There are modal claims we want to make. But the system will not let us do that.

2.3. "Strong" fictionalism

There are, of course, indefinitely many ways to develop possible worlds theory, so we don't know at this point that some variation can't work. But so far the prospects are not encouraging.

Rosen (1990, 353-354) contrasts "timid" fictionalism, that is, a theory that aspires to no more than linking modal claims with a fiction, with "strong" fictionalism, which provides truth conditions – meanings – for modal statements in terms of the fiction. Nolan – O'Leary-Hawthorne (1996) argue that a version of the Brock-Rosen argument tells against any "strong reflexive" fictionalism. They recommend that a fictionalist should develop a "timid" rather than a "strong" fictionalism.

Woodward (2008) and Liggins (2008), in slightly different ways, argue for "strong" fictionalism. They propose that we should construe the fictionalist biconditionals as a translation scheme, and, in particular, that we should interpret modal statements in natural language ("Possibly there are blue swans") as meaning the equivalent claims about the fiction ("According to PW, there is a possible world at which there are blue swans"). Liggins proposes additionally a way of treating sentences using the distinctive vocabulary of possible worlds theory that occur in "ordinary" (mainly philosophical) discourse. Modal fictionalism offers

MF p iff Acc to PW, $p*$

where p is a modal sentence (containing "possibly", "necessarily", etc.) and $p*$ is the possible worlds translation. Where p isn't a modal sentence but does contain possible worlds vocabulary Liggins proposes

$MF2$ p iff Acc to PW, p

Thus "there is a possible world at which there are blue swans" means that Acc to PW, there is a possible world at which there are blue swans, which is true. *MF* together with *MF2* give us "reinforced modal fictionalism". (Woodward (2008, 274) endorses a similar idea.)

Woodward's strategy is to construe the fictionalist theory as metalinguistic: the biconditionals state a truth condition for a mentioned sentence by using a sentence in the metalanguage, namely, one about the fiction, thus:

'There could have been talking donkeys' is true iff according to PW, there is a possible world at which there are talking donkeys.

Liggins's proposal is simpler and more direct. Biconditionals are symmetrical: there's nothing special about the left-to-right direction as op-

posed to the right-to-left direction. Hence to capture the idea of fictionalism, a stipulation must be made: which side "has semantic priority". So we stipulate that the story operator side of the biconditional is the one that has semantic priority.[5] There is no need for the metalinguistic formulation.

The upshot for the Brock-Rosen argument is either: (a) The intermediate step in the argument has this syntactic shape: "Necessarily, there is a plurality of worlds". But the conclusion, that there is a plurality of worlds, does not follow, since the proposition this sentence expresses is that according to PW, at all worlds there is a plurality of worlds. Or else (b) the conclusion does follow, but what it means is that according to PW, there is a plurality of worlds, which is unproblematically true.

The upshot for the Nolan/Hawthorne recommendation in favor of "timid" fictionalism is simply that they made a mistake: they didn't notice that, since the strong fictionalist version of the Brock-Rosen argument doesn't mean what it appears to mean, it doesn't conclude validly with the damaging conclusion.

I think there are two serious problems for "strong" modal fictionalism.

Wrong meanings? Does "Possibly there are blue swans" mean that according to PW, there is a possible world at which there are blue swans? I think the answer is fairly clearly "no". We can talk about possibility and necessity without talking at all either about possible worlds or about a fiction about possible worlds (cf. Nolan – O'Leary-Hawthorne 1996, 29-30).

Rosen (1990) may already have been conceiving modal fictionalism as a "strong" modal fictionalism, so this concern about wrong meanings is not specific to Woodward and Liggins. Liggins points out (personal communication) that we could also imagine a position on which the modal fictionalist is offering a new meaning for modal English claims, the meaning ex-

---

[5]   Woodward (2008, 278) complains that Liggins has merely stipulated that the priority is as he says. Indeed, that is just what Liggins says (cf. Liggins 2008, 153). Two comments: (1) Liggins is free to stipulate what he likes. Just as biconditionals are symmetrical, there is nothing about them that prohibits saying that we intend to use them in one way rather than some other. (2) Woodward says that his metalinguistic approach gets around this objection. But his approach is equally stipulative: we could, however implausibly, take the truth-attribution (left hand) side of his biconditionals to have semantic priority: to present the meaning of the fiction attribution on the right hand side.

pressed on the right hand side of the biconditionals, and so there would be no concern about not capturing the old meaning. I think the difficulties facing such a position are substantial, but discussing them would take me outside the scope of this paper.

Loss of problem? The anti-realist about possible worlds wants to say that there are no possible worlds. But given these translation schemes, the sentence, "There are no possible worlds" means that according to PW there are no possible worlds. And that's false. So if we were somehow to manage exceptionlessly to endorse the translation scheme, the anti-realist's signature claim would be false, and false for a reason that isn't relevant to the problem that originally motivated her. Sentences about possible worlds mean something about the fiction of possible worlds, and are more or less unproblematically true (or false). The problem has vanished.

Woodward (2008, 282-286) has two responses to this problem. The first is agnosticism. Ontological commitment, as he sees it, is incurred by assertions (that is, particular speech acts). Given his conception of commitment, "the fictionalist will succeed in [avoiding ontological commitment] just in case she never assents to a sentence, which on that occasion of use expresses a proposition that entails a plurality of worlds" (Woodward 2008, 283). So all she needs to do is make sure never to assert or assent to sentences like "there is exactly one world" or "there is not a plurality of worlds". She should remain agnostic. (Woodward's suggestion here is related to Divers' agnosticism (see Divers 2004). I'll have more to say about Divers' view below, §4.)

Woodward thinks "it would be a mark against paraphrastic fictionalism if it committed the fictionalist to agnosticism" (Woodward 2008, 283). The second response is to bring back a way to state the problem. We can hold that the translations are not always in effect. Instead, whether they are in effect or not depends on context. When we are engaging in modal reasoning without any special concerns about ontology, the translations are in effect. When the fictionalist wishes to express her ontological commitments, the context has changed, and there, the sentence "There are no possible worlds" means exactly what it appears to mean.

But then can't we express the original Brock-Rosen argument in a context in which the sentences get no special semantic treatment? Woodward offers an intriguing argument that there is no single context in which (2),

"Necessarily, there is a plurality of worlds iff according to PW, at all possible worlds, there is a plurality of worlds," gets the interpretation it needs for the Brock-Rosen argument (cf. Woodward 2008, 286). I think this is incorrect. The "standard" context for evaluating (2) is the one in which all the words mean what they appear to mean. In that context, (2) gets a stable interpretation, precisely the one it needs for the Brock-Rosen argument.

Liggins has a different response. Again, the anti-realist wants to assert "There is not a plurality of worlds". She now confronts this argument (157):

> There is not a plurality of worlds.
> Possibly, there is not a plurality of worlds. (Standard modal logic)
> Possibly, there is not a plurality of worlds iff according to PW, there is a possible world at which there is not a plurality of worlds. (Instance of *MF*)
> Therefore, according to PW, there is a possible world at which there is not a plurality of worlds. (Modus ponens)

and that conclusion is false.

Liggins offers two replies for the anti-realist. First, she does not have to assert the premise of this argument; she can "sit on the fence", that is, not say anything positive or negative about the existence of possible worlds. Second, she can use Divers's proposal about what "Possibly ... worlds ..." should mean (see above, §2.2.2.). This takes such sentences to their unmodalized versions. So the argument collapses: "Possibly, there is not a plurality of worlds" only entails the starting point, "There is not a plurality of worlds".

I have already registered my doubts about Divers's proposal for handling "advanced modalizing" (§2.2.2. above). "Sitting on the fence" doesn't allow the anti-realist to preserve the problem, since she's required to not assert just those sentences that articulate it. It would be preferable to have a position on which we get to say what we think is true, and problematic.

2.4. Pragmatic modal fictionalism

Liggins (2008) outlines another way out of the Brock-Rosen argument, suggested by a remark Rosen makes about fictionalism in mathematics. Liggins calls this "pragmatic modal fictionalism".

Pragmatic modal fictionalism suggests that we shouldn't decide what someone's ontology is by looking at the literal meaning of the things she says. We should look instead to her considered beliefs. The pragmatic modal fictionalist denies that there is a world at which there are zombies, because she denies that there are worlds: she thinks the truth is that there is not a plurality of worlds. But she uses this sentence to express a different belief: plausibly, the belief that according to PW, there is a possible world at which there are zombies.

Consider a pragmatic account of truth in fiction. We might be inclined to say that "Sherlock Holmes lived at 221B Baker Street" is true. The obvious story operator account would say that this sentence means "According to the Sherlock Holmes stories, Sherlock Holmes lived at 221B Baker Street" and so it comes out true. But we might for various reasons want to resist this account. We could hold that in fact "Sherlock Holmes lived at 221B Baker Street" is false. There is no such person, and there is no such place. When someone utters this sentence we can understand them to be expressing the proposition about the fiction, even though the sentence does not mean anything about the fiction. The context and the speaker's communicative intentions together determine that the proposition the speaker communicates is not the proposition that is literally expressed by the utterance.

The pragmatic modal fictionalist holds that "Necessarily, there is a plurality of worlds" is false. Sometimes she uses it to express the proposition that according to PW, at all possible worlds there is a plurality of worlds. Her evaluation of the critical premise in the Brock-Rosen argument

(2)   Necessarily, there is a plurality of worlds iff according to PW, at all possible worlds there is a plurality of worlds

is that the left hand side is false, the right hand side is true, and hence (2) is false. So the Brock-Rosen argument is unsound.

This is not something the "strong" modal fictionalist can say. Saying that (2) is false is saying that it does not state a translation. That is abandoning the whole idea of "strong" modal fictionalism. Similarly, *no* modal fictionalist who thinks of (2) as an instance of the translation schema can deny (2) on the ground that one side is true and the other false.

Pragmatic modal fictionalism denies what I am considering to be a hallmark of modal fictionalism or story operator fictionalism: the biconditional schema provides equivalences or translations between sentences in

natural language or quantified modal logic and possible worlds theory. The reason for denying it is that there doesn't seem to be a good way to hold on to it and avoid commitment to the Brock-Rosen argument.

There are, however, two worrying things about pragmatic modal fictionalism. First: both Rosen and Liggins express the idea in the context of modal fictionalism and the biconditional equivalence schema. Although it is clear enough that they aren't using it as the original modal fictionalist uses it, it is not completely clear in their presentations how far away from the original they have moved. Second: the distinction between semantics and pragmatics in philosophy of language and linguistics is rich, complex and controversial. So far, the pragmatic modal fictionalist has only said that we can say things about possible worlds and mean something else. We will need pragmatic modal fictionalism to say more than "interpret the fictionalist in such a way that she isn't committed to the ontology she appears to be committed to." So a great deal remains to be spelled out.

3. Instrumentalism

Instrumentalism is an alternative to modal fictionalism, and to pragmatic modal fictionalism. It construes the relation between what we say about modality and counterpart theory in a different way than they do, and provides, I think, a clearer account of what we are doing when we use possible worlds talk.

An *instrument* is an object that is used for some purpose or goal, a tool or an implement, a device, utensil, apparatus, contrivance, a gadget. Since being an instrument is a functional characteristic, pretty much anything can be an instrument. A theory can be an instrument. For my purposes here it shouldn't matter a great deal exactly what kind of object a theory is: a set of sentences, a set of models, something more vague like "a story". Theories taken literally are truth-apt: they say something about the world that may be true or false. For a theory to be used as an instrument we have to somehow ignore that truth-aptness. The theory is used in some way to do something, but that use ignores, and is not sensitive to, the truth or reference of the sentences in the theory.

There are several ways we could use a theory without regard to its truth. The most drastic would be to treat the theory as an uninterpreted

calculus of symbols. Less drastic, we could count the symbols as carrying their standard meanings, but hold that the theory is entirely false: its domain of quantification is empty, its predicates have no extensions. We could hold that the ontology of the theory is unproblematic but that much of what the theory says about it, or certain important things the theory says about that ontology, are false. We could also be more cautious and say that we don't know (or don't care) whether the theory is true or not (whether its subject matter is real or not). For all we know it may be, but our purposes are served whether it is true or not.

Let me call these four kinds of instrumentalism "syntactic," "metaphysical," "ideological" and "epistemic" respectively. Here are some traditional examples.

Hume's skepticism about causation is a version of metaphysical instrumentalism. He argued that we have no impression or experience of necessary connection, and that therefore causation is not real. But experience of certain kinds of regularity in the world leads the mind to connect experience of one of the correlated events with the other. We experience this connection (in the mind) as the connection in the world. In this case the ontology of the causation theory is entirely false. But there is a point to talking about causes: it codifies information about what has gone together with what, and hence serves as a guide to what to do next.

Hume *could* be construed as a syntactic instrumentalist. Since we have no impression of necessary connection, Hume thinks we also have no idea of necessary connection. If we think of the meanings of our words as grounded in the ideas we express using them, then our word "cause" has no meaning. We do find ourselves using it in a definite and constrained way in a way important to our lives.

Osiander wrote in his preface to Copernicus' *De Revolutionibus*:

> Since he cannot in any way attain to the true causes, he will adopt whatever suppositions enable the motions to be computed correctly from the principles of geometry for the future as well as for the past. (Copernicus 1992, xvi)

Copernicus' ontology was traditional and conservative: there is a sun, and there are seven planets, and there are celestial spheres connected with the various astronomical objects. His radical proposal was to switch the places of the sun and the Earth. Taken literally, Osiander's recommenda-

tion is to be an epistemic instrumentalist, since he says the true causes cannot be known at all. It is at least as plausible to read him as an ideological instrumentalist: for various reasons, including political ones, it was very difficult to believe in 1543 that the earth moves.

van Fraassen's "constructive empiricism" is an epistemic instrumentalism (see van Fraassen 1980). van Fraassen argued that (a) the only reason we have for belief in unobservables is inference to the best explanation, and that (b) inference to the best explanation is not a truth-guaranteeing form of reasoning, and hence not rationally compelling. We do not have rationally compelling reason to *believe* in unobservables like subatomic particles. But we do have reason to *accept* theories that talk about them, since those are the best theories we have about how the world works.

What, then, should instrumentalism about modality (or modal instrumentalism) look like?

First, it should not be a syntactic instrumentalism. I argued against Woodward above (§2.3.) that in order to articulate the philosophical problems, and to motivate any kind of fictionalism, we need to be able to say what it is that we are finding problematic. So we need to hang on to the distinctive vocabulary of possible worlds theory. The modal instrumentalist can "calculate" with PW as if it were an uninterpreted calculus, but it does still carry its standard meaning.

Second, it shouldn't be an "ideological" instrumentalism. The metaphysical worry about possible worlds isn't that we're confident that there are some and that what we say about them is false. It's rather that we have serious doubts about whether there are possible worlds.

So we want either a metaphysical instrumentalism on which there are no possible worlds at all, or an epistemic instrumentalism, according to which we do not or cannot know whether there are possible worlds. We don't need to make this choice here: either version works as well as an alternative to modal fictionalism.

The modal instrumentalist thinks that "there could have been talking donkeys" is clearly true and that "there is a possible world that has talking donkeys as parts" is either obviously false or there is real doubt about whether it is true. So there is good reason to think those two sentences are not equivalent. It would be misleading to express the modal instrumentalist's idea as a biconditional schema linking them.

Lewis provides detailed guidance for how to "translate" sentences of quantified modal logic into counterpart theory. Instead of thinking of this process as translation, let us think of it as pairing. Lewis gives us an algorithm for a pairing procedure. In this case the algorithm just gives us the pair ⟨"there could have been talking donkeys", "there is a possible world that has talking donkeys as parts"⟩.

How do we use the instrument? It can serve to validate modal reasoning. We might think that

(A) Necessarily p
therefore, (B) Possibly p

is a valid argument. The pairing procedure gives us

(A*) At all worlds p
therefore, (B*) there is a world at which p

which is clearly valid by universal instantiation. Generally, we can use Divers' "safety result" (cf. Divers 1999b): assuming that A* and B* are the possible worlds translations of A and B, then, necessarily, if B* is a consequence of A*, then B is a consequence of A. We can clearly use the instrument to reason: start with some premises A, A', ..., use the pairing procedure to generate their possible worlds images A*, A'*, ..., reason using counterpart theory to various conclusions B*, B'*, ..., then conclude B, B', ...

In the course of a discussion of physicalism and the philosophy of mind we might find ourselves saying "There is a possible world at which there are zombies". This sentence uses the distinctive vocabulary of PW and so (according to the modal instrumentalist) is a part of the instrument. If her instrumentalism is metaphysical, she will claim that it is false; if she is more cautious, she will say that she doesn't know whether it is true or false. The pairing procedure pairs this sentence with the sentence of modal English "zombies are possible". It doesn't mean the same as the possible worlds version, but for various purposes the possible worlds version may be more useful to reason with.

Let's assemble the ingredients for responding to the Brock-Rosen argument. The original argument went like this:

(1)   According to PW, at all possible worlds there is a plurality of worlds

(2)   Necessarily, there is a plurality of worlds iff according to PW, at all possible worlds there is a plurality of worlds

(3)   Therefore, necessarily, there is a plurality of worlds

(4)   Therefore, there is a plurality of worlds

Since the modal instrumentalist expresses her commitment to the falsity (or dubiety) of PW "outside" of her use of PW, there is no need for the story operator. She denies that natural language modal sentences are equivalent to PW sentences. She will write down the results of the pairing procedure rather than a biconditional like (2). So she has:

(1)   At all possible worlds there is a plurality of worlds

(2_i)   ⟨Necessarily, there is a plurality of worlds, at all possible worlds there is a plurality of worlds⟩

(3)   Therefore, necessarily, there is a plurality of worlds

(4)   Therefore, there is a plurality of worlds

According to modal instrumentalism, sentences (1), (3), (4) and the two members of (2_i) are all part of the instrument, since they make positive assertions about possible worlds. They may be false, or their truth may be unknown or unknowable; at any rate, they are not known to be true. (1)-(4) is not an argument at all, since (2_i) is not a sentence. So it is not a valid argument. So the modal instrumentalist has nothing left that would commit her to the truth of (4).

4. Some nearby views

In this section I discuss four views that are particularly near to modal instrumentalism.

Other versions of instrumentalism. Forbes (1983, 271) holds that according to instrumentalism "sentences of [the instrumental theory] are said not to have literal meaning". Forbes thus conceives of at least an important part of the instrumental theory as syntactic in my sense and hence not truth-apt. According to Rosen (2005, 14) and Nolan (2002, 26-27) instrumentalism is a version of noncognitivism: claims about possible worlds are not truth-apt.

The instrumentalism I am proposing here is not syntactic and it is not a version of noncognitivism. I am proposing that possible worlds theory is truth-apt and that its truth conditions are "realist" in the sense that they are about a mind-independent reality.

Nolan's instrumentalist has a device for turning "acceptable" claims in possible worlds theory into truths, much as the modal fictionalist has the story operator to turn acceptable claims in the fiction into truths about the fiction. My version of instrumentalism does not do this. It only provides a mechanism (the pairing procedure) for connecting claims in modal natural language or quantified modal logic with sentences in possible worlds theory.

Pragmatic modal fictionalism. Instrumentalism and pragmatic modal fictionalism are similar in important respects: both assess the truth and falsity of claims about possible worlds independently of the biconditional equivalence schema, and both reject the Brock-Rosen argument as (at least) unsound. They agree on what Rosen counts as the critical fictionalist move: they count PW as acceptable but they do not count it as true.

They differ most clearly in what they say about the biconditional equivalence schema: pragmatic modal fictionalism accepts it and then holds that the instances of the schema are mostly false. Instrumentalism holds that this is just the wrong way to use the false theory: it shouldn't be thought of as providing equivalences at all.

We can capture the difference in a more general way by asking for their answers to the following "structural" question: if PW is in some broad sense to be thought of as a fiction, how is it related to our serious modal discourse? (Compare the question: if there are no numbers, how is talk of number theory related to our serious numerical discourse?) Modal fictionalism (story operator fictionalism) says: our serious discourse is about a fiction. Strong modal fictionalism says our serious discourse literally means claims about a fiction. Pragmatic modal fictionalism expresses what we accept in the same way the modal fictionalist does, in terms of the biconditional equivalence schema using the story operator. Pragmatic modal fictionalism then says that pragmatic factors (as well as semantics) settle what proposition someone expresses when she uses a sentence like "There is a possible world at which there are zombies". Since she believes it is false, she should be taken to be expressing a proposition about the

fiction: that according to PW there is a possible world at which there are zombies.

Instrumentalism suggests that there is a simpler and clearer way here. Let us take PW, and sentences in English using the vocabulary of PW, as instruments. The sentences mean what they appear to mean, and we judge their truth and falsity (or their knowability) by our prior assessment of the truth of the basic claim that there are possible worlds other than the actual world. We don't have to say that they are fictional in order to express how we use them in English. We don't need to make any claims that sentences outside of PW mean the same as sentences in PW. We don't need to interpret what someone says as meaning something other than what it appears to mean.

Agnosticism. An agnostic about possible worlds (the "worldly agnostic") withholds belief in them. (Divers 2004) shows that the agnostic can endorse a surprising number of claims from possible worlds theory. For example, she can assent to sentences of the form "there are no possible worlds ..." and these can give her necessity claims, since "all possible worlds are such that φ" is equivalent to "there is no possible world at which not-φ". There are some modal beliefs she cannot share with the realist about possible worlds ("there might have been no donkeys"; cf. Divers 2004, 675). But it may turn out that these are rationally dispensable: in the cases where the realist reasons with a belief that the worldly agnostic cannot hold, there may be a plausible substitute that the worldly agnostic can hold.

One of Divers' motivations for articulating agnosticism is that the main anti-realist alternatives (story operator fictionalism, Forbes's fictionalism, Blackburn's expressivism) have serious problems (see Divers 2004, 662-667): they "harbour unanalysed modal content", they have the Brock-Rosen problem, they are not adequate to our inferential practice with modality. Divers thinks agnosticism is the best remaining alternative.

My suggestion is that instrumentalism is at least as good an alternative. *One* way unanalysed modal content finds its way into modal fictionalism is through the story operator, since there doesn't seem to be any way to understand "According to PW" except in terms of a strong conditional (cf. Hale 1995a; Divers – Hagen 2006). Instrumentalism doesn't use the story operator, so it does not inherit that set of problems for modal fictionalism. Instrumentalism does not, as I've argued, have the Brock-Rosen problem.

Modal reasoning according to instrumentalism is exactly what possible worlds theory says it is. Finally, even epistemic (or agnostic) instrumentalism continues to reason with all of what the modal realist reasons with, and so does not need to wrestle with what the worldly agnostic can consistently believe about possible worlds.

Amodalism. Cowling (2011) discusses *amodalism*: denying "modal ubiquity", the claim that every proposition has a modal status (cf. Divers 1999a). Some truths are just true *simpliciter*, neither contingent nor necessary. Amodalism has some striking benefits. Some of the arguments *against* possible worlds depend on modal ubiquity as a premise. Cowling recommends that "defenders of possible-worlds theory ought to accept amodalism" (Cowling 2011, 473).

Amodalism would help with the Brock-Rosen argument. It would provide a reason to reject the equivalence "Necessarily there is a plurality of worlds iff Acc to PW at every world there is a plurality of worlds". It would also help with "Hale's dilemma" for modal fictionalism (Hale 1995b), since that dilemma starts by asking whether PW, according to the fictionalist, is contingently or necessarily false. In both cases amodalism denies that PW has any modal status.

Amodalism shares with instrumentalism a non-standard treatment of truth and modality: amodalism says some theories have no modal status, instrumentalism says that the truth of some theories is irrelevant. Amodalism is consistent with instrumentalism but it is not entailed by it. (Amodalism is also consistent with realism about possible worlds.) It seems like a natural addition to instrumentalism: if PW is simply an instrument, then the question of its modal status seems misguided. I think amodalism and instrumentalism could go well together, but the details will have to wait for a different paper than this one.

5. Conclusion

Modal fictionalism (story operator fictionalism) is an appealing way to use possible worlds theory without commitment to its extraordinarily counterintuitive ontology. But the Brock-Rosen argument against modal fictionalism is resilient. There are several strategies for responding to it; I've argued that none works particularly well. I think the core difficulty

comes from the biconditional equivalence schema. Instrumentalism, which does not use that schema, has the resources to respond to the Brock-Rosen argument. So I think the agnostic or anti-realist would be better off being an instrumentalist about possible worlds theory.[6]

References

BRICKER, P. (2001): Island Universes and the Analysis of Modality. In: Preyer, G. – Siebelt, F. (eds.): *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis.* Oxford: Rowman & Littlefield, 27-55.

BROCK, S. (1993): Modal Fictionalism: A Response to Rosen. *Mind* 102, No. 405, 147-150.

COPERNICUS, N. (1992): *On the Revolutions.* Trans. Edward Rosen. Baltimore: The Johns Hopkins University Press.

COWLING, S. (2011): The Limits of Modality. *The Philosophical Quarterly* 61, No. 244, 473-495.

DIVERS, J. (1999a): A Genuine Modal Realist Theory of Advanced Modalizing. *Mind* 108, 217-239.

DIVERS, J. (1999b): A Modal Fictionalist Result. *Noûs* 33, 317-346.

DIVERS, J. (2004): Agnosticism about Other Worlds: A New Antirealist Programme in Modality. *Philosophy and Phenomenological Research* 69, No. 3, 660-685.

DIVERS, J. – HAGEN, J. (2006): The Modal Fictionalist Predicament. In: MacBride, F. (ed.): *Identity and Modality.* Oxford: Oxford University Press, 57-73.

FORBES, G. (1983): Physicalism, Instrumentalism and the Semantics of Modal Logic. *Journal of Philosophical Logic* 12, No. 3, 271-298.

HALE, B. (1995a): A Desperate Fix. *Analysis* 55, No. 2, 74-81.

HALE, B. (1995b): Modal Fictionalism: A Simple Dilemma. *Analysis* 55, No. 2, 63-67.

KIM, S. (2002): Modal Fictionalism Generalized and Defended. *Philosophical Studies* 111, 121-146.

LEWIS, D. (1968): Counterpart Theory and Quantified Modal Logic. *The Journal of Philosophy* 65, No. 5, 113-126.

LEWIS, D. (1986): *On the Plurality of Worlds.* Oxford: Basil Blackwell.

LIGGINS, D. (2008): Modal Fictionalism and Possible-Worlds Discourse. *Philosophical Studies* 138, 151-160.

NOLAN, D. (2011): Modal Fictionalism. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy.* Winter 2011 ed., available at: http://plato.stanford.edu/entries/fictionalism-modal/.

Nolan, D. – O'Leary-Hawthorne, J. (1996): Reflexive Fictionalisms. *Analysis* 56, No. 1, 23-32.

Nolan, D. (2002): *Topics in the Philosophy of Possible Worlds*. New York – London: Routledge.

Noonan, H. W. (1994): In Defence of the Letter of Fictionalism. *Analysis* 54, No. 3, 133-139.

Rosen, G. (1990): Modal Fictionalism. *Mind* 99, 327-354.

Rosen, G. (1993): A Problem for Fictionalism about Possible Worlds. *Analysis* 53, No. 2, 71-81.

Rosen, G. (2005): Problems in the History of Fictionalism. In: Kalderon, M. E. (ed.): *Fictionalism in Metaphysics*. Oxford: Oxford University Press, 14-64.

van Fraassen, B. (1980): *The Scientific Image*. Oxford: Clarendon Press.

Woodward, R. (2008): Why Modal Fictionalism Is Not Self-Defeating. *Philosophical Studies* 139, 273-288.

An Argument for Authorial Creation[1]

ZSÓFIA ZVOLENSZKY

Department of Logic. Institute of Philosophy. Eötvös University
Múzeum krt. 4/i. H-1518 Budapest. Hungary
zvolenszky@elte.hu

ABSTRACT: Artifactualism about fictional characters, positing Harry Potter as an abstract artifact created by J. K. Rowling, has been criticized on the grounds that the idea of creating such objects is mysterious and problematic. In the light of such qualms, it is worth homing in on an argument in favor of artifactualism, showing that it is the best way to include the likes of Harry Potter in our ontology precisely because it incorporates authorial creation. To that end, I will be exploring Kripke's fleeting remarks in the Addenda to his "Naming and Necessity" lectures about expressions like 'unicorn' and 'Harry Potter'. Elsewhere, Kripke motivates artifactualism by suggesting that incorporating authorial creation (as artifactualism does) is a move that is intuitive and natural; but beyond this, he doesn't provide any *arguments* in favor of such a move. My purpose in this paper is to construct such an argument based on considerations about Kripke's general view about proper names, in particular, his seminal causal-historical chain account of reference determination, and its consequences for fictional names as well as nonfictional names without bearers such as 'Vulcan'.

KEYWORDS: Abstract artifacts – Kripke – proper names – realism about fictional characters – causal-historical chain theory of reference determination – semantics of fictional discourse.

---

[1]   Shorter predecessors of this paper are Zvolenszky (2012, Section 2; 2014).

1. Introduction

Artifactualism about fictional characters (*artifactualism*, for short) positing Harry Potter as an abstract artifact created by J. K. Rowling, takes (1) at face value:

(1)   Harry Potter was created by J. K. Rowling.

Like other forms of *realism* about fictional characters, artifactualism posits an ontology that includes the likes of Harry Potter. But realism is not our only option; we could also accept an irrealist analysis of (1) that doesn't take it at face value: "J. K. Rowling wrote a body of fiction in which Harry Potter is a specific character". For the purposes of this paper, I set aside the irrealist alternative and focus on artifactualism and its major realist rivals none of which take (1) at face value. Does the fact that artifactualism – singularly among realists views – incorporates authorial creation constitute an advantage or a disadvantage?

An advantage, say I. In this paper, I will argue that if we accept a certain widely held theory about proper names proposed by Saul Kripke (1972/ 1980) – to wit, that their reference is determined by a causal-historical chain of uses leading back to the introduction of the name – then we have reason to choose artifactualism over its major realist rivals *precisely because it incorporates authorial creation*.

Meanwhile, several philosophers have had serious qualms about authorial creation and taking (1) at face value. Takashi Yagisawa (2001, 154) argues that the most influential creationist views (by John Searle and Peter van Inwagen) "are ultimately unsuccessful in establishing creationism"; more generally, he claims that no view on which fictional characters exist can do justice to our intuition that a claim like "Harry Potter doesn't exist" is true and is entailed by the true "Harry Potter is a fictional character".

Stuart Brock sets out to

...explain why creationism about fictional characters [the view that fictional characters exist by being created by their author(s)] is an abject failure. It suffers from the same problem as theological creationism: the purported explanation is more mysterious than the data it seeks to explain. (Brock 2010, 338)

Brock's charge in a nutshell: creationism fails to provide a satisfactory account of the spatial and temporal dimensions of fictional characters, for example, their moment of creation. Brock (2010, 340-342) also complains that beyond the intuitive appeal of taking (1) at face value, arguments in favor of incorporating authorial creation in theories about the metaphysics of fictional characters are "almost completely lacking".[2] In the light of such doubts and paucity of motivation, it is worth homing in on an argument in favor of artifactualism (a form of creationism), showing that it is the best form of realism one could adopt. The goal of this paper is to expound such an argument.

1.1. Authorial creation and theories of fictional characters

First, let's take stock of the various realist positions. We may, along with Mark Sainsbury (2010, 44-114), distinguish three realist alternatives about fictional characters: there really are such things just as there are ordinary concrete objects occupying space and time; but unlike those ordinary objects like cups, saucers and the Big Ben, …

- fictional characters *don't exist*, according to *Meinongianism* about fictional characters;[3]
- fictional characters *are not actual* but merely possible, according to *nonactualism*;[4] and

---

[2]   More precisely, Brock remarks that "arguments in support of the fundamental thesis are almost completely lacking", where by 'fundamental thesis', he means the following: "Fictional characters, to the extent that there are any, are genuinely created by the authors of the works in which their names (or designating descriptions) first appear" (Brock 2010, 340, 342).

[3]   For brevity's sake, I'll suppress the qualification 'about fictional characters' and will simply talk of realism, irrealism, Meinongianism, nonactualism, artifactualism, Platonism. Whenever these labels appear unqualified, they are shorthand for theories about fictional characters. Parsons (1980) is a contemporary proponent of Alexius Meinong's (1904) eponymous theory.

[4]   Lewis (1978) put forth such a view. This position is sometimes called possibilism about fictional characters. See also Kripke's earlier (1963) view about Sherlock Holmes. Importantly: for nonactualism to provide a distinctive alternative, it has to commit to a Lewisian (1986) metaphysics of possible worlds: modal realism (also called extreme realism), argues Sainsbury (2010, 74, 222, fn. 8).

- fictional characters *are not concrete* but abstract, created by the activities of authors according to *artifactualism*.[5]

How might the various forms of realism handle (1)? On this point, artifactualism appears to show a clear edge relative to its two rivals. A negligible point of advantage is that according to neither rival theories is Potter *created* – going from nonexistent to existent. According to the Meinongian, Potter isn't created – brought into existence – because he doesn't exist (he just is). And according to the nonactualist, Potter had existed all along as a merely possible object and continues to exist as a merely possible object after the novels are written. According to Sainsbury (2010, 61-63, 82-85), the real advantage of artifactualism concerns its response to the so-called *selection problem*: upon introducing the name 'Harry Potter' in her novel, how does J. K. Rowling manage to select one rather than another among the countless candidate objects? According to Meinongianism, there are countless nonexistent candidates; according to nonactualism, there are countless merely possible, nonactual candidates.[6] Sainsbury (2010, 63) doesn't see "how a Meinongian can offer any sensible account of how an author's or reader's thoughts are supposed to engage with one rather than another nonexistent entity".

---

[5]   Kripke (1973/2013), Searle (1979), van Inwagen (1977), Fine (1982), Schiffer (1996), Salmon (1998), Thomasson (1999) are prominent proponents of artifactualism who hold that authors' creative process of writing novels, stories, etc. creates fictional characters. This position is sometimes called creationism about fictional characters.

There is a position in logical space for holding that fictional characters are abstract but exist timelessly, and authors don't create but discover them – we might call such a view Platonism about fictional characters. Zalta's (1983) unorthodox neo-*Meinongian* proposal as well as Wolterstorff's (1980) theory can be considered instances of such an account. The only kind of abstract-object theory I will consider in this paper is artifactualism, given the overwhelming popularity and attention that this position has been enjoying (compared to Platonism), as well as the advantages that I think it has over rival theories (Platonism included) precisely because it treats fictional characters as human-created objects. The arguments expounded here carry over to Platonism also, but I will relegate discussion of that to footnotes.

[6]   Ultimately, Sainsbury (2010) rejects artifactualism in favor of irrealism. For a response strategy that the artifactualist can adopt to fend off Sainsbury's criticism, see Zvolenszky (2013, an earlier version of which appears in 2012).

We are about to see that a far more decisive objection than the selection problem emerges against the Meinongian once we consider the difficulties that the *nonactualist* encounters when it comes to a different set of problems that Kripke (1972/1980) raised.

1.2. Naming and Necessity *on fictional characters*

Throughout this paper, I am accepting Saul Kripke's (1972/1980) proposal about causal-historical chains determining the reference of proper names, exploring what follows from it: that artifactualism has an advantage relative to its major realist rivals.

The core of Kripke's position (from the second lecture of *Naming and Necessity*) about what does and doesn't determine the reference of proper names like 'J. K. Rowling' and 'London' (which refer to concrete objects) can be summarized with the following two claims:

- *Qualitative fit is neither necessary nor sufficient for being the referent of a name.* Suppose individual speakers who competently use a name N associate various descriptions with N. Kripke's claim: to be the referent of N, it is neither necessary nor sufficient that the referent be the unique individual fitting the associated descriptions (or fitting the weighted majority of the descriptions). Call this the *qualitative-fit claim*.
- *A causal-historical connection is necessary for reference.* Competent N users refer to an object o by using N only if there is a causal-historical chain of uses of N in their linguistic community leading back to the introduction of N as a name for o. Call this the *historical connection requirement*.[7]

What does Kripke say about fictional names that don't refer to concrete objects? In the "Addenda" to his "Naming and Necessity" lectures, Kripke

---

[7]   I'm not including here the corresponding sufficiency claim: that a causal-historical chain of uses leading back to an object being given the name is *sufficient* for it to be the name's bearer. In the light of considerations about 'Santa Claus', and 'Napoleon' introduced as a name for a pet (and later, on, also examples like 'Madagascar') indicate that much more elaboration and complexity lies ahead before we get a *sufficient* condition for being the referent of a name. And the fact that Kripke (1972/1980, 93, 96-97) was pointing out such examples makes it clear that he was aware of the additional complexity required while he was delivering the lectures, so Evans' (1973) charge that Kripke's sufficiency claim is unwarranted is itself unwarranted.

(1972/1980, 156-157) motivates two theses for expressions like 'unicorn' and 'Harry Potter':

- The *metaphysical thesis*: There is no basis for counting any *merely possible* object as Harry Potter, Sherlock Holmes, a unicorn, etc.
- The *epistemological thesis*: There is no basis for counting any *actual* object as Harry Potter, Sherlock Holmes, a unicorn, etc.

In the metaphysical thesis, Kripke's target seems to be the nonactualist (given that he is talking about merely possible entities, the nonactualist's candidates for fictional characters). At the end of the paper, we will see, however, that both theses bear on Meinongianism also. Along the way, we will also see (in Section 2) that the two arguments are at root intimately connected.

Elsewhere (see Kripke 1973/2011, 1973/2013), Kripke motivates artifactualism by suggesting that incorporating authorial creation (as artifactualism does) is a move that is intuitive and natural; but beyond this, he doesn't provide any *arguments* in favor of such a move.[8] My purpose in this

---

[8]   The following are telling passages from Kripke in which authorial creation and artifactualism receive motivation from intuition as well as a recurring analogy with how people's activities create nations (this is the only kind of motivation for artifactualism that we find in these two works of Kripke's in which he is focusing specifically on fictional names and other names that lack a concrete referent):

> On my view, to write a novel is, ordinarily, to create several fictional characters, as Twain, by writing *Huckleberry Finn*, brought both a novel and a fictional character into being. It is not that fictional characters exist in one sense but not in another. The fictional character Huckleberry Finn definitely exists, just as the novel does: I would withdraw the statement only if my impression that there was any real novel was mistaken. Thus, their existence is not like that of numbers, abstract entities which are said to necessarily exist, independently of empirical facts. ... *A fictional character, then, is an abstract entity. It exists in virtue of more concrete activities of telling stories, writing plays, writing novels, and so on, under criteria which I won't try to state precisely, but which should have their own obvious intuitive character. It is an abstract entity which exists in virtue of more concrete activities the same way that a nation is an abstract entity which exists in virtue of concrete relations between people.* A particular statement about a nation might be analyzable out in virtue of a more complicated one about the activities of people, or it might not: it might be hard, or maybe, because of problems of open texture, impossible to do so. But, at any rate, the statement about the nation is true in virtue of, and solely in virtue of, the activities of

paper (in Section 3) is to construct such an argument based on considerations about Kripke's general view about the reference of proper names.

2. The intertwining background of the two Kripkean theses

Behind Kripke's metaphysical thesis is what we might call the *insufficient-specificity problem*.[9] The Harry Potter novels specify many details about Harry; but they also leave a lot of other details unspecified, for example, which of various parental cells Harry came from. Due to such lack of specificity in the novels, we have no basis for deciding between two distinct merely possible candidates (they originate from distinct zygotes, say) that are just like Harry is described in the novels, which of them is Harry Potter. (The insufficient specificity problem is different from the selection problem, which (among other things) is about how authors select from countless candidates the one they *set out* to write about.)

The epistemological thesis turns out to generate an even deeper problem for the nonactualist, one that we shall see (at the end of the paper) affects the Meinongian also. Behind the epistemological thesis is what we might call the *coincidental-resemblance problem*, which Kripke discusses in connection with the mythical species of unicorn:

---

the people. [footnote omitted] I hold the same thing to be true of fictional characters. (Kripke 1973/2013, 72-73, emphasis added)

It is important to see that fictional characters so called are not shadowy possible people. The question of their existence is a question about the actual world. It depends on whether certain works have actually been written, certain stories in fiction have actually been told. *The fictional character can be regarded as an abstract entity which exists in virtue of the activities of human beings, in the same way that nations are abstract entities which exist in virtue of the activities of human beings and their interrelations.* [footnote omitted] A nation exists if certain conditions are true about human beings and their relations; it may not be reducible to them because we cannot spell them out exactly (or, perhaps, without circularity). Similarly, a fictional character exists if human beings have done certain things, namely, created certain works of fiction and the characters in them. (Kripke 1973/2011, 63, emphasis added)

[9]   Kaplan also emphasizes insufficient specificity as an obstacle to naming nonexistents (see Kaplan 1973, 506; 1989, 609).

...the mere discovery of animals with the properties attributed to unicorns in the myth would be no means to show that these were the animals the myth was about: perhaps the myth was spun out of whole cloth and the fact that animals with the same appearance actually existed was mere coincidence. In that case, we cannot say that the unicorns of the myth really existed; we must also establish a historical connection that shows that the myth is *about* these animals. (Kripke 1972/1980, 157, emphasis in the original)

Kripke is making two points here: even if we find animals qualitatively like the unicorns of the myth, that wouldn't justify counting them as unicorns given (i) the lack of historical connection between the newly found species and the use of the expression 'unicorn'; and given that (ii) the unicorn myth was "spun out of whole cloth", not created in the right way, to make the term apply to the newly found species. The upshot of (i) and (ii): we would have no more than mere qualitative coincidence between unicorns as described in the myth and the actual species discovered. And for a proper name, reference takes more than coincidental resemblance, so we don't have any candidate actual objects to count as unicorns.[10]

Notice that the idea that coincidental resemblance is *a problematic, insufficient* basis for reference simply echoes one half of Kripke's views about what does and doesn't determine the reference of proper names like 'J. K. Rowling': the qualitative fit claim. Meanwhile, (i), about historical unconnectedness being a problem, echoes the other half of Kripke's claim; the historical connection requirement. We see then that two of the three interconnected problems underlying the epistemological thesis directly rely on Kripke's general claims about proper name reference.

In the case of the expression 'unicorn', the coincidental-resemblance problem thus arises as a result of two distinct problems: (i) *historical unconnectedness* and (ii) *unsuited mode of introduction*. Pure myth-making mode and pure fiction-writing mode both give rise to expressions that aren't introduced in the right way to refer to actual objects.

Notice that so far, the two Kripkean theses, as I formulated them, were about Sherlock Holmes and unicorns, not the *reference of expressions* like 'Sherlock Holmes' and 'unicorn', with which (i) and (ii) are concerned. There is a crucial difference to be drawn here (see Braun 2005): even if one

---

[10]   Kaplan quotes Harry Deutsch: "reference is no coincidence" (Kaplan 1989, 608).

agrees with the artifactualist that Sir Arthur Conan Doyle created the abstract artifact that is Sherlock Holmes, from that it does not follow that *any* uses of the name 'Sherlock Holmes' *refer* to that artifact. Indeed, influential artifactualists like Kripke, Searle, Thomasson and van Inwagen agree (against Salmon 1998) that only *certain* instances of discourse about fiction contain names that refer to abstract artifacts: for example, creation sentences like (1) and sentences like (2) below feature proper names as referring to abstract artifacts, but the sentences created during authors' fiction-making activity (call these instances of *textual discourse*[11]) involve no such reference.

(2)   Sherlock Holmes is a fictional character.

Then a weaker general claim that all forms of artifactualism are committed to is this: On at least some uses – uses like (1) and (2), call these the *artifactualists' focal uses, which exemplify what we might call metatextual discourse*[12] – fictional names refer to abstract artifacts, and whenever fictional names refer to something, they refer to abstract artifacts. Parallel commitments can be constructed for the Meinongian as well as the nonactualist, respectively: on at least some uses, namely, the artifactualists' focal uses, fictional names refer to Meinongian objects, and whenever fictional names refer to something, they refer to Meinongian objects; also, on at least some uses, namely, the artifactualists' focal uses, fictional names refer to merely possible objects, and whenever fictional names refer to something, they re-

---

[11]   Bonomi (1999/2008)'s label, also favored by García-Carpintero (2014). Thomasson (2003) labels this 'fictionalizing discourse'.

Notice that artifactualists (including Thomasson, van Inwagen and Kripke) are under no pressure then to say that Rowling, in the context of her fiction-making activities, aimed to use the name 'Harry Potter' to *refer* to an abstract artifact, and aimed to get her readers to interpret her in this way. Why? Because these artifactualists invoke the abstract artifact Harry Potter as a referent of 'Harry Potter' in the focal uses like (1) and (2) only.

[12]   Bonomi (1999/2008)'s label for sentences like (1) and (2), see also García-Carpintero (2014). Thomasson (2003) uses the label 'external discourse'. Salmon's (1998) label 'meta-fictional discourse' corresponds to a broader category that includes instances of metatextual discourse as well as examples like "According to the short stories, Sherlock Holmes is a detective". Kroon – Voltolini (2011) label the former external metafictional discourse, and the latter, internal metafictional discourse.

fer to merely possible objects. These general formulations suffice to facilitate moving between discussion of fictional characters like Sherlock Holmes (and mythical species like unicorn) and the reference of expressions like 'Sherlock Holmes' (and 'unicorn') on focal uses.[13]

The notion of a myth warrants another brief detour as there is a crucial variation in what various philosophers take myths to be. Should we, in the context of Kripke's exposition, draw a distinction between mythical and fictional labels, like 'unicorn' and 'Holmes', respectively? I think we shouldn't. In particular, there is one commonly drawn distinction between myths and fictions that is inapplicable to Kripke's discussion. Let's clarify that distinction and why it is about a notion of myth other than Kripke's. With respect to modes of introduction, it is natural to expect the *intentions and beliefs* of language users to be highly relevant to determining the mode in which they introduce expressions of their language. On this point, it is customary to note a key difference between myth and fiction:

> The difference between authors and myth-makers is one of propositional attitude: authors *make-believe* their works of fiction, whereas myth-makers do not make-believe their myths; rather, they genuinely *believe* their myths. (Caplan 2004, 334, emphasis in the original)[14]

According to Ben Caplan (and also David Braun, Nathan Salmon and Jeffrey Goodman), myth-makers believe their myths, and these philosophers' paradigm of a myth-maker is an astronomer putting forth a failed hypothesis about the existence of a celestial object that doesn't exist. The astronomer *believes* that the hypothetical object exists and inadvertently creates a myth even though her intention had been to describe reality. Plausibly, Caplan's formulation does not then leave room for the possibility that myth-making is a process in which a myth is *spun out of whole cloth* (as Kripke wrote). Why not? Because spinning out of whole cloth means "to

---

[13] Of course, Meinongians and nonactualists tend to (and unlike the artifactualist, easily can) provide a uniform account of the reference of fictional names, committing to universal claims about *all* uses of fictional names rather than the existential claims formulated here. But it's worth bearing in mind that any Meinongian (and nonactualist) committing to the universal claim is committed to this weaker one. So if I can show (as I aim to) that the weaker formulation is problematic, I thereby have raised a problem for the stronger formulation also.

[14] See also Salmon (1998); Braun (2005), Goodman (2014).

fabricate", "to invent with no basis in fact"; "to create complete fiction". And activities like fabricating are intentionally done, not by accident, and one cannot therefore *believe* what one's fabricating (one could at best *make-believe* it). In short, on Kripke's conception of a myth (*pace* Caplan), a myth-maker *need not* believe what her myth claims.[15] Cases in which she doesn't believe them are ones in which the labels in the myth are introduced in a way that is unsuited for the label to apply to any actual, concrete objects. This makes Kripke's notion of a myth crucially unlike that of Caplan's and other philosophers'.[16] The difference between the two conceptions of myth in a nutshell: *on the Caplanian notion of myth, for mythical expressions, (ii), the unsuited mode of introduction problem simply doesn't arise; but it does arise for the Kripkean notion of myth* (in connection with the term 'unicorn', for example). In the context of this paper, to keep the terminology straight, I will stick with Kripke's notion of myth, and will not use 'myth' in the Caplanian sense. In particular, I won't regard names of hypothetical objects in astronomy (discussed in Section 3) as mythical names. This completes the detour.

Right after the passage about unicorns, Kripke (1972/1980, 157-158) repeats the same point with respect to 'Sherlock Holmes' also: "it is theoretically possible though in practice fantastically unlikely, that Doyle was writing pure fiction with only coincidental resemblance to [an] actual man". A crucial consideration emerges from these fleeting remarks about unicorns and Sherlock Holmes: given (ii) the way the myth/fiction was

---

[15] Plausibly, Kaplan (1989, 609) shares Kripke's conception of a myth, see the quote in Footnote 17 from Kaplan, who mentions the possibility of *pure myth* and pure fiction.

[16] This difference is one that is often unrecognized and the Caplanian notion of myth is attributed to Kripke (see for example Goodman 2014, Braun 2012). See Zvolenszky (2015, footnote 8) and Braun (2015) commenting on Kripke's 2013 preface (Kripke 1973/2013, x) in which he makes quite clear that the notion of myth he had intended in the 1970s was unlike the Caplanian one and like the one I'm attributing to him here: he did not take myths to include failed scientific theories.

> I did not intend to apply the notion [that fictional and mythical characters are abstract objects whose existence depends on the existence or non-existence of various fictional or mythological works] to 'Vulcan', 'phlogiston' or other vacuous theoretical names of a more recent vintage, which are 'mythological' objects only in a highly extended and perhaps even metaphorical sense of 'mythological'. (Kripke 1973/2013, x)

created, and (i) the fact that we encounter historical unconnectedness, the result is that we find no more than coincidental resemblance to actual objects.

The unsuited-mode problem would arise even if we had at hand a myth or a novel specifying mythical beings/characters completely, down to the last bit of information about sock color and origin (it would be mind-numbing to read such a novel).[17] So even in special cases of names from complete fictions in which the metaphysical thesis is circumvented, the epistemological thesis would still present problems. In this way, the scope of the epistemological thesis is broader than that of its metaphysical counterpart. (Given the focus of this paper, in what follows, I will concentrate on characters from fiction, setting myths and mythical beings to the side; the points I make about the various problems can be generalized to names from myths also.)

Both the metaphysical and the epistemological theses and all the problems considered so far have taken it for granted that the candidate objects to count as Harry Potter are *concrete, spatiotemporal objects*. Notice that the artifactualist is claiming precisely that the name "Harry Potter" refers to an actual object, an actual *abstract* object. This seems, at first glance, to contradict the epistemological thesis, but it really doesn't, once we make explicit that both theses claim that it is *concrete* objects that are unsuited candidates as referents for expressions like 'Harry Potter' and 'Sherlock Holmes'. It is therefore well to keep the 'concrete' qualification in mind. For example, for (ii) we get: the fiction-writing mode in which the expression 'Harry Potter' had been introduced into the language is unsuited for the name to

---

[17] Kaplan (1989, 609) makes this point:

Insufficient specificity seems to be Kripke's qualm in *Naming and Necessity* regarding the merely possible species Unicorn and a merely possible referent for 'Sherlock Holmes' [the metaphysical thesis]. However, his discussion of what he calls "the epistemological thesis" (that the discovery that there were animals with all the features attributed to Unicorns in the myth does not establish that there were Unicorns) suggests an entirely different argument, namely that *the way in which these particular names arose* (from pure myth and pure fiction) makes it impossible for *them* to name merely possible entities. This argument is independent of the degree of specificity in the myth or in the fiction. (Kaplan 1989, 609)

refer to an actual concrete, spatiotemporal object.[18] For (i) we get: actual, concrete, spatiotemporal objects as potential referents for the name are historically unconnected to the introduction and subsequent use of 'Harry Potter'.

It's crucial to note that in the context of fictional names, of the two problems (i) and (ii), unsuited mode of introduction is the more fundamental one, explaining historical unconnectedness *of the relevant sort*. Given that (ii) Rowling's intention was to create a *fictional character* rather than refer to a flesh-and-blood person with introducing the name 'Harry Potter', (i) 'Harry Potter' was never historically linked (in the relevant way) to an actual orphaned boy wearing glasses, with a Z-shaped scar on his forehead, growing up in suburban England learning wizardry in a boarding school, and so on, and the name cannot refer to any actual concrete boy with spatiotemporal dimensions.

By the *relevant sort* of historical connection, I mean something quite specific: the kind of connection that fixes to whom or to what the name refers. For example, in writing *Robinson Crusoe*, Daniel Defoe incorporated into his story various details from the adventures of an actual castaway, Alexander Selkirk. So there is a causal-historical link between Selkirk and the novel's protagonist Robinson Crusoe: the former inspired the latter. But because Defoe's intention was to write about a fictional character, the his-

---

[18] In this respect, the mode of introducing proper names in the context of writing a work of fiction varies: Rowling introduced the name 'Harry Potter' intending it to refer to fictional characters; by contrast, she introduced 'London' in her first novel as a name of an already existing city. The unsuited mode of introduction problem arises in the former case, but not in the latter.

Someone might argue that proper names featured in fictional works never refer to actual objects: 'Napoleon' in War and Peace refers to a fictional surrogate of the historical figure, an abstract artifact (Voltolini 2013 proposes such a view). I won't explore such views here except for noting two points. First, such views are difficult to argue for as they are plausibly committed to fictional surrogates for the referents of all proper names even in the case of slightly fictionalized biographies or documentary genres (Voltolini is silent on this issue). Second, such views are affected by the claims I am making in this paper: the unsuited mode of introduction problem arises for such views, except it affects not just *some* proper names introduced in the context of fictional works (like 'Harry Potter'), but all names introduced in the context of fiction writing, including 'London' in the Potter novels and 'Napoleon' in *War and Peace*. I thank two anonymous reviewers for key observations on this matter.

torical connection between Selkirk and Crusoe's character is not of the relevant, reference-fixing sort.[19] So we can say in connection with the name 'Robinson Crusoe' also: lack of a historical connection (of the relevant sort) is due to the unsuited mode of introducing the proper name.

The historical unconnectedness problem and the unsuited-mode problem can be extended to concrete, spatiotemporal objects of all sorts, merely possible ones included; this way, we get an even more general formulation:

> *the historical unconnectedness problem generalized*: all concrete, spatiotemporal objects – whether they be actual or merely possible – bear no historical connection (of the relevant, reference-fixing sort) to the introduction and subsequent use of fictional names;
> *the unsuited-mode problem generalized*: the fiction-writing mode of introducing proper names into the language is unsuited for them to have as their reference concrete, spatiotemporal objects, whether they be actual or merely possible.

It is well to generalize in the same way the coincidental-resemblance problem also:

> *The coincidental-resemblance problem generalized*: there is no more than mere qualitative coincidence between concrete, spatiotemporal objects (whether they be actual or merely possible) and fictional characters as described in works of fiction.

Therefore (in the light of the generalization to merely possible objects), as we dig deeper, the problems behind the epistemological thesis turn out to target nonactualism.

As before, in the case of 'Harry Potter', the generalized unsuited-mode problem and the generalized historical connection problem underlie the generalized coincidental-resemblance problem.

And all three problems are in the background of the metaphysical thesis also: the generalized unsuited-mode problem provides the following *additional* reason for holding the metaphysical thesis: (independently of wheth-

---

[19]   I am claiming then that there is a key difference between Tolstoy's writing about Napoleon in the context of *War and Peace* (in which case his intention is to refer to a historical figure with the name) and Defoe's writing about Robinson (in which case his intention is to refer to a made-up figure with the name). See the previous footnote.

er the character of Harry Potter is fully specified in the novels) what grounds do we have at all for choosing, within a counterfactual scenario, between two distinct merely possible concrete, spatiotemporal objects that are qualitatively indistinguishable from the Harry of the novels which to count as Harry Potter when, given J. K. Rowling's fiction-writing mode of introducing 'Harry Potter', it would be a matter of sheer coincidental resemblance for the name to refer to either of those candidate objects? *With respect to names from fiction, the unsuited-mode problem (and in its wake, the coincidental resemblance problem) therefore raises a key issue underlying both the metaphysical and the epistemological theses discussed by Kripke*; this is a striking detail to bring to the surface given that Kripke mentions the unsuited-mode problem in passing only (saying no more than the two half-sentences quoted above), devoting far more attention to the metaphysical thesis. The upshot is then: in connection with fictional names and concrete candidates as their referents (actual as well as merely possible), the most fundamental problem underlying both the epistemological and the metaphysical theses is the generalized version of the unsuited mode of introduction problem.

3. An argument based on the historical unconnectedness problem

Just how bizarre the idea of reference based on coincidental resemblance is – the conception of reference for 'Harry Potter' to which the nonactualist is committed – can be brought out based on considerations about *nonfictional* names that fail to refer. The French astronomer Le Verrier put forth a hypothesis about the existence of an intra-Mercurial planet which he named 'Vulcan', to explain perturbations in the orbit of Mercury. There were various independent sightings mistakenly believed to be of Vulcan before enthusiasm dwindled; by 1916, Einstein's general theory of relativity confirmed that the perturbations were produced by the gravitational field of the Sun; there was no intra-Mercurial planet at all; the Vulcan-hypothesis was refuted; 'Vulcan' turned out not to refer to anything.

What about a counterfactual situation in which the Vulcan-hypothesis is a success story? Imagine a counterfactual scenario with the laws of physics slightly different, and there being an intra-Mercurial planet affecting the orbit of Mercury; Le Verrier puts forth his hypothesis; there are sightings

converging on the planet, which comes to be called 'Vulcan', the name featured in Le Verrier's prior hypothesis. But compared to *our* use of the name 'Vulcan', in the counterfactual circumstance, 'Vulcan' for the counterfactual planet is introduced and used in a markedly different way. It is preposterous to think that in coining the name in the actual world, Le Verrier managed to name *that counterfactual object even though his naming attempt failed in the actual world*. 'Vulcan' might have been a success story just as 'London' might have been introduced as a name for a river instead of a city; but all that is irrelevant to how and whether these strings, as parts of *our* language, were introduced and subsequently used (see Kripke 1972/1980, 77, 102-103, 109, especially, fn. 51; also 1971, 145). Le Verrier strove to name an actual concrete, spatiotemporal object; due to his failure to do so, he didn't *by coincidence* name a nonactual concrete, spatiotemporal object (as the nonactualist would have it); doing so was no part of his intention. So 'Vulcan' doesn't refer to any concrete objects in any counterfactual situations. Kaplan (1973, 506-508) makes this point eloquently with respect to a mythical name like 'Pegasus'.[20]

---

[20] Kaplan (1973, 506-508) writes as follows:

Suppose we start out by acknowledging that the Pegasus-myth is FICTION. Still it is, in a sense, possible. Should we not take 'Pegasus' to denote what it denotes in the world of the myth? We must be very careful now. ...

The myth is possible in the sense that there is a possible world in which it is truthfully told. Furthermore, there are such worlds in which the language, with the exception of the proper names in question, is semantically and syntactically identical with our own. Let us call such possible worlds of the myth '*M* worlds'. In each *M* world, 'Pegasus' will have originated in a dubbing of a winged horse. The Friend of Fiction, who would not have anyone believe the myth..., but yet talks of Pegasus, pretends to be in an *M* world and speaks its language.

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

I do not object to the inhabitants of one of the *M* worlds remarking that their name 'Pegasus' denotes something with respect to our world that does not exist in our world. But I reserve the right to retort that our name 'Pegasus' does not even denote with respect to their world.

But what is far more interesting is that the point holds for 'Vulcan'! The reason why this is interesting is because one of the problems that precludes 'Pegasus' from referring to a concrete object (actual or possible) is one that is specific to fictional (and mythical) names: the unsuited mode of introduction problem. That problem doesn't arise for 'Vulcan', only the historical unconnectedness problem does. What we see then in the case of 'Vulcan' is that the two core Kripkean claims (the historical-chain requirement and the qualitative-fit claim) are being applied straightforwardly, and being further generalized (in the light of the generalized versions of the historical unconnectedness and the coincidental resemblance problems), extending the core Kripkean claims not just to *actual* concrete objects but also to *merely possible* concrete objects. In effect, unfolding in front of us is the generalization of the qualitative-fit claim and the historical-chain requirement to concrete objects of all sorts, actual as well as merely possible.

We can say the following about the name 'Vulcan' in our language, as well as other proper names intended for concrete objects or for fictional characters: *if it cannot make it here, it won't make it anywhere.* If the name doesn't manage to refer to a concrete, spatiotemporal object here, in the actual world, it doesn't refer to such an object in other possible worlds either. Elsewhere (Zvolenszky 2007), I call this the *inverse-Sinatra principle for proper names.*[21, 22] Notice that this principle, unlike the two core Kripkean claims, goes beyond imposing constraints on the referents of proper names in the actual world, constraining also their referents in merely possible worlds. So it is well to generalize, in the light of the inverse-Sinatra principle, the qualitative fit claim and the historical-chain requirement to characterize the core tenets of a Kripkean stance:

---

[21]   Even an irrealist about fictional characters can, based on the considerations about Vulcan and unicorns above, accept the inverse-Sinatra principle. The principle commits her to the following: fictional names do not refer to anything in the actual world or any possible world. And this is something irrealists already accept.

[22]   Frank Sinatra sang about New York City: "If I can make it there, I'll make it anywhere". In the inverse-Sinatra principle (to keep it parallel with the song), I use the modal auxiliary 'can', by which I mean (as the song's 'can' does) 'is able to'; I don't mean metaphysical possibility. Thanks to Nathan Wildman for prompting me to clarify this.

- In the case of reference to concrete individuals (actual as well as merely possible) qualitative fit is neither necessary nor sufficient for being the referent of a name. Call this the *generalized qualitative-fit claim*.
- A causal-historical connection is necessary for a name to refer to a concrete object (actual or merely possible). Call this the *generalized historical connection requirement*.[23]

The inverse-Sinatra principle is quite general, covering names like 'Vulcan', 'Pegasus', and 'Harry Potter'. And the reason why these names don't make it anywhere given that they cannot make it here (in the actual world), is because nonactual concrete objects are, at best, coincidentally similar to Vulcan, Pegasus and Harry Potter, as these are described in various bodies of text. We thus have a nonfictional variant of the coincidental resemblance problem. Notice that the inverse-Sinatra principle is a name-based counterpart of the metaphysical thesis that is generalized to cover, besides fictional characters, nonfictional objects like Vulcan also. Crucially, however, the justification for the principle is very different from Kripke's (who we saw was focusing on the insufficient specificity problem underlying the metaphysical thesis). It is the generalized core Kripkean claims that justify the inverse-Sinatra principle.

Notice that 'Vulcan' and 'Harry Potter' differ in one crucial detail: the unsuited-mode problem doesn't extend to a nonfictional name like 'Vulcan'. Le Verrier's intention had been to introduce 'Vulcan' for a concrete, spatiotemporal object; so a historical connection, if there had been one,

---

[23] Kripke (1972/1980) did supply a further thesis that, together with the two core claims, yields the generalized versions of the two core claims for proper names that refer to concrete objects. That thesis is a well known one, about proper names being rigid designators: according to one formulation, a rigid designator r is such that if it refers to an object o in the actual world, then it refers to o in every world in which o exists, and in worlds in which o doesn't exist, r doesn't refer to an object other than o. But notice that the claim that proper names are rigid designators leaves open whether a proper name without an actual concrete referent does or doesn't refer to a concrete object in a merely possible world. It is the inverse-Sinatra principle that supplies the needed constraint for names like 'Harry Potter' and 'Vulcan': no concrete object to refer to here (in the actual world) means no concrete object to refer to in other possible worlds either. In this way, the rigid designation thesis about proper names and the inverse-Sinatra principle are two facets of an overarching theory about the reference of proper names across possible worlds.

linking uses of the name to an actual concrete object, could have served to fix the reference of 'Vulcan', circumventing coincidental-resemblance-related qualms. For a name like 'Vulcan', a historical connection can be secured in the actual world only – there is absolutely no historical connection between *our* use of 'Vulcan' and a merely possible concrete, spatiotemporal object. And in the absence of an actual historical connection, qualms about coincidental resemblance do arise, leading to the metaphysical thesis about the hypothetical object Vulcan, quite independently of whether or not the specification of Vulcan had been complete: we have no basis for counting any merely possible concrete object as Vulcan. (Notice that here, as before, my argument leading to the metaphysical thesis for Vulcan was crucially linked to considerations about coincidental resemblance and historical unconnectedness, which were originally identified behind the other thesis – the epistemological one. With respect to 'Vulcan', too, we see that the two theses are intimately connected.) The upshot is then: in connection with names of hypothetical objects that don't exist (like 'Vulcan') the most fundamental problem underlying both the epistemological and the metaphysical theses is the generalized version of the historical unconnectedness problem.

The foregoing observation about 'Vulcan' allows us to highlight a general point of advantage for the artifactualist position over both Meinongianism and nonactualism. This then translates into an argument for favoring artifactualism over its major realist rivals.

According to artifactualism, Harry Potter is an actual object that hasn't always existed. And the fact that he is an *actual* artifact makes room for a certain kind of causal-historical dependence on the physical world: in the 1990s, J. K. Rowling's creative activities bring it about that Potter is an *actual* abstract object.[24] The sort of dependence in place allows Harry Potter

---

[24] We might ask: how can we even make sense of a causal-historical chain between concreta and abstracta? I have considered this issue elsewhere. First, note that Kripke assumes the existence of causal-historical chains linking people and the names of numbers like π (see Kripke 1972/1980, 115-116, footnote 58; Kaplan 1989, 607-608, footnote 101); and numbers are the quintessential candidates for abstract objects; but it is unclear if this makes for a causal-historical chain linking the *abstracta* and the names (see Footnote 27). There is a second, more promising response: it seems overwhelmingly natural to assume that expressions like 'soccer', 'marriage', 'The Magic Flute', 'The Constitution of the United States' denote *abstract artifacts* (see Thomasson 1999;

qua abstract artifact to be the kind of referent for Rowling's name 'Harry Potter' with respect to which issues having to do with historical unconnectedness and, in turn, coincidental resemblance, and, in turn, the epistemological thesis, do not arise.[25] (Notice that before, we noted that for names of fictional characters, no historical connection to concrete, spatiotemporal objects is of the relevant, reference-fixing sort. Meanwhile, the point made here is that for the artifactualist, a historical connection to an *actual abstract artifact* is precisely what fixes the reference of 'Harry Potter'.)

We have already seen that the unsuited mode of introduction problem prevented concrete objects (actual as well as merely possible ones) from being the right sort of candidates to be referents of 'Harry Potter'. What about abstract, author-created artifacts as such candidates? For such objects, the unsuited mode of introduction problem does not arise (in this, the situation is analogous to the one in which merely possible concrete objects are considered as candidates for being the referents of 'Vulcan'). What about the historical unconnectedness problem? The point I was making in the previous paragraph is that that problem is also avoided. So the option of 'Harry Potter' referring to an abstract, author-created artifact does not face *any* of the problems behind the metaphysical and epistemological theses. And the argument emerging is that of the realist alternatives considered, artifactualism is the only view that readily manages to escape all the problems associated with the metaphysical and epistemological theses.

By contrast, alternative realist accounts that make Harry Potter a concrete object whose existence does not causally depend on us either because

---

Zvolenszky 2012, 2013, 2015); the burden of proof is on those who want to claim that these social and cultural objects are *not* abstract artifacts. And once we admit as abstract objects social and cultural artifacts, we already have to secure the possibility of a causal-historical chain for names introduced into our language for these objects. This possibility can then be extended to names of fictional characters also.

[25]   Of course, having said this much leaves unspecified the nature of the character Harry Potter's dependence on authorial activity, the circumstances of and constraints on creating fictional characters; these are issues that Thomasson (1999), an artifactualist, explores at length. To formulate a complete artifactualist account, such details have to be filled in. My purpose in this paper has been more modest than that: I aim to show that certain challenges having to do with historical unconnectedness create difficulties for major realist theories about fictional characters but not for artifactualism.

the object is nonexistent (according to Meinongianism) or because it is nonactual (according to nonactualism), face a challenge. First, these theorists have to explain why those objects are candidates of the right ontological status to count as the referents of 'Harry Potter'. As we have already seen, on this point, the nonactualist founders already (stumbling on the unsuited mode of introduction problem).[26] The Meinongian can get past this hurdle: he may suggest that his nonexistents are objects of thought and hence have just the right sort of ontological status to be suitable targets of authors' intended reference. But on the next hurdle the Meinongian stumbles: if his nonexistent objects are of a suitable sort as objects of fiction-writing, what historical connection is there to account for Rowling's 'Harry Potter' referring to one of countless nonexistent candidate objects (each equally faithful to the way Potter is depicted in the novels but varying in details left unspecified – about sock color, etc.)? The Meinongian cannot provide such a historical connection: causal-historical connection between his timelessly nonexistent objects and actual concreta (like authors) is extremely problematic, downright unintelligible even. And because of historical unconnectedness, the Meinongian is confronted with qualms about having to work with no more than coincidental resemblance between Harry Potter as specified in the novels, and various qualitatively similar but nonidentical Meinongian nonexistents. And, on the one hand, coincidental resemblance does not suffice for reference (in the light of the historical unconnectedness problem), according to the epistemological thesis; and, on the other hand, with insufficiently specified characters like Harry Potter,

---

[26] One version of nonactualism that is worth exploring further in this connection is markedly different from Lewis's: Priest's (2005) (im)possibilism. According to Priest, Rowling, in writing her novels, intends a particular individual that is (i) nonactual, (ii) not created by Rowling, and that (iii) realizes the Harry Potter novels in some other possible worlds. A radical feature of Priest's proposal (making it very different from Lewis's) lies in (iv) his commitment to impossible worlds and impossible objects (to account for impossible fiction). Whether Priest's view can get past the first hurdle (the unsuited mode of introduction problem) and the second one (the historical unconnectedness problem that confronts the Meinongian) is subject to debate (Priest argues that his view clears both hurdles, Kroon – Voltolini (2011), for example, raise doubts on both counts); but what is clear is that the hurdles stand tall before Priest's version of nonactualism but not in front of artifactualism, given the latter's unique combination of two features for fictional characters: their being actual and human-created.

coincidental resemblance leaves room for the metaphysical thesis and the insufficient specificity problem to arise.[27]

---

[27] This line of argument brings to the fore why the only abstract-object theory contender we considered for fictional characters was artifactualism: it is the only view according to which Harry Potter is created and hence historically linked to goings on in the actual world. Platonism, a theory according to which Harry Potter is a *timelessly existing* abstract object (akin to numbers, sets), would, like Meinongianism and nonactualism, run into problems with historical unconnectedness and hence coincidental resemblance, and, in their wake, the metaphysical and epistemological theses. For an attempt to combine the advantages of artifactualism and Meinongianism in a Platonist framework, see Zalta (2000; 2006).

One might then wonder: what about a Platonist about *numbers*, who thinks such objects are abstract timeless existents? In the light of the foregoing arguments, how is it that we can make room for proper names referring to such objects, a name like π, for example? A promising reply: reference to abstract timeless existents is possible but exactly as difficult as reference to merely possible objects is. The merely possible object in question has to be fully specified, in other words, has to escape the insufficient specificity objection (behind the metaphysical thesis). This is probably why Kaplan claims that "ever-unactualized possibilia are extraordinarily difficult to dub" (Kaplan 1973, 505), while leaving it open that unactualized possibilia are *possible* to dub.

Imagine a lectern-kit that never gets assembled. We decide to give the name 'Woody' to the merely possible lectern that would have resulted from assembling the parts by following the instructions in the kit. It would seem unwarranted to deny the possibility of dubbing a nonexistent in this way. To make things even smoother, let us imagine that no parts (not even screws) can be interchanged for others in assembling the lectern – each component has a unique intended spot, and Woody is the lectern that would result from placing all parts where they belong. With such degree of specificity, we can introduce the name 'Woody' to refer to a merely possible object. See Kaplan (1989, 607-608; 1973, 517, n19). Also, Salmon (1981, 39, n41) reports Kaplan's and Kripke's willingness (in lectures and conversation) to allow the introduction of the name 'Noman' to name "the person who would have developed from the union of this sperm and that egg, had they been united".

Now, full specification for numbers is not problematic: one can give a definition that uniquely fits them and specify that a name like π is to refer to the unique object fitting the definition: is the ratio of a circle's circumference to its diameter. The position I am outlining is that names of numbers as Platonic objects refer to whatever fits the definition associated with the name. So in this very special case, reference *is* based on qualitative fit. But it would take further argument to justify an analogous claim for *created* abstracta like the artifactualist's Vulcan and Harry Potter. I doubt such an argument can be made plausible (though I won't argue for this here), and accordingly, I also doubt that insufficient specificity is the main reason why 'Vulcan' and 'Harry Potter' are

In connection with nonfictional empty names like 'Vulcan', we have found that of the three problems underlying the epistemological thesis (as well as the metaphysical one), two can be extended to 'Vulcan' also: the historical unconnectedness problem, and, in its wake, the coincidental resemblance problem. Indeed, it's well to recognize that this pair of problems provides a key pair of tests for referring to actual, concrete objects – about the presence of a historical connection, and about a relation that goes beyond coincidental resemblance – that applies to *all* proper names. 'Vulcan' and 'Harry Potter' fail the pair of tests while a nonempty name like 'J. K. Rowling' passes it. With a causal-historical connection in place between Rowling and uses of the name 'J. K. Rowling', there is more than qualitative resemblance linking the person and the name.

In addition, we have found that the results are the same for the pair of tests construed as testing reference to Meinongian objects. There is no historical connection (and hence no more than coincidental resemblance) to link 'Vulcan' and 'Harry Potter' to such objects. Crucially, once the candidate referent for a name like 'Harry Potter' is construed as an abstract artifact created by the activities of Rowling, a historical connection between creator and the created object *is* secured after all. Hence, 'Harry Potter' (on some occasions of use) referring to an abstract artifact is an option that escapes the historical unconnectedness problem.[28]

---

names for which the inverse-Sinatra principle holds. I thank Karen Lewis for discussion on this.

[28]   Notice that we have arrived at this result without incurring any commitment about the metaphysics of hypothetical objects that are posited by failed scientific theories (including Vulcan). This is an important strategic feat for an artifactualist about fictional characters given that one of the most influential arguments for her view (that sentences like (1) and 'Harry Potter is an abstract artifact' involve reference to an abstract artifact Harry Potter) is widely thought to carry over to names like 'Vulcan' also (see Caplan 2004, Braun 2005, Goodman 2014), yet Goodman argues that there is an objection specifically targeting artifactualism about the likes of Vulcan. This, by *modus tollens*, threatens to undermine the influential argument for artifactualism about fictional characters. But not the argument I presented in this paper.

4. The argument in the light of the causal-historical chain theory

Once fleshed out, Kripke's (1972/1980) fleeting remarks about fictional characters can be summarized as follows: qualitative resemblance is insufficient to determine the reference of a proper name; a causal-historical connection between names and their referents is necessary to determine to whom or to what proper names refer. This is eloquently discussed by Keith Donnellan, who, in parallel with Kripke, also formulated the proposal that the reference of proper names is determined by causal-historical chains of use (Donnellan's paper was published in 1970, shortly after Kripke delivered his *Naming and Necessity* lectures):

> Suppose that Aristotle and Herodotus were … making up the story [about Thales] … Suppose further, however, that fortuitously their descriptions fitted uniquely someone they had never heard about and who was not referred to by any authors known to us. Such a person, even if he was the only ancient to hold that all is water, to fall in a well while contemplating the stars, etc., is not 'our' Thales. (Donnellan 1970, 352)

Notice that Donnellan, like Kripke, is claiming that coincidental resemblance alone (in the absence of a historical connection) is insufficient for reference when the story is "made up".[29]

For names of actual concrete objects like 'J. K. Rowling' and 'London', a parallel overarching lesson transparently emerges from the second lecture of *Naming and Necessity*: a name like 'Gödel' refers to a person only if he is the one being named at the end of the causal-historical chain of uses leading back to the introduction of 'Gödel'; qualitative resemblance is not enough to make a person Gödel (cf. Kripke 1972/1980, 91). It is considerably less transparent that, once we unpack and develop Kripke's fleeting remarks about the names of fictional characters, we arrive at the very same

---

[29]   In other words, Donnellan is noting the problem of historical unconnectedness arising for names featured in stories that are spun out of whole cloth. In the same paper, he also formulated, in parallel with Kripke, the causal-historical chain picture of reference determination for proper names. For a similar idea, see also Kaplan's early proposal that the geneology of a proper name is what determines who a proper name is of. (Kaplan 1967, 197-200); he credits Kripke for the idea: "The … principle for determining who it is that a given proper name, as it is used by some speaker, names, was first brought to my attention by Saul Kripke" (Kaplan 1967, 213, fn. 24).

lesson, but a generalized version of it: concerning not just actual-world reference but reference in merely possible worlds also. And in this way (as highlighted by the inverse-Sinatra principle I proposed), we arrive at a generalized version of the two core Kripkean claims (about qualitative fit being neither necessary nor sufficient for reference and a historical connection being necessary for reference) that applies to the whole spectrum of proper names, whether they have or lack a bearer. And on this basis emerges a pro-artifactualism argument (one that Kripke did not formulate): of the forms of realism considered, artifactualism is the only one that can heed the generalized lesson at hand. Artifactualism heeds this lesson precisely because (unlike alternative views) it incorporates authorial creation.[30]

References

BONOMI, A. (1999/2008): Fictional Contexts. In: Bouquest, P. – Serafini, L. – Thomason, R. H. (eds.): *Perspectives on Context*. Stanford: CSLI Publications, 213-248.

BRAUN, D. (2005): Empty Names, Fictional Names, Mythical Names. *Noûs* 39, 596-631.

BRAUN, D. (2012): Hob, Nob, and Mythical Witches. In: Campbell, J. K. – O'Rourke, M. – Silverstein, H. S. (eds.): *Reference and Referring*. Cambridge (Mass.): MIT Press, 148-188.

BRAUN, D. (2015): Wondering about Witches. In: Brock, S. – Everett, A. (eds.): *Fictional Objects*. New York: Oxford University Press, 71-113.

BROCK, S. (2010): The Creationist Fiction: the Case against Creationism about Fictional Characters. *Philosophical Review* 119, 337-364.

CAPLAN, B. (2004): Creatures of Fiction, Myth, and Imagination. *American Philosophical Quarterly* 41, 331-337.

Donnellan, K. (1970): Proper Names and Identifying Descriptions. *Synthese* 21, No. 3 – 4, 335-358.

Evans, G. (1973): The Causal Theory of Names. *Proceedings of the Aristotelian Society*, Suppl. Vol. 47, 187-208.

Fine, K. (1982): The Problem of Non-Existents. I. Internalism. *Topoi* 1, 97-140.

García-Carpintero, M. (2014): Introduction: The Problem of Empty Representations. In: García-Carpintero, M. – Martí, G. (eds.): *Empty Representations: Reference and Non-Existence*. Oxford: Oxford University Press, 1-24.

Goodman, J. (2014): Creatures of Fiction, Objects of Myth. *Analysis* 74, No. 1, 35-40.

Kaplan, D. (1967): Quantifying In. *Synthese* 19, No. 1 – 2, 178-214.

Kaplan, D. (1973): Bob and Carol and Ted and Alice. In Hintikka, K. J. I. – Moravcsik, J. M. E. – Suppes, P. (eds.): *Approaches to Natural Language*. Dordrecht: D. Reidel, 490-518.

Kaplan, D. (1989): Afterthoughts. In: Almog, J. et al. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 565-614.

Kripke, S. (1963): Semantical Considerations on Modal Logic. *Acta Philosophica Fennica* 16, 83-94.

Kripke, S. (1971): Identity and Necessity. In: Munitz, M. K. (ed.): *Identity and Individuation*. New York: New York University Press, 135-164.

Kripke, S. (1972/1980): Naming and Necessity. In: Davidson, D. – Harman, G. (eds.): *Semantics of Natural Language*. 2nd ed. Dordrecht: D. Reidel, 253-355; Reprinted with Preface as Kripke, S. (1980): *Naming and Necessity*. Cambridge (Mass.): Harvard, page references are to the reprint.

Kripke, S. (1973/2011): Vacuous Names and Fictional Entities. In: *Philosophical Troubles: Collected Papers, Volume 1*. New York: Oxford University Press, 52-74.

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

Kroon, F. – Voltolini, A. (2011): Fiction. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition), available at: http://plato.stanford.edu/archives/fall2011/entries/fiction/.

Lewis, D. (1978): Truth in Fiction. *American Philosophical Quarterly* 15, 37-46. Reprinted in: Lewis, D. (1983): *Philosophical Papers*, Vol. 1. Oxford: Oxford University Press, 261-275.

Lewis, D. (1986): *The Plurality of Worlds*. Oxford: Blackwell

Meinong, A. (1904): Über Gegenstandtheorie. In: Meinong, A. (ed.): *Untersuchungen zur Gegenstandtheorie und Psychologie*. Leipzig: Barth. Translation by I. Levi, D. B. Terrell, and R. M. Chisholm in: Chisholm, R. (ed.) (1960): *Realism and the Background of Phenomenology*. Glencoe: Free Press, 76-117.

Parsons, T. (1980): *Nonexistent Objects*. New Haven: Yale University Press.

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

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

Salmon, N. (1981): *Reference and Essence*. Princeton: Princeton University Press.

SALMON, N. (1998): Nonexistence. *Noûs* 32, 277-319.

SEARLE, J. R. (1979): The Logical Status of Fictional Discourse. In: French, P. A. et al. (eds.): *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minneapolis Press, 233-243.

SCHIFFER, S. (1996): Language-Created Language-Independent Entities. *Philosophical Topics* 24, 149-166.

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

THOMASSON, A. L. (2003): Speaking of Fictional Characters. *Dialectica* 57, No. 2, 205-223.

VAN INWAGEN, P. (1977): Creatures of Fiction. *American Philosophical Quarterly* 14, 299-308.

VAN INWAGEN, P. (2000): Quantification and Fictional Discourse. In: Everett, A. – Hofweber, T. (eds.): *Empty Names, Fiction, and the Puzzles of Nonexistence*. Stanford: CSLI Publications, 235-247.

VOLTOLINI, A. (2013): Probably the Charterhouse of Parma Does Not Exist, Possibly Not Even That Parma. *Humana Mente, Journal of Philosophical Studies* 25, 235-261.

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

YAGISAWA, T. (2001): Against Creationism in Fiction. *Philosophical Perspectives* 15, 153-172.

ZALTA, E. N. (1983): *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht: D. Reidel.

ZALTA, E. N. (2000): Pretense Theory and Abstract Object Theory. In: Everett, A. – Hofweber, T. (eds.): *Empty Names, Fiction, and the Puzzles of Nonexistence*. Stanford: CSLI Publications, 117-147.

ZALTA, E. N. (2006): Deriving and Validating Kripkean Claims Using the Theory of Abstract Objects. *Noûs* 40, No. 4, 591-622.

ZVOLENSZKY, Z. (2007): Naming with Necessity. In: *Modality, Names and Descriptions*. Ph.D. thesis. New York University.

ZVOLENSZKY, Z. (2012): Against Sainsbury's Irrealism about Fictional Characters: Harry Potter as an Abstract Artifact. *Hungarian Philosophical Review (Magyar Filozófiai Szemle)* 2012, No. 4, Special Issue: Realism, 83-109.

ZVOLENSZKY, Z. (2013): Abstract Artifact Theory about Fictional Characters Defended – Why Sainsbury's Category-Mistake Objection is Mistaken. *Proceedings of the European Society for Aesthetics* 5, 597-612.

ZVOLENSZKY, Z. (2014): Artifactualism and Authorial Creation. *Proceedings of the European Society for Aesthetics* 6, 457-469.

ZVOLENSZKY, Z. (2015): Fictional Characters, Mythical Objects and the Phenomenon of Inadvertent Creation. To appear in *Res Philosophica*.

Modal Empiricism and Knowledge of *De Re* Possibilities: A Critique of Roca-Royes' Account

Duško Prelević

Faculty of Philosophy. University of Belgrade
Čika Ljubina 18-20. 11000 Belgrade. Serbia
dusko.prelevic@f.bg.ac.rs

ABSTRACT: Accounting for our knowledge of *de re* modalities is probably the main reason why the proponents of modal empiricism think that their view should be preferred to modal rationalism. In this paper, I address Sonia Roca-Royes' account, which is taken to be a representative modal empiricist view, in order to show that modal empiricism faces serious problems even in explaining our knowledge of possibility *de re*, something which seems to be the easiest thing to explain on this view. I argue that Roca-Royes' account does not prove what she claims it does, that it can hardly be articulated in a non-redundant way, and that her account of our knowledge of possibility *de re* can hardly be reconciled with the essentiality of origin principle, to which modal empiricists sometimes appeal while criticizing the modal rationalist account.

KEYWORDS: *De re* modal knowledge – modal empiricism – modal epistemology – modal rationalism.

1. Modal rationalism and our knowledge of modality *de re*

Modal epistemology examines possibilities and boundaries of our modal knowledge. One of the most intriguing questions in this philosophical discipline is related to the knowledge of unrealized possibilities, that is, the knowledge that something is possible even if it is not realized in the actual world (or at least we do not know whether it is realized or not; see Van In-

wagen 1998, 74). Modal rationalism and modal empiricism are two principal competing views on this issue. The former is the view that there is an *a priori* access to metaphysical modality, while the latter is the view that at least some modal beliefs require *a posteriori* justification, primarily those beliefs that are related to modalities *de re*, that is, to modalities that are attached directly to objects. In what follows, it will be argued that modal empiricism meets serious problems in explaining our knowledge of possibility *de re*, something which seems to be the main reason why this view is worth considering. In §2, I present Sonia Roca-Royes' account of the knowledge of possibility *de re*, which is taken to be a representative modal empiricist view, and in §3 I will pose several objections to this account that I find hard to answer.

But first, let us sketch briefly modal rationalist approach, proposed by David Chalmers, in order to understand better why some philosophers are willing to search for alternative approaches. Namely, Chalmers has articulated the relevant senses of conceivability and possibility, and introduced the epistemic version of two-dimensional semantics in order to handle ordinary Kripkean cases of necessary *a posteriori* statements (see Chalmers 2010, for more details). By this maneouvre, he has tried to support the link between conceivability and possibility that was questioned by Kripke's examples (see Kripke 1972). According to Chalmers, the relevant notion of conceivability is what he calls 'ideal positive primary conceivability', which consists in conceiving of a counter-actual scenario (or a counter-actual situation) that verifies the statement one is conceiving of and which is undefeatable by a better reasoning. Understood in this way, conceivability, according to Chalmers, entails primary (counter-actual) possibility, and, depending on the semantics of concepts that are involved in propositions, it might entail secondary (counterfactual) possibility, giving us interesting results in metaphysics.

Although Chalmers is sometimes credited for explaining successfully *a posteriori* cases, some philosophers think that he remained silent on the cases of modality *de re*. For example, Sonia Roca-Royes thinks that there are two sorts of Kripkean examples rather than one, *a posteriori* and *de re* reading respectively, and that Chalmers' modal rationalism is capable of explaining *a posteriori* reading of necessary *a posteriori* statements, but not capable of explaining successfully *de re* reading (Roca-Royes 2011, endnote 28; see also Vaidya 2008, 206). For example, necessary *a posteriori* statement

'Water is H$_2$O' can be read, according to Roca-Royes, either as 'Necessarily, water is H$_2$O' (*a posteriori* reading), or as 'Water is necessarily H$_2$O' (*de re* reading). The similar holds for the principle of the essentiality of origin, which is one of the main essentialist principles. It seems that a relevant reading of this principle, like in the case of necessary *a posteriori* statement 'This table is necessarily wooden', is *de re* reading. According to Kripke, origin is an essential property of an object that enables us to individuate it in every possible (counterfactual) world. Namely, the principle of the essentiality of origin enables us to distinguish two qualitatively similar objects (for example, two tables) by their origin (the material they are carved from; see Kripke 1972, footnote 56). Given that *de re* modality is attached directly to *objects*, it is not clear at all how conceptual analysis based on conceivability can be of any use here. If so, then modal rationalism, according to the objection, cannot be the whole story about our knowledge of metaphysical modality.

However, modal rationalists can bite the bullet and argue that their view is quite in accordance with the existence of modality *de re*. Namely, it seems that the existence of modality *de re* is not by itself an obstacle to modal rationalist approach, because at least some trivial *de re* properties are obviously knowable *a priori*, like the property *being self-identical* (cf. Hossack 2007, 440). If so, then the conceivability method seems reasonable to apply to the cases of non-trivial *de re* possibilities as well, given that conceivability can be defined through *a priori* knowledge ('*p* is conceivable' is usually defined as 'non-*p* is not *a priori*', while '*p* is *a priori*' can be defined as 'non-*p* is not conceivable'). Probably more should be said in favor of this account,[1] yet, setting this aside, it is interesting to check whether modal empiricists themselves are capable of explaining successfully our knowledge of modality *de re*, because if not, modal rationalism would have a greater initial plausibility concerning the explanatory power even before elaborating their own account. In this paper, the emphasis will be on assessing one popular modal empiricist account of our knowledge of modality *de re*, proposed by Sonia Roca-Royes, and showing why it does not work.

---

[1]   One such account can be found in Chalmers (2010, footnote 3).

2. Roca-Royes' account

Now, let us turn to modal empiricist approach to metaphysical modality. Here, Roca-Royes' account of our knowledge of possibility *de re* will be addressed mainly because her account offers a solution to Benacerraf's dilemma (see Benacerraf 1973), which is usually considered to be more pressing challenge to modal empiricists than to modal rationalists. This is mainly because Benacerraf's dilemma is based on the idea that the causal theory of knowledge, which plausibly holds for *a posteriori* knowledge, is not in accordance with the principle that abstract entities, including modality, are not causally related. This means that in the case of our knowledge that something is possible it is hard to see how one can empirically know in the actual world that something holds in a possible world, if there is no causal connection between these two worlds.

Here is Roca-Royes' purported example that illustrates her account of our knowledge of possibility *de re*. One might believe, that it is possible for a table, let us call it 'Messy', to be broken, although this had never happened in the actual world. Roca-Royes (2007, §4; 2014) thinks that we are justified in believing that this possibility is a real one once we have found out somehow that some other table in the actual world, let us call it 'twin-Messy', which is similar enough to Messy, was broken. Here, twin-Messy is Messy's counterpart in the actual world. Roca-Royes introduces counterpart relation, which, in contrast to standard Lewisian counterpart relation (see Lewis 1979, 113) that holds only for individuals in different worlds, relates two individuals in *the same* (actual) world. Namely, both Messy and its counterpart twin-Messy inhabit the same (actual) world. Yet, the counterpart relation that Roca-Royes uses has the same relevant features that Lewis' counterpart relation has, like non-identity, non-symmetry and non-transitivity. As for twin-Messy, which is actually broken, when it is said that it is possible to it to be broken, it does mean that twin-Messy *itself* inhabits a possible world in which it is broken.[2] Now, given that actuality entails possibility,[3] it is possible for twin-Messy to be broken. If so, then, ac-

---

[2]   By this maneouvre, Roca-Royes avoids the problem of non-transitivity of counterpart relation that might be posed if her example includes twin-Messy's counterpart (thanks to an anonymous referee for drawing my attention to this problem).

[3]   Roca-Royes (2014) thinks that this is a conceptual truth.

cording to Roca-Royes, the same is possible for Messy itself, because twin-Messy is Messy's counterpart. Thus, it seems that there is a sense in which an empirical evidence might justify our modal beliefs concerning possibility *de re*, which is exactly what modal empiricists claim. On the other hand, Roca-Royes thinks that in the absence of empirical evidence that a counterpart of an object has such-and-such properties we should restrain our judgment on what is possible for that object, because these are cases in which possibilities are not anchored in our experience.

Roca-Royes' account seems to be an elegant way of explaining our knowledge of possibility *de re*, as well as a way how to avoid Benacerraf's dilemma. Her solution consists in avoiding a *direct* empirical access to *de re* possibilities by means of introducing a counterpart relation that is empirically established in the actual world, and by using *a priori* inference that is based on the premise that actuality entails possibility.

Although the abovementioned account, proposed by Roca-Royes, is restricted to the cases of our knowledge of *possibility de re*, and does not hold either to the cases of our knowledge of *de re* necessities, or to the essentialist principles, Roca-Royes claims in her other papers that any good theory of the epistemology of modality should account for our knowledge of essentialist truths as well, if there is such a knowledge (Roca-Royes 2011, 23). Given that she argues that various forms of modal rationalism are not capable of explaining our knowledge of essentialist principles, it seems that she herself is committed to the claim that such a knowledge exists, and obliged to provide an explanation of it. In the next section, it will be argued that Roca-Royes' account cannot achieve this.

3. Some problems for Roca-Royes' account

In what follows, I will pose three objections to modal empiricist account, presented in the previous section, which I find hard to answer. My first objection to this account is that it is not clear at all that Roca-Royes' example with Messy and its real world counterpart, twin-Messy, shows that one is justified in believing that it is possible *de re* for Messy to be broken. In other words, it is not clear that the possibility for twin-Messy to be broken tells us anything about the possibility for Messy to be broken. After all, we should remind ourselves that the initial aim of modal empiricists was to

explain our knowledge of possibility *de re*, in our case, the possibility for *Messy* being broken, not the possibility for twin-Messy being broken. Given that counterpart relation is not identity, information about what is for Messy's counterpart possible does not entail by itself that Messy might be broken. Thus, contrary to modal empiricists account, the possibility for twin-Messy to instantiate a property does not entail the possibility for Messy to instantiate the same property, and therefore such an account does not succeed.

Related to this, it is possible to construct counterexamples to modal empiricist account of our knowledge of possibility *de re*, and show that such an account does not ensure reliable method of getting the knowledge that something is *de re* possible. Let us suppose, for example, that a person *x* has a real world counterpart that is almost exactly the same as *x* in all physical respects, except that she differs from *x* with respect to just one particular DNA molecule. As it is well-known,[4] this might cause huge differences in abilities between these two beings, which would serve as a good evidence that what is possible for one of them might not be possible for the other. Moreover, given that abilities (or disabilities) need not manifest themselves, the differences in (dis)abilities between these two persons need not be empirically revealed as well. If so, then what holds for a real world counterpart of a person needs not hold for the person herself, which would contradict Roca-Royes' proposed account.

As a response to this objection, Roca-Royes might restrict her account to artifacts, without applying it to biological kinds. Maybe this is the reason why artifacts (the table called 'Messy' and its counterpart) are mentioned in her example. This would be in accordance with the intuition, shared by some philosophers, that artifacts, in contrast to biological kinds, allow of small changes without changing their nature. For example, there is an intuition that an artifact can origin from a slightly different piece of material, without ceasing to be the same object (see, for example, Vaidya 2008, 202-203). Perhaps this is mainly because the nature of artifacts is determined by their functions rather than by the life-cycle that is characteristic for organisms (see Hale 2013, 278-279, for more details).

However, this fallback seems to be still liable to another sort of counterexamples. Given that dispositions seem to be a good guide to possibility,

---

4   For example, in the case of Dawn syndrome, etc.

one can, *mutatis mutandis*, appeal to the cases of *finking*, introduced by Charles Martin for some other purposes.[5] According to Martin, 'a disposition and a change of disposition need not manifest themselves' (Martin 1994, 1). For example, disposition can be changed due to changes in environment, like in the cases of temperature change, and so on. A usual counterfactual explanation of why a glass is fragile consists in asserting that it would be broken, were it to fall down. Yet, suppose that the very glass itself has a disposition of transforming into steel when struck. This would falsify the counterfactual analysis of dispositions, because in the purported example the glass would not break, were it to fall down. Now, let us apply this insight to Roca-Royes' account. Let us suppose that there is *a unique* glass in this world that belongs to the case of finking, that is, which would not be broken, were it to fall down. Its counterpart in the actual world would plausibly be a glass that is qualitatively the most similar to it, yet which is, in contrast to it, fragile. Suppose that this counterpart falls down at one point and breaks. Is this reliable evidence that the glass with the above-mentioned finking property is fragile as well? It is hard to believe so. Thus, it seems that Roca-Royes' account faces problems in the case of artifacts as well.

As a response to such an objection, the proponents of modal empiricism might claim that their account is based on inductive evidence, as well as on Humean regularity theory of natural laws, which is something that Roca-Royes would be inclined to accept (see Roca-Royes 2007; 2014). This probably would include *ceteris paribus* clause, which enables us to justify the claim that it is possible for Messy to be broken in the light of *a posteriori* discovered examples in which Messy's counterparts were broken, and such an account seems to be in accordance with common scientific practice.

However, this answer leads to similar problems that are involved in counterfactual analysis of dispositions stressed by Martin. Namely, Martin

---

[5]   Martin constructs his examples in order to show that counterfactual analysis of dispositions does not succeed. Also, his point was mainly that an object can have a disposition that needs not be manifested and analyzed by means of counterfactual conditionals. In my modified example, it is presupposed that an object does not have a disposition, and that this needs not be manifested itself. This is, in my opinion, in accordance with Martin's general idea that dispositions and a change of dispositions need not manifest themselves.

notices that adding *ceteris paribus* clause to the counterfactual analysis of dispositions would trivialize the whole analysis (cf. Martin 1994, 6). Similarly, if one adds *ceteris paribus* clause to Roca-Royes' account, the whole account will become question-begging. For example, if modal empiricists claim that it is possible for Messy to be broken because we empirically know that twin-Messy is broken, and because we can expect, *ceteris paribus*, that Messy would break when struck, they have already presupposed that there is no finking.

Let us turn to my second objection to Roca-Royes' account, which is closely related to the previous one. Namely, one would expect that modal empiricists can pick out directly a pertinent possibility *de re*, through, let us say, perception. Yet Roca-Royes purported example with twin-Messy suggests that the possibility *de re* is picked out by means of a counterpart relation. This counterpart relation is arguably based on providing a pertinent description that justifies our belief that two objects stand in such a relation. To be more precise, we establish that the two objects stand in a counterpart relation once we have realized that they are *similar* to each other. And such a similarity can be grasped only by description. In Roca-Royes' example, one's belief that it is possible for Messy to be broken is based on one's knowledge that it is possible for *the table that is similar enough to Messy* to be broken. While the former possibility is the possibility *de re*, the latter, which enables us to find out the former, is the possibility *a posteriori*. The upshot is that the knowledge of modality *de re* presupposes the corresponding knowledge of modality *a posteriori*. This turns us back to modal rationalist justification of our knowledge of possibilities *de re*,[6] which makes modal empiricist explanation of *de re* modal knowledge *redundant*.

My third objection to modal empiricism is that Roca-Royes' explanation of our knowledge of possibility *de re* is not in accordance with some main essentialist principles, such as the essentiality of origin, and so on. Namely, we have seen in §1 that the essentiality of origin implies that a particular hunk of matter, from which a table is carved, is an essential property of the table. Modal empiricists typically try to provide an epistemic justification of this principle by saying that we find out *a posteriori* which particular piece of matter is the origin of a particular object. Roca-Royes

---

[6]   In §1 it is mentioned that many critiques of modal rationalism agree that this view is capable of explaining *de dicto* cases.

(2014) admits that her proposed account of our knowledge of possibility *de re* does not provide us with an epistemology of essence, that is, that her account remains silent on whether essentialist principles hold or not. Her view is that we cannot decide in a non-question-begging way whether the essentiality of origin (or the essentiality of kind) principle holds or not, mainly because we do not have *anchoring* in experience that entitles us to know such things. This, of course, does not mean that modal empiricists should not search for an explanation of our knowledge of essentialist principles, especially because, as it was mentioned in §1, Roca-Royes argues that rival modal rationalist approach is not capable of explaining such a knowledge.

Yet, by using Roca-Royes' account one might come to the conclusion that it is possible for the same object to origin from a slightly different piece of material, which, if true, would falsify the essentiality of origin principle. Namely, one can be justified in believing that it is possible for particular piece of material to be an origin of, let us say, the table called 'Messy', simply because one knows *a posteriori* that its counterpart (in Roca-Royes' sense), twin-Messy, is carved from a very similar, but slightly different piece of material. Given that the two pieces of material stand in the counterpart relation, what holds for the actual origin of twin-Messy also holds for the possible origin of Messy. Therefore, it is not necessary for Messy to be carved from the piece of material from which it actually is carved, and so the essentiality of origin principle fails. Thus, the same methodology that Roca-Royes adopts leads us to the conclusion that origin is not the essential property of an object. It seems that the same holds, *mutatis mutandis*, for other metaphysical principles, so, modal empiricists cannot justify them by means of Roca-Royes' proposal. Perhaps this objection is not sufficient by itself for rejecting Roca-Royes' account of our knowledge of possibility *de re* as such, but it poses an uncomfortable tension inside modal empiricist camp concerning the possibility of providing a unifying account of all *de re* modal knowledge.

4. Conclusion

Let us summarize. Modal empiricism faces serious problems even in explaining our knowledge of possibility *de re*, something which *prima facie*

seems to be the easiest to explain on this view. Roca-Royes' account of our knowledge of possibility *de re* is liable to several counterexamples, and it is also hard to formulate it in a non-redundant way. The whole account is also hard to accommodate with the essentiality of origin principle, to which modal empiricists sometimes appeal while criticizing the modal rationalist account. On the other hand, the cases of possibility *de re* are not by themselves an insurmountable obstacle to modal rationalist approach, for at least some trivial *de re* modal properties are *a priori* knowable. These considerations suggest that modal empiricism can hardly be a good substitute to modal rationalist account of *de re* cases of modal knowledge.[7]

References

BENACERRAF, P. (1973): Mathematical Truth. *The Journal of Philosophy* 70, 661-679.

CHALMERS, D. (2010): The Two-Dimensional Argument against Materialism. In: Chalmers, D. (ed.): *The Character of Consciousness*. Oxford: Clarendon Press, 141-205.

HALE, B. (2013): *Necessary Beings: An Essay on Modality, Ontology, and the Relations Between Them*. Oxford: Oxford University Press.

HOSSACK, K. (2007): Actuality and Modal Rationalism. *Proceedings of the Aristotelian Society* 107, 433-456.

KRIPKE, S. (1972): Naming and Necessity. In: Harman, G. – Davidson, D. (eds.): *Semantics of Natural Language*. Dordrecht: D. Reidel, 253-355.

LEWIS, D. (1979): Counterpart Theory and Quantified Modal Logic. In: Loux, M. (ed.): *The Possible and the Actual*. Ithaca – London: Cornell University Press, 110-128.

MARTIN, C. B. (1994): Dispositions and Conditionals. *The Philosophical Quarterly* 44, 1-8.

Roca-Royes, S. (2007): Mind Independence and Modal Empiricism. *Latin Meeting in Analytic Philosophy Genova*, 117-135.

Roca-Royes, S. (2011): Conceivability and *De Re* Modal Knowledge. *Noûs* 45, 22-49.

Roca-Royes, S. (2014): An Epistemology of *De Re* Modality (talk). Available at: http://media.rcub.bg.ac.rs/?p=3809.

Vaidya, A. (2008): Modal Rationalism and Modal Monism. *Erkenntnis*, 68, 191-212.

van Inwagen, P. (1998): Modal Epistemology. *Philosophical Studies* 92, 67-84.

Why You Can't Actually Imagine the Impossible (but Think that You Can)

LUKE MALIK

Kyoto, Japan
`lukemalikjapan@gmail.com`

ABSTRACT: There are philosophers who think that it is possible to imagine the metaphysically impossible. On the one hand, there are philosophers that think that only knowledge limits what one can imagine. Prior to knowledge of certain facts the imagination is unbounded. On the other hand, there are philosophers who think that the imagination is unconstrained whatever. I shall argue that (a) it is not actually possible to imagine what is metaphysically impossible, though (b) it appears to be possible. I take this to be a defence of the Kripkean view. I aim to develop an understanding of the imagination that can accommodate this view.

KEYWORDS: Exemplification – Goodman – imagination – knowledge – Kripke – mental representation.

1. Introduction

There are philosophers who think that one can imagine what is *metaphysically impossible*. Consider the negation ¬E, where E is, for example, the statement 'water is H_2O'. This kind of statement is a special kind of statement: if it is true, it is necessarily true but if false, necessarily false.[1] Here the modality is metaphysical. The philosophers in question then think that it is possible to imagine the negation of E. Given E is true, these philoso-

---

[1]   Here and below, 'E' abbreviates this kind of statement.

phers think that it is possible to imagine the metaphysically impossible. Such philosophers fall into two camps. We have those who think that what one knows or doesn't know matters to what one can imagine (call them "epistemologists") and those who think that one can imagine the impossible whatever (call them "anarchists").

1.1. The epistemologist view

Alex Byrne writing about philosophers who think that one can't imagine that necessarily true a posteriori statements are false states that it is "more plausible" that such statements "might start out as imaginable but turn unimaginable once the empirical information about [them] comes in" (Byrne 2007, 8).[2] In this way, empirical knowledge has the power to turn a once imaginable statement into an unimaginable one.

Peter Kail argues that Hume holds a similar view. The following explanation is attributed to him: "The explanation, then, of [someone] finding some metaphysically *impossible* state of affairs conceivable rests on his [or her] ignorance of key facts" (Kail 2003, 51). Without adequate acquaintance with these facts, there is no knowledge of them. Without such knowledge, the negation of, for example, a necessary cause-effect relation is conceivable, that is, it is "separable in the imagination" (cf. Kail 2003, 52-53).

1.2. The anarchist view

Peter Kung is one philosopher who has recently argued that one can imagine almost anything: "Imagining impossibilities isn't unusual", he says (see Kung 2010, 626). For Kung, "When we think about imagination in its own right, and aren't biased by philosophical considerations about modal epistemology, it is plausible that we can imagine the impossible" (Kung 2010, 626-627).

Tamar Gendler's claims her story, *The Tower of Goldbach*, allows the reader to imagine the impossible, specifically that 5 + 7 is and is not 12. She writes, that the story shows that the "conceptually impossible proposition that...twelve suddenly ceases to be the sum of two primes becomes – for the moment at least – imaginable" (Gendler 2000, 68).

---

2   But Byrne suggests that he is an anarchist. See Byrne (2007, 2-3).

Kendal Walton seems to lean towards the anarchist position, too. In *Mimesis and Make-Believe* he suggests a player of a game of imagination is able to imagine the impossible. A player of a game in which she imagines a bear every time she sees a tree stump is able to do so, Walton (1990, 26) claims, "even if [the player] believes that the stump *could not* be a bear, that its being a nonbear is one of its essential properties."

These are examples of philosophers who hold that it is possible to imagine the metaphysically impossible. It seems it is something like the common view and it may even be the common sense one. I'll run somewhat against the grain, then, by arguing that it is actually not possible to imagine the metaphysically impossible—regardless of knowledge, or anything else. Since so many philosophers hold the opposing thought, there is some motivation to produce an understanding of the imagination that can accommodate the heretical view. This view I attribute to Kripke and it is the view that I hope to defend. This paper provides a way to understand the imagination and the underlying mechanisms that accommodate the perspective argued for.

To get to the conclusion I wish to draw, we need to be sure that we know what it is that we are talking about. I will start by defining the particular sense of the verb "imagine" I have in mind. The verb seems to suffer polysemy, and this may lead to confusion. So it is important to get the terms of our discussion straight. Second, I will introduce the view that I want to defend, which, I think, can be founded on a reading of Saul Kripke. Third, after some preliminary remarks, which introduce what I assume is the vehicle of imagining, I will introduce some ideas and insights gleaned from Nelson Goodman and Catherine Elgin. Finally, I will use these ideas to outline an understanding of the imagination which accommodates the view championed.

2. The verb "imagine"

The verb "imagine" suffers polysemy. It has several related but distinct meanings and it is difficult to distinguish these systematically. Nevertheless, we need to do what we can to differentiate the primary sense of the verb here from other senses of the verb in order to avoid talking past each other. This need can be made clear by considering the following. The verb "im-

agine" can be synonymous with the verbs "expect" and "suppose". These two terms are semantically distinct. They are, also, grammatically distinct, the former is not always stative, but when it is, it has no imperative form. The latter is stative, and does have a widely used imperative form. What's more there are pragmatic differences. For example, if one knows that P is true, then one ordinarily doesn't *suppose* that P is true, but rather false. On the other hand, if one knows that P is true, one does not rationally *expect* that it is false, but only true. For all these reasons there is a need to differentiate at least these two senses of the verb.

In the context of this essay, not differentiating these senses from the actual sense of the verb that is relevant to our discussion will only cause confusion. I hope to defend the claim that it is not possible to imagine the metaphysically impossible. This claim is attributed to Kripke. But if Kripke was actually using "imagine" synonymously with "suppose", he would have been making an obviously false claim. But, if he was using "imagine" synonymously with "expect", he would have been making another false claim unless modified by knowledge in which case it would have been a trivially true claim. For these reasons, then, it is, also, important to differentiate the use of "imagine", here, from uses which are synonymous with expectation and supposition (and like synonyms).

But we can provide a positive sense to the verb in question. The verb is not just used to say something is expected or supposed, etc. The primary use of the verb that is important here is the one that links thinking of something to thinking of it relative to a sensory modality: That sense of imagining that is thought to provide one with the same kind of information that one would get if one were actually seeing, hearing, smelling, tasting or touching something that one is not actually seeing, hearing, smelling, tasting or touching. This is entirely consistent with Kripke's writings when talking about modality. For example, Kripke (1980, 121) asks us to imagine: animals that *appear* or *look* a certain way. Other examples abound (e.g., Kripke 1980, 112, 114, 118, 128, 142, 150).

Given that it is a Kripkean position about imagining that I am defending, it is not untoward to think that the primary sense of "imagine" and related imaginings relate to thinking of something relative to a particular sensory modality in the way described. This does not necessarily mean all other senses of imagining are excluded, and this is what is meant by talking of the "primary sense" of "imagine" here.

3. Kripkean reasons to think one can't actually imagine
what is metaphysically impossible

A key passage in *Identity and Necessity* sums up a good deal of what Kripke says when he uses the verb "imagine". Here he is talking of a lectern standing in his near vicinity:

> What I am saying is, given that it [the lectern] is in fact not made of ice, in fact is made of wood, one cannot imagine that under certain circumstances it could have been made of ice. So we have to say that though we cannot know a priori whether this [lectern] was made of ice or not, given that it is not made of ice, it is necessarily not made of ice. In other words, if P is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis, some conditional of the form "if P, then necessarily P". (Kripke 1971, 153)

From this passage, and others like it,[3] I draw three claims. First, there is the logical claim: If E is true, it is necessarily true. Second, there is the factual claim: It is the *facts* that determine whether E is true, regardless of how knowledge of the facts is ascertained. Last, the psychological claim: If it is necessarily true that E, then it is unimaginable that E is false. The key claim is the last one: given a metaphysical necessity, one cannot imagine that its negation is true. This is the Kripkean claim to be defended. To do so also involves explaining how it is that we come to *think* that we can imagine the impossible because we just do as often as not think in this way. That is something that Kripke also says something about.

4. Saul Kripke on a certain confusion

Kripke offers us an idea of what kind of confusion is involved when we think we imagine the impossible: One can come to think that E is possibly false (when true) when 'qualitative analogues' are involved. One may confuse a qualitative analogue for the real thing. And one may, thereof, conclude that the negation of E is possible. Suppose E abbreviates 'water is H_2O'. If we discover or imagine a substance that is qualitatively identical to actual water, but not composed of H_2O molecules, then we will have dis-

---

3   Also see Kripke (1980, 46, 47, 112, 113, 114, 126, 127, 129-131).

covered a qualitative analogue of actual water. One may draw the erroneous conclusion that, under the circumstances in question, E is false. That is, that there is a possible situation (or world) in which the negation of E is possible. Kripke's suggests that "We [should] say instead that just as there is fool's gold there could be fool's water" (Kripke 1980, 128). The "fools water" here is a qualitative analogue of actual water, qualitatively indistinct from actual water, but essentially different. To infer the negation of E would be based on confusing fool's stuff for actual stuff.

Talk of qualitative analogues is obscure, however, how they get mixed up in the imagination vague, and, perhaps, even inconsistent (cf. Byrne 2007). However, Kripke does, I think, point us in the right direction; when one thinks one imagines the impossible one is merely confused. I'll try to provide an understanding of the imagination that accommodates this kind of confusion without relying on talk of qualitative analogues. Before we get to this, I briefly introduce what I take to be the vehicle of the kind of imagining in question: the mental image.

5. Mental imagery

I assume that imagining, in the relevant sense, involves mental imagery and that imagery involves a set of mental images. Also, I take these mental images to be representations where that just means they provide information about objects, their qualities and relations. So imagining in the relevant sense involves mental representations that provide information about how objects, their qualities and relations would be experienced under certain circumstances in the absence of external stimuli.

Psychologists today take mental images as (a) literal or functional depictions or (b) descriptions. A debate has sprung up around which of these interpretation is correct (see Kosslyn 1995; 2010; Pylyshyn 1981; 2003; Thomas 2014; Tye 1991). On one side, we have the work of Stephen Kosslyn and his collaborators, on the other, that of Zenon Pylyshyn and his supporters. Kosslyn sometimes argues that mental images are literal depictions realised in the physical matter of the brain (cf. Kosslyn 1995, 290-291; Kosslyn 2010). Mostly though, Kosslyn argues for the less ambitious position that mental images are functional depictions. Functional depictions are not realised in the physical matter of the brain but are encoded in

the way that pictures are encoded. In contrast, description-theorists advocate the view that mental imagery is encoded in the way language is encoded. The difference between the two is one of 'format' (cf. Kosslyn 1995, 280). If mental imagery is depictive, then a mental image of, for example, *a ball on a box* is encoded in an array format – a mapping of the spatial points of relevant objects and their spatial relations to each other. The analogy is to the way information about graphical representations in a computer is stored. If mental imagery is descriptive, then the mental image of a ball on a box is encoded in a 'propositional' format, something like: ⟨on⟨ball, box⟩⟩. That is similar to the way language is encoded in a formal representation of language.

One can remain neutral on what mental images are, whilst noting that mental imagery is taken by both parties to be *representational*. This is all that is needed for the theory presented below. Each format encodes information about an object and its qualities/relations. Thus the mental image in question is representational in the sense that I outlined above.

Note that the view that associates imagining with the possession of mental images that are mental representations is not universally loved. The standard arguments against it can be found in the work of Gilbert Ryle (1951), Jean Paul Sartre (2004) and Ludwig Wittgenstein (1967), and recently Colin McGinn (2006). Rebuttals are found in a number of places.[4]

We can also note that there are alternative accounts. One alternative account is offered by Gilbert Ryle: imagining is a kind of sophisticated pretending (or 'fancying', 'rehearsing', etc.; see Ryle 1951). For example, imagining *a ball on a box* is pretending that one is having the experience that one would have if one were actually seeing a ball on a box (in the absence of the relevant external stimuli). Imagining in this sense is a kind of activity and does not entail the existence or possession of any actual mental imagery. The problem with this is that one can pretend that one is seeing, etc. absolutely anything. The analogy is to supposition. And the reason to dismiss this as irrelevant here is the same reason to dismiss the claim that Kripke is talking about supposition when he talks about imagining: Kripke would have been making *obviously* false claims about the

---

[4]   The interested reader should consult Kosslyn (2010) and Tye (1991).

imagination. So Kripke could not have meant to equate imagining with pretending.

Because the arguments against mental imagery can be rebutted and influential alternative views do not sit well with the relevant Kripkean understanding of imagining (as well as the contemporary scientific understanding of imagining – see Thomas 2014), I will proceed on the assumption that imagining involves mental imagery and mental images, and that these are representation in the sense stated.

6. Nelson Goodman

The next question to ask is how do mental images represent? Nelson Goodman provides us with some important ideas which might be used to make sense of such representations.

According to Nelson Goodman, symbol systems (descriptive, depictive, or otherwise) are composed of symbols which are of two types: labels and samples. There are linguistic labels, like names, predicates, etc. There are theoretical labels like variables, models, etc. There are pictorial labels, like portraits, caricatures, etc. Labels denote. Catherine Elgin explains the relation in the following manner.

> Denotation is a two-place semantic relation between a symbol and the objects to which it applies. A symbol denotes whatever complies with, or satisfies, or is an instance of it. Thus, a name denotes its bearers; a variable, its values; and a portrait, its subject. A predicate denotes severally the objects in its extension. (Elgin 1983, 19)

Samples are also quite numerous and their uses varied:

> The samples we encounter are various, and the uses to which they are put diverse. The model home on a development site, the prototype of a jet plane, and the free bottle of shampoo which arrives in the mail are integral parts of sales campaigns. A sample problem worked out in a text is an illustration of characteristic problems and acceptable modes of solution in a given discipline. And an example of the way you can expect to be treated or the sort of person you are likely to become, may serve as a promise or a threat. (Elgin 1983, 71)

Labels, then, are symbols that denote but samples are symbols that exemplify. Labels denote things. Samples exemplify predicates. Both are forms of reference.

Denotation is relatively easy to understand. Exemplification, on the other hand, is slightly more troublesome. Generally, if a sample, S, exemplifies a predicate F, then F applies to S. But not every predicate that can be applied to S is exemplified. That is, too many predicates may apply to an S. For example, a tailor's swatch may be described as 'threadbare' without, thereby, exemplifying that predicate. To understand something as a sample, then, context, intention and use are essential.

Another problem veers in the opposite direction; there are too few predicates. That is, something may be exemplified which has no corresponding linguistic predicate, a musical note, or a dance move, for example. To solve this problem Goodman and Elgin allow for non-linguistic predicates (see Elgin 1983, 78). In the special cases mentioned, the exemplifying object (note, movement, etc.) may act as its own predicate.

The most serious problem has to do with how to *define* exemplification: if a sample exemplifies F, then the sample is said to refer to F. But what this reference relation is is a matter of dispute. Does it mean that the sample *denotes* F, *exemplifies* F or *refers* to F, or does it have some other reference relation to F, which is not denotative or exemplificational?

Dempster (1989) argues that one is faced with two basic alternatives. A sample, S, exemplifies F iff F denotes S and either (a) S exemplifies F or (b) S denotes F. The former option is circular, so must be rejected. Unfortunately, the latter fares no better. Vermuelen et al., for example, provide ample cause for concern: First, Goodman says that denotation and exemplification are two distinct kinds of reference, the first goes from a label to object, the latter in the other direction. But the definition in question makes it the case there is only one: denotation (unidirectional and bidirectional). Second, the definition makes exemplification symmetrical, so it turns out that the exemplified predicates must themselves exemplify the sample. But, 'the first adjective I will use after 5pm tomorrow' may well exemplify 'predicate' without 'predicate' exemplifying 'the first adjective I will use after 5pm tomorrow', even though the two denote each other by definition (cf. Vermeulen et al. 2009).

Similar objections count against defining exemplification in the other ways mentioned. The following alternative, for example, leads to the same

set of issues: S exemplifies F iff F denotes S and S *refers* to F. But "refer" must mean either exemplify or denote, which just takes us back to the same definition dismissed in the previous paragraph. The last option is just to say there is some other sense of refer in question. However, that has no exegetical support. Although Goodman acknowledges other forms of reference, he mentions none other than denotation and exemplification.

Shouldn't we just give up on the whole notion of exemplification? The thing is that it is a very useful concept. So what is the alternative? According to Vermeulen the concept must be taken as basic. One reason to think that this is the right thing to do is that Goodman does not seek to define denotation or exemplification but takes them as basic forms of reference. Indeed, influential commentators like Elgin proceed in like manner. She stresses that she is following in Goodman's footsteps and that both denotation and exemplification are *basic* forms of reference: "Like Goodman," she writes, "I recognise two basic modes of reference – denotation and exemplification" (Elgin 1995, 66).

All this is no cause for alarm according to Vermeulen. One reason is that every theory involves basic terms that are left undefined. For example, denotation is often left undefined. I'll follow this line. However, following Vermeulen, it can be noted:

(a) The concept is not obscure or out of the ordinary: "In ordinary language, the reference of "man" to Churchill, and of "word" to "man", is unequivocally denotation; while if Churchill symbolises "man", and "man" symbolises "word", the reference is unequivocally exemplification" (Goodman 1976, 58).

(b) Neither does its formal introduction obfuscate: "Denotational reference goes from a label to the objects to which that label applies. Exemplificational reference goes from an object to labels that apply to it" (Elgin 1983, 73).

(c) It is a very fruitful concept. We have already seen the range of its use which includes discussion of: models, prototypes, consumer samples, illustrations, etc. Elgin has also made use of the concept in analysing the sciences (see Elgin 2011). And Goodman applied it to the understanding of art, dance, fiction, expression, metaphor, description- and representation-as, etc. (see Goodman 1976). It can, thus, be considered to be quite fruitfully deployed.

While what has been said has not provided a formal definition, hopefully, the concept is understood as the form of reference that exists between a sample and a predicate that describes it.

Denotation and exemplification are important to Goodman's understanding of how representations represent and what they represent things as (which is important for us below): representation involves denotation; representation-as involves both denotation and exemplification at the same time (cf. Elgin 1983, 141). It is simple enough to think of representations as denoting and as such standing for whatever it is in particular (or general) that they stand for much as names do:

> If a [representation] represents *k* as (or the) soandso, then it denotes *k* and is a soandso [representation]. If *k* is identical to *h*, the [representation] also denotes *h*... To represent the first Duke of Wellington is to represent Arthur Wellesley... (Goodman 1976, 30)

On the other hand: "Description-as and representation-as, though pertaining to labels, are... matters of exemplification rather than of denotation" (Goodman 1976, 66). What this means is that a representation *represents* something *as* something by exemplification: The representation, then, must not only refer to a predicate but that predicate must also denote the representation in question.

This allows a representation to not merely denote what it represents, but to refer to predicates that can be used to label and classify the representation in question. Taking a cue from Elgin (1983, 90), consider the following representation, 'The death penalty is morally inexcusable'. We can say, this representation represents a certain kind of act in that it predicates something of it. The representation exemplifies the predicate 'an-immorally-inexcusable-act-description', which classifies *the representation*. Thus, the act is represented and it is represented as morally inexcusable. But, still, it is a further issue to say whether that act just has been represented as it is; for purveyors of extreme justice may take the contrarian view. That is, they may take the view the representation in question is a *misrepresentation*.

Here is another example. Consider a description that claims dolphins are fish. The description denotes dolphins since it describes or predicates something of them. It exemplifies 'a-fish-description', since that predicate applies to the representation in question. Yet since dolphins fail to fall into

the extension of things that are fish, the representation in question fails to classify dolphins correctly – it is a *misdescription*.

Here is another example (important below). Assume, first, that water is denoted by a representation. Next that the predicate that is exemplified is 'a-water-like-substance-that-lacks-H_2O-representation', so that that predicate denotes the representation faithfully. Since water necessarily does not fall under the extension of this representation, it is just not represented correctly. But we might also say, it is necessarily not represented *as it is*, but rather as something that is water-like but lacks H_2O.

Water does not fall into the extension of things that the representation designates *because it is impossible for water to lack molecules of H_2O*. We don't need to know anything about any object whatsoever *except water* to draw this conclusion. Note it is correct to say that water is represented as something that is, in actual fact, essentially not water without needing to accept that there is any substance that is, in actual fact, watery but lacking in H_2O molecules, i.e. without talk of qualitative analogues. With this in mind, we can make sense of the kind of confusion that Kripke gestures towards.

7. Goodman based confusion

The last case mentioned is the kind of case that is apt to cause confusion. A representation of water that exemplifies the predicate 'a-water-like-substance-that-lacks-H_2O-representation' may be understood as a representation of water *that* lacks H_2O. This representation, then, may be taken to ground the inference that water could have existed without having the molecular make-up that it does actually have. And, even, that there are possible worlds in which water lacks its constitutive identity. This process of reasoning is erroneous. There is a gap between representing water *that* lacks H_2O and representing water as something water-like that lacks H_2O. A representation of water may, indeed, exemplify the predicate 'a-water-like-substance-that-lacks-H_2O-representation', but though and, in fact, just because the predicate is faithful to the representation, the representation in question fails to represent water correctly on the basis that water does not fall into the extension of the representation. What's more, it is necessarily true that water fails to fall into that extension and, thereof, it is necessarily true that water is not represented as it is.

So why is there a confusion? There may be many reasons, but the fundamental reason is that the gap between what is denoted and what it is represented as is obscured. Why? Speculative answers are: just not knowing that something has the molecular structure it has. If I do not know that water is in actual fact necessarily composed of H_2O molecules, then I may draw the conclusion that what is represented is represented as *it* is. Another is not knowing that objects, like water, are identical to their molecular structure rather than other sets of properties. Heat is another example.

8. Why you can't imagine the impossible but think that you can

It is assumed that when one imagines, mental imagery and mental images are involved. Also that these images are representations as defined. Following Goodman, that means they denote and exemplify. And this means that the same kind of story that was told above can be retold below with respect to the imagination.

The kinds of cases that are important to us are those in which familiar items are imagined in a more or less a familiar way, most often when they are imagined as one thinks they would be experienced. When one imagines something and tries to represent one's understanding or experience of it in the way one thinks one would experience it one draws upon familiar predicates. One may, for example, imagine water and try to represent it as one would experience it by drawing upon predicates that are related to one's workaday understanding and experience of *it* (its wetness, transparency, tastelessness, etc.). If so, the associated images or set of images exemplify predicates that classify them as 'a-water-description'. The relevant set of predicates may, unintentionally or intentionally, occlude water's actual molecular structure. If so, as in the example above, the image exemplifies the predicate 'a-water-like-substance-that-lacks-H_2O-representation'. But water is not something that could ever lack H_2O molecules. So it doesn't fall into the extension of the denoted representation. And, thus, water is not represented *as* something it actually is. In fact, it is necessarily not represented as something it is, *even if it is represented as it is ordinarily experienced*.

As above, one may conclude that one has been able to imagine water *that* is not constituted by its actual molecular structure. This may lead one

to infer that one has represented a situation in which water that lacks H_2O could be experienced. And this may even lead to the conclusion that it is possible for the statement that 'water is H_2O' to be false. But, as above, this would be wrong. As said, there is a gap between representing an object and representing it as it is.

This more or less explains just what we wanted to explain: In general, to imagine is to represent. To represent is to denote and exemplify. Denoting k always denotes h on the basis necessarily k is h. There is nothing impossible here. For the representation to exemplify is to classify the representation. To classify the representation determines whether it represents k faithfully or misrepresents it. If it doesn't represent k as h, it necessarily misrepresents h. But there is nothing impossible referred to, here, either, since it cannot follow that k has been represented as k *that* is not h. One can *think* one can imagine the impossible by missing the gap between what is represented and the denoted representation. That is, by confusing the representation of k *as* something that is not h for a representation of k *that* is not h. It is speculated this is based on epistemic frailty or misunderstanding the relationship between objects and phenomenal properties. Objections and further examples will be considered below.

9. Objections

Consider another set of statements: E abbreviates 'cats are animals' and F abbreviates 'cats are demons'. Knowing that E is true, one knows that E is necessarily true. Nevertheless, one may understandably think that one can still imagine a situation in which F is true. One may think, for example, that one can imagine a situation in the future when we find out that E is false and F is true. Doesn't it follow from this that I can imagine the impossible? After all, F must be impossible given E is true. All this can be explained in exactly the same way as the water case was explained. To have a mental image of a cat that represents a cat as being something that is less critter than fiend is not to have a mental representation of a cat that represents a cat *that* is a demon. The representation in question if a representation of a cat, given E, denotes an animal since it is necessarily true cats *are* animals. That representation may represent its object in a certain manner and in that manner exemplify the predicate 'a-cat-like-entity-that-is-a-

demon-representation', which correctly classifies the representation, but a representation that necessarily misrepresents the facts. Such a cat representation necessarily fails to represent its object as it is. It does not follow that a cat that is a demon has been represented nor that a cat has been represented as a cat that is a demon.

But, still, Putnam *and* Kripke suggest something possible just is imagined when one imagines 'water is not H_2O', 'cats are demons', etc. This is where "textbook Kripkeanism" takes up the story (see Yablo 2000). Textbook Kripkeans say the sense in which F could turn out to be true (or could have turned out to be true) is logical or conceptual. The possibility associated with F, here, understood as a linguistic representation, though, is the possibility associated with *using F to say something true rather than the possibility associated with what F actually says!* That former kind of possibility is primarily associated with a situated epistemic reality but, nevertheless, presents one with a metaphysical possibility, though that possibility is not to be confused with the actual metaphysical possibility associated with the statement. The theory presented can be given a similar sheen. The representation discussed above represents a cat (and, thus, an animal), but represents it as something that it is not (e.g. as a demon). All this is true whether or not there is something or nothing in the extension of the representation denoted. Something seems possible, though, right? But this latter possibility is not the same as the metaphysical possibility of F being true. It is the possibility of the denoted representation representing something true in some situation, e.g. representing something less critter than fiend experienced as we experience actual cats.

Another objection may seem to come from an interpretation of David Chalmers. Chalmers provides us with a theory of conceivability. There are various kinds of conceivability some of which involve the imagination. Chalmers talks about two central kinds. First, we have 'primary positive conceivability' and, second, 'secondary positive conceivability'. Positive conceivability is another name for imagining (see Chalmers 2002). Thus, we have what can be called 'primary imaginings' and 'secondary imaginings'. Imagining that 'water is not H_2O', 'cats are demons', etc. are cases of primary imagining. To say that one cannot imagine these things is to talk about secondary imagining, one cannot imagine that 'water is not H_2O', 'cats are demons', etc. It seems that one 'passes' from the former to the latter once one is endowed with the actual facts. In any case, primary imagin-

ing is a kind of imagining, and if 'water is not H_2O', etc. is what one imagines when one imagines that such statements are true, the impossible really is being imagined.

But is this a good reading of Chalmers? Chalmers is a textbook Kripkean (see Yablo 2000). The key textbook Kripkean move is to say that when one thinks one can imagine the possibility of an impossibility one is confusing a different *metaphysical* possibility for the actual possibility in question. We've already seen how this move can be accommodated by the theory being defended here. It is also possible to see how it is accommodated in Chalmers theory. One may say one imagines that water is not H_2O in the primary sense (based on the conceptual relations), which makes us think we can imagine the impossible. One may come to see that one cannot imagine that, that is, in the secondary sense, when the facts are in. However, when we consider the metaphysical possibilities associated with each kind of imagining, we see, on Chalmers view, since there is only one set of worlds, such that those worlds are not distinguished at the primary and secondary level (cf. Chalmers 1996), that the conceptual possibility associated with imagining that 'water is not H_2O' is just the metaphysical possibility associated with what one can imagine being true of *some* state of affairs (i.e. there being a watery substance that is not H_2O) and not *the* actual state of affairs. This explains the metaphysical possibility that the conceptual possibility associated with a sentence such as 'water is not H_2O' etc. implies. And similar explanations can be given for sentences like 'cats are demons', etc. The related conceptual possibilities, then, do register the *metaphysical* possibility of something just not the metaphysical possibility associated with actual water, cats, etc. The same can be said of primary imaginings. They do register the metaphysical possibility of something but just not those associated with actual water, actual cats, etc.

It seems to me that if this line of reasoning is not available to Chalmers, it is difficult to see just how his Conceivability Argument for Dualism (or his defence of Jackson's Knowledge Argument) is supposed to work (see Chalmers 1996; 2004). Contrast two primary imaginings: 'water is not H_2O' and 'pain is not the firing of C-fibres'. For the physicalist either is impossible. But Chalmers wants to show that the latter is metaphysically possible and he wants to do so by showing that he is justified in *moving from a primary conceivability to a secondary conceivability*. So he has it that we can primarily imagine both (based on not knowing enough). When the

facts arrive (and we do know enough), we realise that we cannot imagine such cases and, thus, *that we were not imagining what we thought we were imagining*. The conceptual possibility is just that of being able to think that 'water is not H_2O' based on the primary imagining. But what that conceptual possibility turns out to be is just the possibility of being able to say that 'water is not H_2O' based on the metaphysical possibility of there being some water-like substance that is not H_2O. It is now available for Chalmers to say that line of reasoning is not available to the physicalist when the primary imagining associated with imagining pain that is not the firing of C-fibres is concerned. To say that one can primarily imagine that pain is not the firing of C-fibres where that primary imagining points to a conceptual possibility where that possibility is the possibility of saying that the sentence in question can be used to say something true is all well and to the good. But when we look for what the related metaphysical possibility is, that is, what the statement in question could be used to say that is metaphysically possible, it turns out that it could not be used to say anything but pain is not the firing of C-fibres since there is nothing else pain could plausibly refer to: there is no possibility of pain-like phenomenon that is not pain as there is of a water-like substance that is not water (or a cat-like fiend that is not a critter, etc. – see, e.g., Chalmers 1996, 67-69, 73, 146-149). We have come to the conclusion that all the sentence, in fact, could say is just what it does say. And, thereof, mind-body materialism is false.[5]

Last, note Chalmers (1996, 146) introduces Kripke's argument against mind-body identity theory in exactly the manner laid out and notes his arguments affinities to Kripke's argument with respect to their shared modal and logical core.

One last objection, what if we don't know which of E and F is necessarily true and which necessarily false. In such a case, we surely seem to be able to imagine either one is true. If so, surely that is to imagine the impossible. Say, now, E = 'Goldbach's Conjecture is proved', and F = 'Goldbach's Conjecture is disproved'. We may think we can primarily imagine either. But, according to Chalmers (1996), if E actually turns out to be true, then F will have been a misdescription of a possible world. And, so, if there is a conceptual possibility associated with F, it will be the possibility asso-

---

[5]   Strictly speaking, Chalmers only draws the conclusion that the present form of materialism is false.

ciated with F being used to say something true and that might be, for example, the metaphysical possibility of a mathematical outcome that involves a miscalculation; *not the negation of E*. On the theory presented here, Goldbach's Conjecture will have been represented, thus, its true status will have been represented. But it just will have exemplified the predicate 'a-Goldbach-Conjecture-is-false-representation'. The denoted representation will, then, have misrepresented Goldbach's Conjecture. It will have represented it as something it is not. It does not follow that a mathematical truth *that* was false was represented or anything was represented as a mathematical truth *that* was false.

10. Conclusion

It has been claimed that one cannot imagine the impossible because the imagination involves mental images, that they represent objects and their qualities/relations, and that the relevant images can be analysed as Goodman analyses representation. One can *think* one can imagine the impossible by missing a gap between what is represented and what it is represented as. Reasons for such confusion may be due to epistemic frailty or misunderstanding the relation between objects and phenomenal properties. Epistemologists are wrong, then, to say knowledge regulates the imagination and anarchists are wrong to think nothing constrains the imagination.

References

BYRNE, A. (2007): Possibility and Imagination. *Philosophical Perspectives* 21, No. 1, 125-144, available at: http://web.mit.edu/abyrne/www/possandimag.pdf

CHALMERS, D. (1996): *The Conscious Mind*. Oxford: Oxford University Press.

CHALMERS, D. (2002): Does Conceivability Entail Possibility. Available at: http://consc.net/papers/conceivability.html

CHALMERS, D. (2004): Phenomenal Concepts and the Knowledge Argument. In: Ludlow, P. – Nagasawa, J. – Stoljar, D. (eds.): *There's Something about Mary: Essays on Frank Jackson's Knowledge Argument against Physicalism*. The MIT Press.

DEMPSTER, D. (1989): Exemplification and the Cognitive Value of Art. *Philosophy and Phenomenological Research* 49, 393-412.

ELGIN, C. E. (1983): *With Reference to Reference*. Indianapolis: Hackett Publishing Company.

ELGIN, C. E. (1995): Metaphor and Reference. In: Radman, Z. (ed.): *From a Metaphorical Point of View. A Multidisciplinary Approach to the Cognitive Content of Metaphor.* Berlin – New York: de Guyter, 53-72.

ELGIN, C. E. (2011): Making Manifest: The Role of Exemplification in the Sciences and the Arts. *Principia: An International Journal of Epistemology* 15, No. 3, 399-413.

GENDLER, T. S. (2000): The Puzzle of Imaginative Resistance. *The Journal of Philosophy* 97, No. 2, 55-81.

GOODMAN, N. (1976): *Languages of Art.* Indianapolis: Hackett Publishing Company, Inc.

KAIL, P. (2003): Conceivability and Modality in Hume: A Lemma in an Argument in Defence of Sceptical Realism. *Hume Studies* XXVIX, No. 1, 43-61.

KOSSLYN, S. M. (1995): Mental Imagery. In: Osherson, D. N. – Kosslyn, S. M. (eds.): *An Invitation to Cognitive Science: Visual Cognition. Vol. 2.* The MIT Press.

KOSSLYN, S. M. ET AL. (2010): *The Case for Mental Imagery.* Oxford: Oxford University Press.

KRIPKE, S. (1971): Identity and Necessity. In: Munitz, M. K. (ed.): *Identity and Individuation.* New York: New York University Press, 135-164.

KRIPKE, S. (1980): *Naming and Necessity.* Harvard (Mass.): Harvard University Press.

KUNG, P. (2010): Imagining as a Guide to Possibility. *Philosophy and Phenomenological Research* 81, No. 3, 620-663.

McGINN, C. (2006): *Mindsight: Image, Dream, Meaning.* Cambridge (Mass.): Harvard University Press.

PUTNAM, H. (1962): It Ain't Necessarily So. *The Journal of Philosophy* 59, No. 22, 658-671.

PYLYSHYN, Z. W. (1981): The Imagery Debate: Analogue Media Versus Tacit Knowledge. *Psychological Review* 87, 16-45.

PYLYSHYN, Z. W. (2003): Return of the Mental Image: Are There Pictures in the Brain? *Trends in Cognitive Sciences 7*, 113-118.

RYLE, G. (1951): *The Concept of the Mind.* Watford, Herts: The Mayweather Press.

SARTRE, J. P. (2004): *The Imaginary: A Phenomenological Psychology of the Imagination.* Webber, J. (Trans), London – New York: Routledge.

THOMAS, N. J. T. (2014): Mental Imagery. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Available at http://plato.stanford.edu/archives/fall2014/entries/mental-imagery/.

TYE, M. (1991): *The Imagery Debate.* Cambridge (Mass.): The MIT Press.

VERMEULEN, I. – BRUN, G. – BAUMBERGER, C. (2013): Five Ways of (Not) Defining Exemplification. In: Ernst, G. – Steinbrenner, J. – Scholz, O. R. (eds.): *From Logic to Art: Themes from Nelson Goodman.* Berlin: De Gruyter.

WALTON, K. (1990): *Mimesis as Make-Believe: On the Foundations of the Representational Arts.* Cambridge (Mass.): Harvard University Press.

WITTGENSTEIN, L. (1967): *Zettel.* Anscombe, G. E. M (trans.), von Wright, G. H. (ed.). Oxford: Blackwell.

YABLO, S. (2000): Textbook Kripkeanism and the Open Texture of Concepts. *Pacific Philosophical Quarterly* 81, No. 1, 98-122.

Quine and Quantified Modal Logic:
Against the Received View

ÁDAM TAMAS TUBOLY[1]

Department of Philosophy. Faculty of Humanities. University of Pécs
Ifjúság Street 6. Pécs 7624. Hungary
tubolyadamtamas@gmail.com

ABSTRACT: The textbook-like history of analytic philosophy is a history of myths, received views and dogmas. Though mainly the last few years have witnessed a huge amount of historical work that aimed to reconsider our narratives of the history of analytic philosophy there is still a lot to do. The present study is meant to present such a micro story which is still quite untouched by historians. According to the received view Kripke has defeated all the arguments of Quine against quantified modal logic and thus it became a respectful tool for philosophers. If we accept the historical interpretation of the network between Quine, Kripke and modal logic, which is to be presented here, we have to conclude that Quine's real philosophical animadversions against the modalities are still on the table: though Kripke has provided some important (formal-logical) answers, Quine's animadversions are still viable and worthy of further consideration.

KEYWORDS: History of analytic philosophy – quantified modal logic – Saul Kripke – Willard van Orman Quine.

Modality and especially (quantified) modal logic (abbreviated as QML) has played an important role in twentieth-century analytic philosophy. In the early years it was considered suspicious, sometimes even meaningless. Among the main proponents of this skeptical attitude we find Willard van Orman Quine, whose considerations against the modalities (or any intensional items) date back to his doctoral dissertation (1932/1990), when he "reworked the foundations of *Principia Mathematica* in strictly extensional terms, and propositional functions to the winds" (Quine 1990/2004, 55). His very first explicit 'attack' directed against modal logic is to be found in his contribution to the Schilpp volume of A. N. Whitehead (Quine 1941, 141-142, especially n26), where he argued that the modal operators violate certain basic rules of extensional predicate logic, therefore it is dubious whether the modal notions could be integrated into the regimented language of science (a language based on extensional first-order predicate logic with identity) which was one of the few conditions of meaningfulness.

This kind of skepticism of Quine surfaced in most of his writings from the 1940s until his death, and posthumously published *Confessions of a Confirmed Extensionalist and Other Essays*.[2] One of his major problems was the following reduction-challenge: "There are logicians, myself among them, to whom the ideas of modal logic (e.g. [C.I.] Lewis's) are not intuitively clear until explained in non-modal terms" (Quine 1947, 43). Besides empiricism, the source of this conception is that Quine (2001/2008, 500) used to mark the boundaries of meaningfulness around classical extensional predicate logic with identity: "I doubt that I have ever fully understood anything that I could not explain in extensional language."

Though only a minority of logicians and philosophers (like Ruth Barcan Marcus, Rudolf Carnap, Alonzo Church and Arthur Smullyan) have responded explicitly to Quine's animadversions, QML gradually became a respectful tool after the 1960s and one could hardly imagine any metaphysical or linguistic analysis that was not using the modal vocabulary as a respectful tool.

As the typical text-book-like story goes, it was Saul Kripke and his possible-worlds semantics of modal logic that played a crucial role in

---

[2]   See Quine (2008); the relevant works are Quine (1941; 1943; 1947; 1951; 1953; 1953/1966; 1960, chapter 6; 1963/1966; 1977; 1992/2008; 2001/2008).

changing the philosophical atmosphere. In the 1980s, Hintikka formulated this point according to the lines of the familiar story:

> The possibility of a reasonable modal logic was denied by Quine on philosophical grounds, but *his objections have been dead* for a while, even though they have not yet been completely buried. What has made a crucial difference is the development of what has generally been taken to be a *viable semantics (model theory) for modal logic*. This semantics has provided a basis from which Quine's objections can apparently *be answered satisfactorily* and which yields a *solid foundation* for the different axiom systems for modal logic. (Hintikka, 1982/1989, 1)

Recently Joseph Melia has framed this idea in a very similar way:

> We shall see that such worries [Quine's worries] were largely misguided and that the possible worlds machinery provides us with the conceptual tools to see off all such objections. (Melia 2003, 63)

> We have examined Quine's arguments against modal logic, and we have found them all lacking. The defenders of modal logic, be it propositional or predicative, have nothing to fear from Quine and are quite within their rights to take modal truths and modal logic seriously, and to search for a respectable theory of modality. (Melia 2003, 79)

I will claim that one part of Hintikka's report was adequate, but the other hasn't captured the force and aims of Quine's original animadversions. The adequate part of the claim is that Kripkean semantics indeed yields a solid foundation for the various axiom systems of modal logic – in some of his less known papers Quine has also accepted the formal results of Kripke's investigations. But the other claim of Hintikka – which is just the received, text-book-like story about Quine and QML – that says that Kripkean semantics has *fully answered* the Quinean arguments against the meaningfulness of the modalities cannot be maintained in such a simple form. We can detect various textual evidences where both Kripke and Quine state that the Kripke-semantics in itself is *just* mathematics, a formal investigation which cannot yield *philosophically interesting* results, i.e. it is not the required interpretation that Quine has demanded. Though there seems to be a consensus about the major theses of Kripke's philosophy (about rigid designation, natural kinds and metaphysical necessity) the field is still open for further debate.

By bringing together these historical claims we can uncover a highly different story of the twentieth-century analytic philosophy and modal logic. We can also show – though I will do no more than sketch the main idea – that the later philosophical works of Kripke and David Lewis could be integrated into this picture – they both provide the required form of intended or philosophical *interpretation* that Quine was demanding, though neither of them is acceptable for Quine for various reasons.

1. Methodological considerations

The present study summarizes partly the recent ideas about Quine and QML. Though one can still find some further "received-view-like" claims about these issues, in the last few years scholars like John Burgess (1998/2008; 2013), Stephen Neale (2000), Kit Fine (1989; 1990), Zsófia Zvolenszky (2006; 2007; 2010), Roberta Ballarin (2004) and earlier David Kaplan (1969; 1986) did a lot to rewrite the story and to reorient our thinking.

Taking a special approach to history and historiography we can distinguish three approaches: (i) systematic-narrative, (ii) argumentative, (iii) micro-historical. The first one results in a certain *history of ideas*, in a comprehensive and systematic *narrative*, usually based on a singular point of view, which tries to reconstruct the given story on a unified line. A typical example is that analytic philosophy is united by the *ideas* of the linguistic turn. The second embraces an *argumentative* point of view according to which we have to investigate the explicit statements and arguments of the individual thinkers, taking them to be hypothetical participants in our contemporary philosophical debates (this method is used by Scott Soames in his *Philosophical Analysis in Twentieth Century*). Such an approach is much less sensitive to the contextual factors offered by a narrative account. It is receptive instead to bare claims and arguments, lifted from their historical contexts. These have to hold up by themselves, facing the contemporary critics in our current space of reasons.

The third (iii) approach makes a far more moderate claim – it does not want to argue directly with the figures depicted *as* contemporaries, nor to generate a systematic and unifying *narrative*. Instead it sees its task in writing certain well defined *micro histories* – in such a suitably restricted story

the heroes can be certain thinkers and their oeuvre which are unknown and isolated from the point of view of the 'big picture', various scientific controversies and disputes and their participants, or the reconstruction of the (un)published works in the light of their reception. Erich Reck formulates a similar point of view:

> [...] we carefully recover – using tools borrowed from history, philology, as well as from philosophy (archival research, close textual exegesis, and attention to context) – what the philosopher's core concepts, basic assumptions, and main project *actually* were. We also refrain, at least initially, from evaluating the recovered ideas by using current standards. What we do, on the other hand, is to think them through *internally*, i.e., to evaluate them by using the standards and the understanding of the time. The latter is what makes the approach *philosophical* (not just 'historical reconstruction *as history*', but 'historical reconstruction *as philosophy*'). It is also what makes it a form of 'historicism', albeit a relatively modest one. (Reck 2013, 12)

The revisionary points of the aforementioned scholars are connected to approaches (i) and (ii) thus I will formulate my points on the base of their results. Some more quotations, however, from the primary materials are also provided to strengthen their claims from a more historical point of view. I shall join their efforts by calling attention to certain micro-stories regarding Quine, Kripke and the impact of their work and explaining some features of the received view on the basis of these considerations.

Considering logical and philosophical textbooks along with the relevant secondary literature from the second half of twentieth century, we can abstract a certain *received* or *official view* about Quine and QML. Talking about the 'received view' does not mean that there really is a unified story about Quine and QML which is shared by everyone working on this topic – it is only a useful formulation of certain well known ideas, and it has a certain heuristic value for the later argumentation.

The received view can be characterized with various dogmas as shown by Burgess (1998/2008) but I will restrict my focus and will discuss only two of them:[3]

---

[3] For a much comprehensive list see the groundbreaking work of Burgess (1998/2008).

(RV1) QML does not commit one to essentialism, since the essentialist theses cannot be derived in the usual systems.

(RV2) The (model-theoretic) possible-worlds semantics pointed out that QML could be interpreted in a meaningful way; therefore Quine's critique (that QML is meaningless) has been defeated.

2. Quine and essentialism

According to (RV1) the essentialist-charge states that QML is committed to the untenable view of *essentialism*. Quine claimed that essentialism comes into play when one tries to combine, for example, the notion of necessity to open sentences like: $(\exists x)\Box(x > 7)$. At this point Quine was mainly concerned with the concepts of *linguistic* (or *verbal*) *modalities*. Focusing on necessity, linguistic necessity, as Zvolenszky has put it (following Burgess), covers both analytic and logical necessity: "for both concern truth in virtue of the meanings of certain expressions; the difference is only whether we consider the meanings of all vocabulary items or just the logical ones" (Zvolenszky 2010, 43).

Having this in mind, Quine's problem could be formulated as follows. If one tries to substitute for the variable x 'nine' one gets that 'necessarily nine is greater than seven' which is true. On the other hand if one substitutes for x 'the number of planets' one gets 'necessarily the number of planets is greater than seven' which is presumably false. Given that nine *is* just the number of planets, the two expressions are coreferring, and hence substituting them for the same variable should yield sentences with the same truth-value.

In some of his writings Quine concluded that in certain contexts it *does* matter how we name things, that is, the form of the names. As he says:

> An occurrence of the name in which the name refers simply to the object designated, I shall call *purely designative*. Failure of substitutivity reveals merely that the occurrence to be supplanted is not purely designative, and that the statement depends not only upon the object but on the form of the name. For it is clear that whatever can be affirmed about the *object* remains true when we refer to the object by any other name. (Quine 1943, 114)

For Quine, the not purely designative constructions are the typical cases of the modalities: while the object, the number nine (in and of itself) under the name 'nine' is necessarily greater than seven, under a different name, for example, 'the number of planets', it is not necessarily greater than seven, be it either the case that it is not *analytic* that "the number of planets is greater than seven" or that it is not a *logical* truth.

But one of the most important metaphilosophical commitments of Quine is that he reads the quantifiers *objectually*. The objectual interpretation of the quantifiers means that the admissible values of the variables bounded by the quantifiers are *objects* simpliciter. Quine used to call this interpretation the *ordinary sense* of quantification, "corresponding to the sense of the parallel pronominal constructions of ordinary language."[4] Given this reading of the quantifiers it seems that they are neutral with respect to the form of the names of the objects since they are concerned with the values of the variables which are objects *simpliciter*.

Recall Quine's (1953/1966) famous three grades of modal involvement: in the first grade we attach the modal operators to sentences as *metalinguistic predicates*; in the second we attach the modal notions only to *closed formulas* – it is parallel with the use of negation; but in the third grade the modal operators are capable of attaching to *open formulas*. When we are concerned with open sentences the primary elements are just the *objects*, hence we attach the modal items (for example properties as "necessarily being such-and-such") to the *objects*.[5]

Now we would like to say that there is an object x (namely the number nine) which is necessarily greater than seven, and there is exactly x planets and it is not necessarily that there is x planets. We should be able to say that the following sentence (even using a hybrid language) is meaningful:

$$(\exists x)[\Box(x > 7) \land \text{ there are just } x \text{ planets} \land \neg\Box \text{ (there are just } x \text{ planets)}]$$

and in general that

---

[4]   Quine's letter to Carnap, QC/105/1943-5-7, in Creath (1990, 328).

[5]   It is important to note that we are dealing with the objectual interpretation since, as Quine emphasized, "if [...] we do not have quantification in the old sense, then I have nothing to suggest at this point about the ontological implications or difficulties of modal logic" (Quine's response, in Marcus *et al.* 1962/1993, 27).

$(\exists x)(\Box Fx \wedge Gx \wedge \neg\Box Gx)$.

According to our earlier example, it is necessary that nine is greater than seven and it is true that the number of planets is nine but it is not necessary that the number of planets is also nine, since it could have been five (if there were only five planets). Quine thought that in QML "quantifying in" could be meaningful only if we distribute modal properties with respect to the *form* of names (cf. Quine 1953, 147-148). That is, while under the label '9' it is necessary that nine is greater than seven, the same holds for the label '5+4', but under 'the number of planets' it seems that it is not. Under certain descriptions things bear their properties necessarily while under different descriptions they do not.

But regarding the quantifiers, in order to do their job we shall not deal with the mode or form of naming: "quantification, ordinarily understood, abstracts from the mode in which objects are designated."[6] The truth of a formula like '$(\exists x)(x > 7)$' depends on the existence of an *object* which is greater than seven and not on the naming of things. On Quine's view this means that we do not have to tie the accidental and necessary properties to the linguistic descriptions of objects (or to the meanings) but to the objects *themselves*.

So the tension in QML can be formulated as follows: according to the usual objectual interpretation of the quantifiers the mode of naming is not important, while the modalities, when we try to understand them as *verbal*, or *linguistic* modalities, are to be attached to the names (or generally to the linguistic considerations) of the objects. Therefore QML has to take account of the modes of naming and leave them out of consideration *at the same time*. Since for Quine the usual quantificational strategies have priority, we have to settle with the latter option, namely to leave behind the form of the names. But this means that we have to attach the modalities directly to the *objects* and hence we have to say that the number nine in and of itself bears some of its properties necessarily and some of them accidently.[7]

---

[6]   The formulation is from Leonard Linsky who owes it to one of his students. See Linsky (1969, 695, n10).

[7]   Quine considered the possibility of reading the quantifiers, for example, in an intensional way (as, according to him, Carnap did in his *Meaning and Necessity*) but he

This view is usually called (Aristotelian) essentialism:

> This is the doctrine that some of the attributes of a thing (*quite independently of the language in which the thing is referred to, if at all*) may be essential to the thing, and others accidental. (Quine 1953/1966, 173-174; italics added)

> This is how essentialism comes in: the invidious distinction between some traits of an object as essential to *it* (by whatever name) and other traits of it as accidental. (Quine 1962/1966, 182)

> An object, of itself and *by whatever name or none*, must be seen as having some of its traits necessarily and others contingently, despite the fact that the latter traits follow just as analytically from some ways of specifying the object as the former traits do from other ways of specifying it. (Quine 1953/1961, 155; italics added)

This kind of essentialism was incompatible with Quine's (1953/1961, 156) scientific naturalism: "Such a philosophy is as unreasonable by my lights as it is by Carnap's or [C. I.] Lewis's." According to Quine the existential questions are to be dealt with by science and scientific inquiries: since philosophy is not in a position to answer these questions *science* has to inform us whether there are essential properties or not.[8] Inasmuch the latter does not claim that there are such things, their conception is just plain metaphysics for Quine (1953/1966, 174): "[QML] leads us back into the *metaphysical jungle* of Aristotelian essentialism."[9]

---

claimed that "if you take quantification in some such new sense, you depart from the topic of my article" QC/105/1943-5-7, in Creath (1990, 328).

[8]   About Quine's naturalism see Gregory (2008) and Weir (2014).

[9]   There is an interesting parallel between Neurath and Quine on this issue. As Quine used to attack Carnap's semantics and considerations of modal logic in the 1950s, Neurath has argued against Carnapian semantics as early as the 1930s but mainly in the 1940s after the publication of Carnap's *Introduction to Semantics*. In a letter to Carnap Neurath said that "I am really depressed to see here all the Aristotelian metaphysics in full glint and glamour, bewitching my dear friend Carnap through and through" (Neurath to Carnap, January 15, 1943, ASP RC 102-55-02). For Neurath semantics in Carnapian (and Tarskian) style entails metaphysics and thus it undermines empiricism; he thought that "it [is] rather dangerous to speak of the DESIGNATUM of an expression" and therefore tried to reformulate the semantical concepts in a behavioristic way and

We can reconstruct a typical argument from the secondary literature which tried to capture the main points of Quine's considerations (cf. Parsons 1967):

(P1)   QML is committed to essentialism
(P2)   Essentialism is an untenable, incoherent view.
(C)   Therefore QML is untenable and incoherent.

This argument did not strike Quine as having a devastating consequence:

> [...] in conclusion I say, as Carnap and [C. I.] Lewis have not: *so much worse for quantified modal logic* as well; for if we do not propose to quantify across the necessity operator, the use of that operator ceases to have any clear advantage over merely quoting a sentence and saying it is analytic. (Quine 1953/1961, 156, italics added)

Philosophers in the 1960s used to attack the first premise of this argument pointing out that QML is not committed to essentialism. The very first defender of QML against the essentialist-charge was Ruth Barcan Marcus (Marcus 1967/1993), who was followed by Parsons (1967; 1969) and Linsky (1969). They have argued that QML is committed to essentialism only in a very *trivial* sense; we can prove, for example, statements that claim self-identity, or that everything bears a certain property or not – that is we are committed in a modal system only to *non-substantial*, *trivial* essentialist claims like, everything is necessarily red or not-red (they are, so to say, logico-essentialist claims) but we cannot prove such statements as 'something is essentially a mathematician' (see Marcus 1962/1993). This trivial sense of essentialism, as Dagfinn Føllesdal has claimed,

> is no more objectionable than the modal operators themselves, when applied to closed formulae. In the case of other types of nonextensional contexts [...] the corresponding notions are required in order to permit quantification into such contexts. So, in this extended sense of 'essentialism', we are all essentialists. (Føllesdal 1986/1998, 104)

---

claim that "I suggest to speak of an ACCEPTED SENTENCE (or of a designating sentence), instead of a denotatum I suggest to speak of an 'acknowledged' expression [...]" (Neurath to Carnap April 1, 1944, ASP RC 102-55-05).

Marcus and Parsons tried to strengthen their arguments by pointing out certain mistakes in Quine's argumentation. Parsons argued, for example, that in the reconstruction of Quine's argument we find an equivocation with respect to 'essentialism'. In (P1), according to Parsons, we have to consider only a trivial essentialism, while in (P2) we find a different and stronger sense of it which cannot be proven in QML. Therefore Quine's argument simply fails to reach the required conclusion.

Nonetheless a more subtle reconstruction of Quine's claims shows that essentialist-charge could be upheld even in the face of the Marcus-Parsons arguments. Dagfinn Føllesdal (1986/1998) argued for the thesis that in the writings of Quine we have to deal with two different essentialist-charges. On the one hand there is a certain *weak-essentialist-charge* (WEC) and a *strong-essentialist-charge* (SEC):

> The first, weak notion was developed in response to Carnap, Lewis, and others, who championed quantified modal logic while at the same time rejecting as metaphysical nonsense the traditional Aristotelian view *that necessity inheres in things and not in language*. [...] Quine saw that Carnap and Lewis's linguistic conception of necessity was untenable if one wants to quantify into modal contexts, and that their position therefore was incoherent. (Føllesdal 1986/1998, 98; italics added)

According to (WEC) we have to make a distinction between a thing's necessary and accidental properties and this distinction inheres not in the way we talk about things but *in the things themselves, in and of themselves*. Quine claims that if we want to interpret QML, then we must tie essentialism to the things but do not have to say anything about what could be the essences. At this point there is no need to commit oneself to *any non-trivial* essential property, what matters is only the distinction and that it inheres in the things, not in the language.

Essentialism, however, got a stronger sense, (SEC). It says that things have in themselves *individual, unique essences*; it says that there are certain essences which belong only to one individual: *x* has a certain *x*-ness, which is *x*'s essence. (SEC) appears, for example, when we are dealing with trans-world identity, and we don't have to consider it here. Føllesdal (1986/1998) states that for Quine in the 1940s-1960s what was at stake is (WEC), but after that in the 1970s Quine was putting forward (SEC) and hence causing certain interpretational confusions for others.

What is more important is that Marcus and Parsons haven't met the Quinean challenges. The reason is that Quine, already in 1962 at the Boston Colloquium, has made his arguments explicit; i.e. he is not talking about proving essentialism in QML:

> I've never said or, I'm sure, written that essentialism could be proved in any system of modal logic whatever. I've never even meant to suggest that any modal logician even was aware of the essentialism he was committing himself to, even implicitly in the sense of putting it into his axioms. *I'm talking about quite another thing – I'm not talking about theorems, I'm talking about truth, I'm talking about true interpretation.* And what I have been arguing is that if one is to quantify into modal contexts and one is to interpret these modal contexts in the ordinary modal way and one is to interpret quantification as quantification, not in some quasi-quantificatory way that puts the truth conditions in terms of substitutions of expressions, then in order *to get a coherent interpretation* one has to adopt essentialism and I already explained a while ago just how that comes about. *But I did not say that it could ever be deduced in any of the S-systems or any system I've ever seen.* (Quine's response, see Marcus et al. 1962/1993, 32; italics added.)

Earlier in the discussion Marcus argued that we cannot prove any disturbing essentialist claims in QML, thus Quine's arguments fail against QML. Quine pointed out two things in his response. He said that he wasn't talking about proving essentialism in the systems of QML. From this point of view the articles of Marcus and Parsons pointed out rightly that such theorems are not available in QML, but they miss the point. Keeping this in mind, it is even stranger that though Quine made these claims in 1962, Marcus and Parsons in their published articles at the end of the 1960s were attributing to Quine a non-Quinean position.

On the other hand, Quine did not specify the list of essential properties and the non-trivial essences have not appeared at this point either. They will be relevant only after the 1970s and not in the sense of provability – Quine has localized the essentialist commitments in the *use* of QML. Since the language of QML is stronger than that of classical logic, we can formulate in it certain statements that cannot be stated in classical logic. That is, when we are talking in the language of QML and *using* modal operators in the formalization of statements of ordinary language, we can formulate sentences whose interpretation shows the signs of essentialism. Modal logic

commits us to essentialism when we are using it to *express* ourselves and not at the level of provability and deducibility. This problem is closely tied to what Quine has called 'true interpretation', but to discuss it, we have to move on to the second point of (RV).

3. Quine and interpretation

(RV2) states that Quine's arguments against QML have been defeated mainly due to Kripke's model-theoretic possible-worlds semantics. One of the main reasons behind Quine's animadversions against QML was that it violates the rules of extensionality.[10] Kripke was well aware of the usual extensionalist tendencies of his time and tried to motivate his approach from that point of view:

> It is noteworthy that the theorems of this paper can be formalized in a *metalanguage* (such as Zermelo set theory) which is 'extensional', both in the sense of possessing set-theoretic axioms of extensionality *and* in the sense of postulating no sentential connectives other than the truth-functions. *Thus it is seen that at least a certain non-trivial portion of the semantics of modality is available to an extensionalist logician.* (Kripke 1959a, 3; italics added)

It seems that Kripke has indeed defeated Quine's arguments (since he has provided the required purely extensional means for the semantics of the modalities) and seemingly Quine accepted this point which could be based on the fact that he has not published much about the modalities after Kripkean semantics has appeared in the early 1960s.[11]

In the 1970s, however, two publications appeared from Quine which show a different picture and attitude:

> The notion of possible world did indeed *contribute to the semantics of modal logic*, and it behooves us to recognize the nature of its contribu-

---

[10]   About Quine's extensionalism see Bar-Am (2012) and Kemp (2014).

[11]   The last longer discussion of the modalities could be found in his *Word and Object* from 1960 (§41). In his most famous later writings, *Pursuit of Truth* (see Quine 1990/1992, 73-74) and *From Stimulus to Science* (see Quine 1995, 99 and 90-95), he discussed the modalities only a few pages altogether.

tion: it led to Kripke's precocious and significant theory of models of modal logic. Models afford consistency proofs; also they have heuristic value; but they *do not constitute explication*. Models, however clear they be in themselves, may leave us still at a loss for the primary, *intended interpretation*. When modal logic has been paraphrased in terms of such notions as possible world or rigid designator, where the displaced fog settles is on the question when to identify objects between worlds, or when to treat a designator as rigid, or where to attribute meta-physical necessity. (Quine 1972, 492-493; italics added)

A rigid designator differs from others in that it picks out its object by essential traits. It designates the object in all possible worlds in which it exists. Talk of possible worlds is a *graphic way of waging the essentialist philosophy*, but it is only that; it is not an explication. Essence is needed to identify an object from one possible world to another. (Quine 1977, 8)

In these passages Quine claimed that Kripkean semantics is *indeed* a huge step in order to interpret QML meaningfully. He admitted that the model-theoretic semantics provides various technical insights and results (completeness, soundness, paradox free structures etc.). He also pointed out, however, that (RV2) builds upon an equivocation.

We can talk about 'semantics' in the sense of *mathematical theory of models*, where the modelling of the relevant semantical notions are based on mathematical structures and formal relations; and on the other hand we can talk about 'semantics' as the *philosophical theory of meaning*. As Burgess (1998/2008, 216, italics added) says, both approaches are important but serve different purposes: "A mathematical theory of models could *refute a technical claim* to the effect that the common systems are formally inconsistent, but without some further gloss it cannot say anything against a *philosophical claim* that the common systems are *intuitively unintelligible*."[12] My

---

[12] Cf. with Quine (1962/1966, 176-177): "Still, man is a sense-making animal, and as such he derives little comfort from quantifying into modal contexts that he does not think he understands." Note the case of the Barcan-formulas. In certain formal semantics we can validate both the Barcan-formula and its converse. However, not everyone would admit the both of them are legitimate from a philosophical and metaphysical point of view. We can build up model-theoretic semantics in such a way that we validate only one of them but we certainly have to motivate this step from *outside* of the model theory.

claim is not simply that philosophical semantics, considered as a theory of meaning, is something radically different from mathematical semantics, considered as a branch of model theory, and therefore the second type of semantics cannot resolve directly problems stemming from the first type of semantics, since this would be almost trivially true. I would like to show rather that (i) the received view seems to claim that Kripke's model-theoretic semantics has indeed defeated Quine's arguments and (ii) that this view is mistaken and that (iii) neither participant of the 'debate' held the mistaken view.[13]

Quine is quite consistent on this point since he shows that even if the model-theoretic semantics of modal logic is viable, we still have to account for those possible worlds and individuals which seem to surface in these settings. He claims, for example, that if we try to rely on possible worlds, we risk that our approach will be non-reductive or even circular:

> [...] let us come to grips with necessity as such. It is not easy. A leaf that latter day philosophers have taken from Leibniz's book explains necessity as truth in all possible worlds. Whatever clarity can be gained from explaining necessity in terms of possibility, however, can be gained more directly: a sentence is necessarily true if it is not possibly false. 'Necessarily' means 'not possibly not'. And we can equally well explain possibility in terms of necessity: 'possibly' means 'not necessarily not'. We understand both adverbs or neither. (Quine 1987, 139-140; italics added)

Even though when this passage from 1987 was published a reductive analysis of modality, put forward by David Lewis who has been emphasizing his reductive approach at least from his 1968 article about counterparts, was already in play, Quine's remark could be read as an important warning: using merely *possible worlds* is just defining necessity in terms of possibility and negation. If we are satisfied with circular non-reductive definitions we could use our adverbs without advocating such entities like possible worlds.[14]

---

[13]   As Burgess (1998/2008, 216) claims, all the earlier literature "involve[s] essentially the same error, confusing Quine's philosophical complaint with some formal claim".

[14]   Of course for Quine (1972, 493), as a committed Humean, non-reductive definitions of modality won't work since our world is wholly non-modal: "everything is what it is, ask not what it may or must be". This phrase is a response to Kripke (1971, 160),

Besides his misgivings about possible *worlds*, on the other hand, in his notorious "On What There Is" from 1948, Quine formulated his skeptical remarks about (individuating) possible *individuals*:

> Take, for instance, the possible fat man in that doorway, and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? [...] Or, finally, is the concept of identity simply inapplicable to unactualized possibles? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another? These elements are well high incorrigible? (Quine 1948, 23-24)

As we noted earlier, this problem could be fixed with the help of strong essentialism, but Quine couldn't accept it for the mentioned reasons.

So, in the context of Quine's writings we can formulate the problem as follows: even if we provide a suitable formal semantics for modal logic, we still have to give *philosophical* reasons for motivating our formal approaches; we have to provide an answer for what are possible worlds, what are possible individuals, why the thesis of rigid designation holds, how to differentiate between proper names and definite descriptions, what are the conditions and commitments of essentialism etc.[15]

Despite the fact that the received view claims that Kripke has answered Quine's animadversions, the whole point of the (RV) is just becoming more complex, since according to Ballarin (2004, 609) "surprisingly, Quine's best ally on these matters turns out to be the very philosopher who

---

who cites the words of Bishop Butler in his "Identity and Necessity": 'everything is what it is and not another thing.'

[15]   At this point one could note again the parallel with Neurath, since he also claimed that "[...] as long as semantics appears as pure calculus I have nothing to say, assumed that your calculus is consistent" (Neurath to Carnap, April 1, 1944, ASP RC 102-55-05). For Neurath, formal and technical achievements were one thing, but 'philosophical' and scientific considerations were another. I say 'philosophical' since Neurath thought that philosophy wasn't something higher than science and he always tried to get rid of philosophy. In this case Neurath was a naturalist like Quine (1995/2008, 467) who claimed that "naturalized philosophy is continuous with natural science".

engineered the possible worlds model theory: Saul Kripke." In one of his articles Kripke formulates this point as follows:[16]

> *Philosophers should not confuse their own particular philosophical doctrines with the basic results and procedures of mathematical logic.* (Kripke 1976, 408; italics in the original)

> Philosophers often become so overjoyed, however, when they have found formal criteria for the success of some project that its intuitive basis is often disregarded like a ladder which can easily be kicked away after it has been climbed. [...] it is as if it were thought that any technical criterion, however loosely defended, is superior to a mere (!) philosophical argument. (Kripke 1976, 411)

> *Philosophers should maintain a proper skepticism of attempts easily to settle linguistic or other empirical questions by quick a priori formal considerations.* (Kripke 1976, 412; italics in the original)

> *Philosophers should have a better sense of both the power and the limitations of formal and mathematical techniques.* (Kripke 1976, 413; italics in the original)

> Logical investigations can obviously be a useful tool for philosophy. They must, however, be informed by a sensitivity to the philosophical significance of the formalism and by a generous admixture of common sense, as well as a thorough understanding both of the basic concepts and of the technical details of the formal material used. It should not be supposed that the formalism can grind out philosophical results in a manner beyond the capacity of ordinary philosophical reasoning. There is no mathematical substitute for philosophy. (Kripke 1976, 416)

These passages show that Kripke's attitude and metaphilosophical stance towards these matters were just the same as Quine's. Kripke is emphasizing that though model-theoretic formulations could provide suitable formal criterions with important mathematical results we have to be cautious when we are using formal logic in our argumentations. Formal logic

---

[16] See further Zvolenszky (2007) who also considers these quotations and their importance in the Quine-Kripke debate.

is just the *first step* when we try to defeat certain points of view but surely not the last. "There is no mathematical substitute for philosophy."

Interestingly Kripke was fully aware of the status of his model-theoretic semantics. For example, in his earliest formal articles, where he first formulated possible-worlds semantics, Kripke mentions Quine only two times – but what is more important, he did it only in certain insignificant contexts.[17] Firstly Kripke is mentioning Quine's *Mathematical Logic* with respect to the question that we shall formulate quantificational theory in a way which allows only closed formulae but there is no mention of the modalities (Kripke 1963/1971, 69, and note 14).

The other reference can be found in "A Completeness Theorem in Modal Logic" (see Kripke 1959a, 9) where Kripke is citing Quine's "Three Grades of Modal Involvement". In the citation Kripke is *not* responding to Quine's animadversions but mentions one of Quine's formulae, namely $(\forall x)(\forall y)(x = y) \supset \Box(x = y))$.[18] Regardless of the overall importance of this formula, it shows that Kripke was aware of Quine's writings about modality in 1958 when he was preparing his article while still in high school. Later Kripke applied to Harvard University where Quine was teaching. It is interesting since that time Quine was already famous for his remarks against modal logic so Kripke should have known this.[19]

After these publications, however, we can find no trace in Kripke's logical articles from the time that he was considering his formal semantics as a response to Quine's philosophical expectations.[20] Consequently we can plausibly assume that Kripke's intention *was not to undermine the whole of Quine's position with respect to the modalities*. What is more interesting is

---

[17] Kripke is referring to Quine quite a few times in his first two lectures of *Naming and Necessity* but I will just sketch this issue at the end of the paper since they belong mainly to the second phase of their debate, namely to the philosophical considerations of the modalities. Cf. Zvolenszky (2007).

[18] Burgess (1998/2008, 212, n8) notes that virtually this formula was the only explicit contribution of Quine to the formal theory of modal logic.

[19] Cf. Marcus (2010), who reports a similar attitude against modal logic from Harvard.

[20] Kripke (1963/1971, 65-66) discussed the various philosophical theories (Frege, Russell, Strawson) of descriptions but he considers them as merely a starting point of the formal investigations. Similarly none of Kripke's early model-theoretic works (cf. Kripke 1959b; 1963a; 1963b) contains any reference to Quine or his animadversions.

that the logical community was not seeing Kripke's writings as a response to Quine that time either despite the suggestion of the received view.

Firstly none of the (relatively few) reviews of Kripke mentions the name or the animadversions of Quine – and especially none of them articulates that Kripke has provided an answer to Quine (see Bayart 1966b; Kaplan 1966; Gabbay 1969). Arnould Bayart's review of Kripke's "A Completeness Theorem in Modal Logic" was put in *The Journal of Symbolic Logic* on exactly the same page where Bayart's other review of E. J. Lemmon has been published. In the latter review Bayart (1966a) doomed modal logic to be too obscure and vague without the relevant semantical constructions and metalogical results. This is important because in his review of Kripke (on the same page) Bayart did not mention that Kripke would provide all of these results while the received view indeed claims this.

It is also curious that in the literature from the late 1950s until the late 1960s almost none of the relevant logical journals (like *The Journal of Symbolic Logic, Notre Dame Journal of Formal Logic, Bulletin of the American Mathematical Society, Studia Logica, Zeitschrift für mathematische Logik und Grundlagen der Mathematik*) contains any reference to Quine's animadversions or to Kripke's alleged answers. The logicians were occupied with the systems of J. C. C. McKinsey, J. Łukasiewicz and it is hard to find any article on modal logic which does not refer to B. Sobociński. This suggests that the relevant scholars have discovered the works of Kripke relatively late. The usually discussed logician, Sören Halldén (1969, 306), for example, in his 1969 review of the debate between Lemmon and Henderson about "Is there only one correct system of modal logic?" argues that Lemmon's attempts to show that while the M, S4 and S5 systems require different strategies and all of them are legitimate is just "[...] impressionistic, but suggestive." The problem is that while the debate between Lemmon and Henderson took place in 1969 we usually think that Kripke's article which settled the questions about the legitimacy of the various modal systems was published already in 1963.

Of course there were exceptions in the literature. For example, M. J. Cresswell (1967; 1968) published several (quite positive) articles explicitly about Kripke in the *Notre Dame Journal of Formal Logic* in the 1960s. Cresswell's role is important because his book on modal logic with G. E. Hughes from 1968 was one of the first textbooks about modal logic and it used the Kripke-semantics substantively. The other exception is the afore-

mentioned E. J. Lemmon (1966a; 1966b) who tried to synthetize the algebraic and semantical approaches to modal logic that discussed and analyzed Kripke's works in a positive and subtle manner.

All of these suggest that the philosophical and logical atmosphere of the 1950s and 1960s wasn't after all so receptive with respect to Quine animadversions and Kripke's alleged answers. Two more points, however, need to be mentioned. First, if the logical community was inactive in answering Quine's philosophical animadversions (and for the right reason) who is responsible for the confusion of the formal and philosophical points? Burgess (1998/2008, 216) suggests that the confusion "is represented in the compendium by the suggestion that disputes about quantified modal logic should be conducted with reference to a 'semantic construction' in which connection the now superseded approach of Carnap is expounded (with the now standard, then unpublished, approach of Kripke being alluded to as an alternative in the discussion)." Burgess is referring here to the 1962 Boston colloquium where Quine and Marcus (along with Kripke and Føllesdal) discussed the questions of modal logic. Though he is not saying explicitly but he is quoting from Marcus' presentation and in order to understand the point let's see what was at stake.

Marcus (1961, 316) said that "I would like in conclusion to suggest that the polemics of modal logic are perhaps best carried out in terms of some explicit semantical construction. As we have seen [...] it is awkward at best and at worst has the character of a quibble, not to do so." The mentioned semantical construction corresponds to the formal considerations of the debate since Marcus (1961, 322, n19.) refers to Carnap's *Meaning and Necessity* which was viewed then as a formal statement of modal logic.[21] But she is mentioning Kripke's early abstract about the semantics of modal logic (see Kripke 1959b) too which contains only the formal aspects of the question.[22]

---

[21]   Whether Carnap's *Meaning and Necessity* is rightly considered as a purely formal work in the context of the formal-philosophical distinction of Quine is a further question in the light of the recent works of Carnap-scholars who highlight the philosophical character of Carnap's ideal of explication which is used in his book.

[22]   Interestingly while the English editions of the article (cf. Marcus 1961, 322, n20; 1962/1993, 23, n20) contain the reference to Kripke, one cannot find it in the Hungarian translation (cf. Marcus 1962/1985).

The confusion surfaces also in the closure of Marcus' article (1961, 321): "In such a model modal operators have to do with truth relative to the model, not with things. On this interpretation, Professor Quine's 'flight from intension' may have been exhilarating, but unnecessary." Though the 1993 edition of the paper does not contain the "not with things" phrase it was in the first edition and shows well the intention of Marcus, just like her footnote (cf. Marcus 1961, 322, n21): "If one wishes to talk about possible things then of course such a construction is inadequate." This remark is about *things* and not models, so it seems to suggest that Marcus is concerning with the philosophical and not the formal issues. But the context (and actually the whole paper) makes it clear that for Marcus "possible things" are related to the varying domains conception of quantification (in her semantics the domains of 'possible' individuals and quantification coincide). Thus in her paper (mentioned by Burgess) Marcus did not response to the formal/philosophical distinction of Quine and while she talks about both types of problems at the same she considers only the possible formal solutions under the aegis of semantics and thing.

The second point is that, as I suggested along with Zvolenszky (2007) and Burgess (2013, chapters 3-4), Kripke was fully aware of what he *is* doing and what he *has* to do in order to meet the interpretational challenges of Quine. After showing the formal correctness of modal logic he moved on and tried to give philosophical motivations and justifications for his main moves. This is to be found in his famous lectures of *Naming and Necessity*.

Regarding the above mentioned argument of Quine against modal logic based on the essentialist charge, the usual method was to attack Quine's first premise, pointing out that QML is not committed to essentialism because the problematic essentialist theses cannot be deduced in the systems of QML. As we saw, Quine has never claimed that it is the case. Kripke took part in those discussions from 1962 where Quine made his points explicit against Marcus. While Marcus didn't seem to grasp what the core of Quine's animadversions against essentialism were, Kripke responded more sympathetically to his ideas. In *Naming and Necessity* he didn't claim that one cannot deduce the essentialist theses in QML because it was just irrelevant for Quine's challenges. What he did is to show that Quine's second premise, that essentialism is an untenable, incoherent view, does not hold.

To show this, Kripke is using a fictive dialogue:

Suppose that someone said, pointing to Nixon, 'That's the guy who might have lost'. Someone else says 'Oh no, if you describe him as "Nixon", then he might have lost; but of course, describing him as the winner, then it is not true that he might have lost'. Now which one is being the philosopher, here, the unintuitive man? It seems to me obviously to be the second. The second man has a philosophical theory. The first man would say, and with great conviction, 'Well, of course, the winner of the election *might have been someone else*. The actual winner, had the course of the campaign been different, might have been the loser, and someone else the winner; or there might have been no election at all. So, such terms as "the winner" and "the loser" don't designate the same objects in all possible worlds. On the other hand, the term "Nixon" is just a *name* of *this man*'. When you ask whether it is necessary of contingent that *Nixon* who won the election, you are asking the intuitive question whether in some counterfactual situation, *this man* would in fact have lost the election. If someone thinks that the notion of a necessary or contingent property (forget whether there *are* any nontrivial necessary properties [and consider] just the *meaningfulness* of the notion) is a philosopher's notion with no intuitive content, he is wrong. (Kripke 1980/1990, 41-42)

Kripke is pointing out that the distinction between accidental and essential properties is not an *ad hoc* technical distinction of philosophy but we recognize and use it already in our daily discussions and language games. We understand those claims intuitively which consider the modal features of objects. He also mentions that the question is not whether there are *any* (nontrivial) essential properties – what he is interested in is whether it is meaningful to talk about such things: and that is what Quine was doing in his weak essentialist charges as we have seen above.

Naming and Necessity, however, contains a lot more material where Kripke explicitly aims to meet the philosophical challenges of Quine: he tries to motive his distinction between proper names and descriptions with a lot of examples and thought experiments based on everyday linguistic practices; he tries to rehabilitate some of the modal notions and show their intuitive content but we cannot consider here all of them.

4. Summary

The textbook-like history of analytic philosophy is a history of myths, received views and dogmas. Though mainly the last few years have witnessed a huge amount of historical works that aimed to reconsider our narratives of the history of analytic philosophy there is still a lot to do.

The present study is meant to present such a micro story which is still in many ways untouched by historians. In section 1 I characterized the received view about Quine and modal logic with six dogmas and considered here the first two of them. In section 2 I tried to show that while the received view claimed QML does not commit one to essentialism, since the essentialist theses cannot be derived in the usual systems, Quine's animadversions were based on different grounds. He was talking about the intended meaning of modal notions and the philosophical interpretation of modal language. To shed some light on this point in section 3 I placed the early writings of Kripke in the context of Quine's requirements. While the received view states that Quine's arguments against modal logic have been met by the famous possible-worlds semantics of Kripke we saw that his formal semantics was never meant to defeat Quine's more philosophical argumentations and Quine has accepted the results of Kripke's formal-mathematical investigations. What were at stake were no less than the philosophical ideas and motivations of modality and essentialism and these are just the topics of *Naming and Necessity*.

If we could accept this historical interpretation of the network between Quine, Kripke and modal logic, we have to conclude that Quine's *real* philosophical animadversions against the modalities are still on the table: though Kripke has provided some important considerations which were widely accepted among philosophers in the second half of twentieth century, Quine had his own positive philosophical ideas about modality based on his Humean basic stance which didn't get any attention in the literature yet.

Also there is the account of David Lewis in his *On the Plurality of Worlds* which is just as extensional and reductive as Quine has always required it. Lewis and Quine shared almost the same metaphilosophical commitments and attitudes (see Soames 2015) and though Lewis did not mention the name of Quine, his efforts in *On the Plurality of Worlds* could be read as an answer to the Quinean challenges:

When I say that possible worlds help with the analysis of modality, I do not mean that they help with the metalogical 'semantical analysis of modal logic'. Recent interest in possible worlds began there, to be sure. But wrongly. For that job, we need no possible worlds. We need sets of entities which, for heuristic guidance, 'may be regarded as' possible worlds, but which in truth may be anything you please. We are doing mathematics, not metaphysics. Where we need possible worlds, rather, is in applying the results of these metalogical investigations. Metalogical results, in themselves, answer no questions about the logic of modality. (Lewis 1986, 19)

Lewis claims just the same as Quine did: formal-logical investigations answer the technical problems while for his *interpretational* challenges one requires more substantial 'philosophical-scientific' investigations. As Lewis (1986, 20) said, "It is the substantive theory, not the metalogic, for which we need possible worlds."[23]

But the metaphysics of possible worlds is just one possible answer to Quine and in these philosophical questions there aren't any final answers. In a different context Kripke (1973/2013, 155) said that we have a certain tendency at similar point to throw up our hands. Whether we throw them up or not, the debate about the modalities is still on the table and the ideas of Quine should deserve a fresh start.

References

BALLARIN, R. (2004): The Interpretation of Necessity and the Necessity of Interpretation. *Journal of Philosophy* 101, No. 12, 609-638.

BAR-AM, N. (2012): Extensionalism in Context. *Philosophy of the Social Sciences* 42, No. 4, 543-560.

BAYART, A. (1966a): Review of Lemmon's 'Quantifiers and Modal Operators'. *The Journal of Symbolic Logic* 31, No. 2, 275-276.

BAYART, A. (1966b): Review of Kripke's 'A Completeness Theorem in Modal Logic.' *The Journal of Symbolic Logic* 31, No. 2, 276-277.

BURGESS, J. (1998/2008): Quinus ab Omni Naevo Vindicatus. In: *Mathematics, Models and Modality – Selected Philosophical Essays*. Cambridge: Cambridge University Press, 203-235.

BURGESS, J. (2013): *Kripke*. Cambridge: Polity Press.

---

[23]   Similar ideas are to be found in Lewis' *Convention*, see Lewis (1969/2002, 207-208).

CREATH, R. (ed.) (1990): *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*. Berkeley: University of California Press.

CRESSWELL, M. J. – HUGHES, G. E. (1968): *An Introduction to Modal Logic*. Methuen.

CRESSWELL, M. J. (1967): Alternative Completeness Theorems for Modal Systems. *Notre Dame Journal of Formal Logic* 8, No. 4, 339-345.

CRESSWELL, M. J. (1968): Completeness without the Barcan Formula. *Notre Dame Journal of Formal Logic* 9, No. 1, 75-80.

FINE, K. (1989): The Problem of *De Re* Modality. In: Almog, J. – Perry, J. – Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 197-272.

FINE, K. (1990): Quine on Quantifying in. In: Anderson, C. A. – Owens, J. (eds.): *Proceedings of the Conference on Propositional Attitudes*. Stanford: CSLI, 1-26.

FØLLESDAL, D. (1986/1998): Essentialism and Reference. In: Hahn, L. E. – Schilpp, P. A. (eds.): *The Philosophy of W. V. Quine*. Expanded Ed. Chicago – LaSalle: Open Court, 97-113.

GABBAY, D. (1969): Review of Kripke's 'Semantical Considerations on Modal Logics'. *The Journal of Symbolic Logic* 34, No. 3, 501.

GREGORY, P. A. (2008): *Quine's Naturalism: Language, Theory, and the Knowing Subject*. Continuum.

HALLDÉN, S. (1969): Review of Lemmon – Henderson's 'Symposium: Is There Only One Correct System of Modal Logic?' *The Journal of Symbolic Logic* 34, No. 2, 306.

HINTIKKA, J. (1982/1989): Is Alethic Modal Logic Possible? In: Hintikka, J. – Hintikka, M. B. (eds.): *The Logic of Epistemology and the Epistemology of Logic – Selected Essays*. Kluwer Academic Publisher – Springer, 1-15.

KAPLAN, D. (1966): Review of Kripke's 'Semantical analysis of modal logic: I. Normal Modal Propositional Calculi'. *The Journal of Symbolic Logic* 31, No. 1, 120-122.

KAPLAN, D. (1969): Quantifying In. In: Davidson, D. – Hintikka, J. (eds.): *Words and Objections. Essays on the Work of W. V. Quine*. Dordrecht – Boston: Reidel, 206-242.

KAPLAN, D. (1986): Opacity. In: Hahn, L. E. – Schilpp, P. A. (eds.): *The Philosophy of W. V. Quine*. La Salle: Open Court, 229-289.

KEMP, G. (2014): Quine's Relationship with Analytic Philosophy. In: Harman, G. – Lepore, E. (eds.): *A Compnaion to W. V. O. Quine*. Wiley-Blackwell, 69-88.

KRIPKE, S. (1959a): A Completeness Theorem in Modal Logic. *The Journal of Symbolic Logic* 24, No. 1, 1-14.

KRIPKE, S. (1959b): Semantical Analysis of Modal Logic (Abstract). *The Journal of Symbolic Logic* 24, 323-324.

KRIPKE, S. (1963/1971): Semantical Considerations on Modal Logic. In: Linsky, L. (ed.): *Reference and Modality*. Oxford: Oxford University Press, 63-72.

KRIPKE, S. (1963a): Semantical Analysis of Modal Logic: I. Normal Modal Propositional Calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 67-96.

KRIPKE, S. (1963b): Semantical Analysis of Intuitionistic Logic I. In: Crossley, J. N. – Dummett, M. A. E. (eds.): *Formal Systems and Recursive Functions*. Amsterdam: North-Holland, 92-129.

KRIPKE, S. (1971): Identity and Necessity. In: Munitz, M. K. (ed.): *Identity and Individuation*. New York University Press, 135-164.

KRIPKE, S. (1973/2013): *Reference and Existence*. Oxford University Press.

KRIPKE, S. (1976): Is There a Problem about Substitutional Quantification? In: Evans, G. – McDowell, J. (eds.): *Truth and Meaning*. Oxford: Oxford University Press, 325-419.

KRIPKE, S. (1980/1990): *Naming and Necessity*. Basil Blackwell.

LEMMON, E. J. (1966a): Algebraic Semantics for Modal Logics I. *The Journal of Symbolic Logic* 31, No. 1, 46-65.

LEMMON, E. J. (1966b): Algebraic Semantics for Modal Logics II. *The Journal of Symbolic Logic* 31, No. 2, 191-218.

LEWIS, D. (1986): *On the Plurality of Worlds*. Oxford: Basil Blackwell.

LEWIS, D. (1969/2002): *Convention*. Oxford: Blackwell.

LINSKY, L. (1969): Reference, Essentialism, and Modality. *The Journal of Philosophy* 66, No. 20, 687-700.

MARCUS, R. B. (1961): Modalities and Intensional Languages. *Synthese* 13, No. 4, 303-322.

MARCUS, R. B. (1962/1985): Modalitások és intenzionális nyelvek [Modalities and intensional languages. In Hungarian]. In: Copi, I. M. – Gould, J. A. (eds.): *Kortárstanulmányok a logikaelmélet kérdéseiről* [*Contemporary Readings in Logical Theory*. In Hungarian]. Budapest: Gondolat, 485-512.

MARCUS, R. B. (1962/1993): Modalities and Intensional Languages. In: *Modalities – Philosophical Essays*. Oxford University Press, 3-23.

MARCUS, R. B. (1967/1993): Essentialism in Modal Logic. In: *Modalities – Philosophical Essays*. Oxford University Press, 45-51.

MARCUS, R. B. (2010): A Philosopher's Calling. Dewey Lecture 2010. Available at: http://leiterreports.typepad.com/blog/2010/09/ruth-barcan-marcuss-philosophical-autobiography.html

MARCUS, R. B. ET AL. (1962/1993): Discussion. (Føllesdal, Kripke, Marcus, McCarthy, Quine) In: *Modalities – Philosophical Essays*. Oxford University Press, 24-38.

MELIA, J. (2003): *Modality*. Chesham: Acumen.

NEALE, S. (2000): On a Milestone of Empiricism. In: Orenstein, A. – Koťatko, P. (eds.): *Knowledge, Language and Logic: Questions for Quine*. Dordrecht: Springer, 237-346.

PARSONS, T. (1967): Grades of Essentialism in Quantified Modal Logic. *Noûs* 1, No. 2, 181-191.

PARSONS, T. (1969): Essentialism and Quantified Modal Logic. *The Philosophical Review* 78, No. 1, 35-52.

QUINE, W. V. O. (1932/1990): *The Logic of Sequences: A Generalization of 'Principia Mathematica'*. Harvard Dissertations in Philosophy. New York: Garland.

QUINE, W. V. O. (1941): Whitehead and the Rise of Modern Logic. In: Schilpp, P. A. (ed.): *The Philosophy of Alfred North Whitehead*. Evanston: Northwestern University Press, 125-163.

QUINE, W. V. O. (1943): Notes on Existence and Necessity. *The Journal of Philosophy* 40, No. 5, 113-127.

QUINE, W. V. O. (1947): The Problem of Interpreting Modal Logic. *The Journal of Symbolic Logic* 12, No. 2, 43-48.

QUINE, W. V. O. (1948): On What There Is. *Review of Metaphysics* 2, No. 5, 21-36.

QUINE, W. V. O. (1951): Two Dogmas of Empiricism. *The Philosophical Review* 60, 20-43.

QUINE, W. V. O. (1953): Reference and Modality. In: *From a Logical Point of View*. 1st ed. Cambridge (Mass.): Harvard University Press, 139-159.

QUINE, W. V. O. (1953/1961): Reference and Modality. In: *From a Logical Point of View*. 2nd ed. Cambridge (Mass.): Harvard University Press, 139-159.

QUINE, W. V. O. (1953/1966): Three Grades of Modal Involvement. In: *The Ways of Paradox and Other Essays*. New York: Random House, 156-174.

QUINE, W. V. O. (1960): *Word and Object*. The MIT Press.

QUINE, W. V. O. (1962/1966): Reply to Professor Marcus. In: *The Ways of Paradox and Other Essays*. New York: Random House, 175-182.

QUINE, W. V. O. (1963/1966): Necessary Truth. In: *The Ways of Paradox and Other Essays*. New York: Random House, 48-56.

QUINE, W. V. O. (1972): Review of Munitz, M. K.'s 'Identity and Individuation'. *The Journal of Philosophy* 69, No. 16, 488-497.

QUINE, W. V. O. (1977): Intensions Revisited. *Midwest Studies in Philosophy* 2, No. 1, 5-11.

QUINE, W. V. O. (1987): Necessity. In: *Quiddities: An Intermittently Philosophical Dictionary*. Cambridge – London: Belknap Press of Harvard University Press, 139-142.

QUINE, W. V. O. (1990/1992): *Pursuit of Truth: Revised Edition*. Harvard University Press.

QUINE, W. V. O. (1990/2004): Two Dogmas in Retrospect. In: *Quintessence – Basic Readings from the Philosophy of W. V. Quine*. The Belknap Press of Harvard University Press, 54-63.

QUINE, W. V. O. (1992/2008): Pressing Extensionality. In: Quine (2008), 172-175.

QUINE, W. V. O. (1995): *From Stimulus to Science*. Cambridge (Mass.): Harvard University Press.

QUINE, W. V. O. (1995/2008): Naturalism; Or, Living with One's Means. In: Quine (2008), 461-472.

QUINE, W. V. O. (2001/2008): Confessions of a Confirmed Extensionalist. In: Quine (2008), 498-506.

QUINE, W. V. O. (2008): *Confessions of a Confirmed Extensionalist and Other Essays*. Føllesdal, D. – Quine, D. B. (eds.), Harvard University Press.

RECK, E. (2013): Introduction: Analytic Philosophy and Philosophical History. In: Reck, E. (ed.): *The Historical Turn in Analytic Philosophy*. Palgrave MacMillan, 1-36.

SOAMES, S. (2015): David Lewis's Place in Analytic Philosophy. In: Loewer, B. – Schaffer, J. (eds.): *The Blackwell Companion to David Lewis*. Wiley-Blackwell, 80-98.

WEIR, A. (2014): Quine's Naturalism. In: Harman, G. – Lepore, E. (eds.): *A Companion to W. V. O. Quine*. Wiley-Blackwell, 114-147.

ZVOLENSZKY, Z. (2006): Analytic Truths and Kripke's Semantic Turn. *Croation Journal of Philosophy* 6, No. 2, 327-341.

ZVOLENSZKY, Z. (2007): Megnevezés és szükségszerűség – Négy évtized távlatában [Four decades of *Naming and Necessity*. In Hungarian.] In: Zvolenszky, Z (ed.): *Saul Kripke: Megnevezés és szükségszerűség* [*Naming and Necessity*. In Hungarian]. Budapest: Akadémiai, 151-218.

ZVOLENSZKY, Z. (2010): Ruzsa on Quine's Argument against Modal Logic. *Hungarian Philosophical Review* 54, No. 4, 40-48.

Sense, Reference, Speech Acts, Norms and Other Issues. Replies to *Organon F* Papers (Part I)

PETR KOŤÁTKO

Department of Analytic Philosophy. Institute of Philosophy. Czech Academy of Sciences
Jilská 1. 110 00 Prague 1. Czech Republic
kotatko@flu.cas.cz

If I found the first 2015 Supplementary Issue of *Organon F* in the library shelf, I would be struck by its strong assembly of authors: brilliant philosophers and at the same time people whom I have had the pleasure and honour of personally knowing for quite a long time. But I have been presented the volume on an occasion which made it clear that it is somehow (rather intimately) connected with me – despite my repeated pleading not to be reminded about my (regrettable) anniversary. Now, after some time, with all the solemnity belonging to my age, I would like to say how grateful I am for this generous gift to all my friends involved: the authors, the editor (and author of the warm introductory words), the editor-in-chief of *Organon F* and his colleagues. To the latter I would also like to thank for their kind readiness to publish my (tedious but still too brief) reactions on all the papers. I have learned a lot from them, not to speak about the pleasure of an intellectual adventure they involved me into. And although my replies don't – by far – exploit all the impulses they offered, I hope that they at least indicate how inspiring this reading has been for me.[1]

---

[1]   In what follows I preserve the order in which the papers appeared. The replies will come out in two parts: the present one focuses on the philosophy of language, the next one (to appear in *Organon F*, 1/2016) on the philosophy of literature.

Peter Pagin[2]

Peter critically examines an argument I have presented for the first time in Koťátko (1998). In that paper, like in some later texts, I have focused on Donald Davidson's concept of utterance meaning.[3] I took (and still take) it as powerful in two respects: it is applicable to essential part of everyday communication, and it is extremely economical by giving us three valuable things by the same stroke: the general account of utterance meaning (as being constituted by the match between communicative intention and interpretation), the notion of communicative success (as consisting in reaching the match) and the principle of identifying the meaning of particular utterances (an utterance means that p if and only if it was so meant by the speaker and interpreted by the audience).

Inspired by these attractions, I have tried to analyze the requirements this triple notion imposes on the correlations between the speaker's and the interpreter's attitudes (those which can be relevant for the choice of communicative strategies)[4] and found them very modest. I agree with Peter (p. 8) that this should not worry Davidsonians at all: quite on the contrary. But I have argued that the Davidsonian account cannot be generalized for all kinds of communication and that the Trinity I have celebrated above should be dissolved, however blasphemous such a suggestion may appear: what we need is a general account of utterance meaning which allows for various notions of communicative success and various principles of the determination of utterance meanings, both depending on the type of discourse in question or even on specific features of the communicative situations.

As a prelude to this suggestion I have offered an example: Paul says "Martial wrote witty epigrams" intending (in agreement with what he takes

---

[2]   Reply to Pagin (2015). All the page references which appear without the author's name refer to this paper.

[3]   In one of its most concise expositions, Davidson speaks about the central importance of "situations, in which someone intends (or assumes or expects) that his words will be understood in a certain way, and they are. In such cases we say without hesitation: how he intended to be understood, and was understood, is what he, and his words, literally meant on that occasion" (Davidson 1994, 11-12).

[4]   I have a bit more developed and in some respects modified the analysis, as well as conclusions drawn from it, in Koťátko (2000).

to be the standard meaning of "epigram") to assert that Martial wrote witty epigrams. At the same time he hopes to be interpreted by John (in agreement with what John, according to Paul's belief, takes to be the standard meaning of "epigram") as asserting that Martial wrote witty epitaphs. Paul's aim is to embarass John, granted that his misinterpretation comes to light. Peter (not another character in this fiction but my actual opponent Peter Pagin) classifies this desired effect as something Paul is "primarily interested in" (p. 13), which is apt with respect to the position of this aim in the chain of intentions ascribed to Paul in the story. But we should keep in mind that this is a *non-communicative* attitude motivating Paul's communicative intention to express certain proposition with assertive force. The non-communicative attitudes may help us in identifying and understanding the communicative intentions but should not themselves be included into the analysis of utterance meaning. (I mention this just because in the main part of Peter's argument the term "primary" plays an important role in the analysis of *communicative* intentions.)

Peter analyzes the Martial case[5] in a way which should show that it cannot serve as a counter-example to Davidson's account of utterance meaning: more precisely, that it is not a case where the following condition (extracted from Davidson's concept of utterance meaning) is violated:

(a)   If *S* makes an utterance in order to perform certain speech act, she intends and expects that act to be assigned to the utterance in A's interpretation.[6]

Let me try to summarize Peter's interpretation of the Martial example (needless to stress, anybody interested in this issue would do better to read his full analysis in its original version). On the *primary* level, Paul intends to express the standard meaning of "epigram" and John interprets "epigram" (as uttered by Paul) as having its standard meaning. So, on this level, the communication is successful in Davidson's sense. Moreover, "Paul pri-

---

[5]   He suggests two alternative construals of the situation I describe in the Martial example: for short, I will focus on the first one, which, according to Peter, corresponds to the scenario I rely on in my polemics with Davidson.

[6]   Peter's objection against my use of the term "intend" here will be discussed later. A correlative condition (b) concerning the interpreter's attitude (which also plays a role in my original argument) will be omitted here for the sake of brevity.

marily intends *epigram* to mean just what he expects John to primarily interpret it as meaning, in accordance with the condition (a)" (p. 11). According to Peter, a clash appears just on the *secondary* level, due to the difference between Paul's and John's beliefs about the standard meaning of "epigram" and to Paul's relying on this difference in his strategy. Paul secondarily intends "epigram" to mean epigram, while John secondarily interprets (and is expected by John to interpret) "epigram" as epitaph. And Peter argues that the primary level is what should be taken as relevant from the Davidsonian point of view: hence Davidson's position is not threatened by Martial example.

I have to admit that I have a problem with the „primary-secondary" distinction in Peter's analysis, in particular with the conclusion he draws from his specification of Paul's primary intention and John's primary interpretation. It would make good sense to me if I could read Peter's specification of Paul's primary intention and of John's primary interpretation (cf. (i) in (Paul) and (i) in (John) on p. 10), so that both Paul and John *defer*[7] to the standard meaning of "epigram". Then it would be natural to say that it should be this primary intention and interpretation what matters in our account of the situation and of the meaning of Paul's utterance. But since John and Paul have certain (mutually incompatible) beliefs about what the standard meaning of "epigram" is (as it is put in (ii) both in (Paul) and (John) on p. 10 in Peter's paper), I find it appropriate to say that instead of deferring to the standard meaning, Paul *straightforwardly* means epigram by "epigram", taking himself to conform thereby to the standard, and John *straightforwardly* interprets "epigram" as epitaph, taking this to be in conformity with the standard. The result is a *straightforward* discrepancy between communicative intention and interpretation: this is what I take to be a more plausible description of the communicative situation in question,

---

[7]   In his note 2, Peter relates the term "deference" to his alternative analysis of Martial example, which I am not discussing here. Our ways of using this term do not seem to coincide. In my use, the speaker defers to the standard meaning of a term if she does not know what that meaning is, is aware of her ignorance, but intends her utterance to have that meaning, whatever it may be. This is certainly not the case in my exposition of the Martial case, nor in Peter's reconstruction of it that I have been just discussing. But I would not take his second construal of the Martial case as an example of deference either: there I would say that Paul and John are submissive with respect to the standard, rather than that they are deferring to it.

than Peter's conclusion that there is an agreement between the speaker and the interpreter. Peter evaluates the situation as a communicative success, I take the same situation as a clear example of misunderstanding. Because what matters from the communicative point of view is, according to my view, the difference in the literary genre Paul and John take the utterance to mention (since this is likely to make the following conversation confused) rather than the fact that each of them takes himself as conforming to the standard. Moreover, Paul expects and intends John to *straightforwardly* interpret "epigram" as epitaph, *in accordance* with John's presupposed belief about the standard meaning, rather than to *defer* to the standard meaning.

In case of genuine deference the situation and its implications for our discussion would be indeed different. If we (as I think we should) include the deference, whenever we assign it to the speaker, into the construction of the content the speaker intends to express,[8] we would get for Paul's case something like: "Martial wrote witty pieces of poetry belonging to the genre referred to by experts as 'epigram'". And we would get the same construal of the content John assigns to Paul's utterance in his interpretation. Moreover, granted that they both belong to the same linguistic community, the description "the genre referred to by experts as 'epigram'" would pick out the same literary genre in Paul's and John's case: so we would have a two-level match between intention and interpretation. This would be a clear example of communicative success. And granted that Paul expects and intends John to defer to the standard meaning of "epigram" in his interpretation of Paul's utterance, the condition (a) is not violated.

But, as I have suggested, the situation is not like that. Neither John nor Paul are actually *deferring* to the standard: rather, Paul means his utterance and John interprets it *in accordance with what they take to be* the standard. And Paul wishes John to interpret the utterance in a non-standard way, *in accordance* with his (presupposed) false belief about the standard, rather than to *defer* to the standard.

This is how I have construed the situation in the Martial example. Obviously, if Paul, under the conditions just specified, intends to make a mea-

---

[8]   For details see Koťátko (2012). I have had a discussion about some implications of this account with Peter earlier, profited from it a lot and tried to reply to (what I took to be) the main objection in Koťátko (2006).

ningful utterance, he must count with some principle constituting the utterance meaning independently of the audience's interpretation: then the utterance meaning he relies on also cannot depend on the match between the communicative intention and interpretation. My claim was that this kind of communicative strategy makes sense and, moreover, it can be successful. To admit this means to assume (as Paul does in our example) that the speaker can intend to assert that p without intending to be interpreted as asserting that p. It is quite likely that the people who witness the conversation would be disposed to describe it so that Paul asserted that Martial wrote witty epigrams and has been misunderstood by John (provided that John's way of interpreting the utterance becomes manifest in his reaction).

In general, I believe that our account of utterance meaning should be flexible enough to provide a space for various kinds of communicative situations differing in the role they assign to factors like intention, interpretation, linguistic conventions (or other kinds of social standards) in the determination of utterance meanings. These differences will depend either on the attitudes of the communicants (e.g. on the degree of their submissiveness or indifference to linguistic conventions) or on the social setting within which the conversation takes place. Examples taken from highly institutionalized or ritualized types of communication, like issuing legal enactments or military orders (cf. the army-example in Koťátko 1998, 235), pronouncing wedding formulas, nomination formulas etc. show perhaps more convincingly than the Martial case that the match-account of utterance meaning cannot be generalized any more than its conventionalist rival.

In this discussion, as well as in the original paper, I have been speaking about Paul as intending to be interpreted in certain way: now I should say a few words concerning Peter's criticism of this use of the verb "to intend" (p. 6). Having in mind my underdeveloped linguistic intuition in English, I do not have anything to object against Peter's characteristics of the standard (ordinary) use of this word. On the other side, I think I have not deviated from the way it is used in relevant parts of the philosophical literature, including Gricean semantics (here the speaker is taken to mean something by her utterance if she *intends* to produce certain effect in the audience on the basis of her recognition of this intention and *intends* to...), Searle's speech act theory (cf., e.g., the condition 8 of his famous definition

of promise) and also in Davidson's way of speaking about utterance meaning. For instance, in the passage quoted above from (1994), Davidson speaks about the central importance of "situations, in which someone intends (or assumes or expects) that his words will be understood in a certain way, and they are". Surely, one cannot intend to produce some effect E by his act A unless one *believes* that doing A is, in given circumstances, an appropriate way for producing E:[9] but to require that one is *certain* about A's leading to E (as Peter does on p. 6), does not seem to be compatible with the examples just given.

In accordance with this, I suppose that the readers will understand me as writing these lines with the intention to meet Peter's objections, without suspecting that I do not realize how easily I can fail. Let me add that Peter's brilliant paper is much richer than my response could reveal and I can only recommend it for careful reading.

Alberto Voltolini[10]

Alberto has been generous enough to put aside our disagreements concerning the language of fiction and turned instead to the Carnap-Heidegger controversy. I can only admire his subtle and elaborate discussion of Carnap's analysis of Heidegger's famous sentence (in Carnap's version: "The nothing nothings") and agree with his conclusion that it is by far not as devastating as intended. Under the proper analysis, as Alberto shows, the sentence is logically well formed and (under certain non-trivial assumptions) even comes out as true. This is in fact all I can say in reaction to this brilliant paper – except rather general moral concerning rules of philosophical discussion, which I cannot avoid voicing in confrontation with this famous case.

I have to admit (as a confession rather than as a matter of programatic declaration) that I never went into such a systematic interpretative enterprise concerning any piece of Heidegger's texts. But as a Carnap's reader,

---

[9]   A well-known Schiffer's argument against one version of Grice's definition of speaker's meaning heavily relies on this platitude, cf. Schiffer (1972, 19-21).

[10]   Reply to Voltolini (2015). All the page references which appear without the author's name refer to this paper.

I could not have failed noticing that his analysis could hardly count as an attempt to understand Heidegger's text, which means to find its *most charitable interpretation*: the reading which maximizes its philosophical value (including, of course, its coherence and truthfulness) and is compatible with the text itself plus the available non-textual (but interpretatively relevant) evidence.

According to Quine's famous sentence, you, as an interpreter, should keep in mind that "your interlocutor's silliness is less likely than your bad interpretation" (Quine 1960, 59). Similarly, if a philosopher of language comes with a criterion of meaningfulness which is not met by a considerable part of our everyday conversation and even by sentences in which that very criterion has been introduced, she should take it as a good reason to look for another theory (rather than to turn to rhetorical tricks, e.g. some variants of the "leather metaphor"). And if we apply our analytical apparatus on a piece of theoretical text and the outcome is a plain nonsense, our first thought should be that there may be something wrong with the apparatus chosen or the way we have applied it – in particular if the text in question is a work of one of the best educated and most influential thinkers of our time.

I suppose that everybody who read for the first time Kant, Hegel and of course Heidegger, found herself involved in the adventure of *radical interpretation* and hence must have recognized that applying previous interpretational training and routines acquired in it would not do. I do not see why this should not apply to the skills in operating with the kind of technical apparatus we are accustomed to. Moreover, here we should keep in mind what Frege said (and what Fredrik Stjernberg commented in the paper we will shortly come to) about the danger that "the formulaic mechanism would take over to such an extent that it suffocated the thought completely" (Frege 1980, 4f). In our case the "complete suffocation of thought" would have the form of paralysis of our interpretative abilities. Interpretation of a philosophical text, in particular not belonging to the intellectual tradition we are acquainted with, may be a serious challenge for our intellect, intuition, imagination and sensitivity: the strenuous work they are supposed to do cannot be reduced to operations of our technical apparatus. Nothing like this has happened in Alberto's text, despite all its technical precision.

Fredrik Stjernberg[11]

Fredrik has done an admirable work in order to properly expose and clarify the tension between Frege's "intuitive" criterion of the identity of senses (relying on the speakers' authority concerning senses, cf. p. 44 f.) and his admitting that senses of some expressions are not known to anybody. Moreover, he has explored possible ways out and found the most promising one in a (specific kind of) pragmatic account of Fregean senses. Like in the preceding case, I find the author's arguments convincing and have nothing to add to them – except few general remarks concerning the possible functions of Fregean senses.

Confronted with the kind of inquiry to be found in Fredrik's paper and with its results, a philosopher of language can still ask whether she really has a useful application for Fregean senses in the specific field of her study. What is important from my point of view is that, despite Frege's claims to the opposite, postulating senses as ideal entities in Fregean strong reading (i.e. as inhabitants of the Third Realm) is neither necessary nor sufficient to explain the possibility of successful communication.[12] It is not sufficient, since we are still left with the task of explaining how it comes that the speaker and the audience understand the expressions uttered in the same way. And no matter whether we see the explanation in their sharing of linguistic conventions or in their creatively coordinating their communicative strategies (concerning the choice of expressions on the part of the speaker and their interpretation on the part of the audience), the ideality of senses does not make our task easier at all. And if we succeed, in both cases heavily exploiting psychological notions (like intention, belief and preference), we have all we need to explain what is the communicative success based on and it comes out that the ideal senses "do not oil the wheels" of the theory (to borrow a well-known metaphor from Davidson).

Another doubts about the usefulness of senses concern the presumption that senses (no matter how we interpret their ontological status), construed as complete modes of presentation of the entities referred to, is what we have to know to be able to successfully communicate. As Fredrik reminds

---

[11]   Reply to Stjernberg (2015). All the page references which appear without the author's name refer to this paper.

[12]   I have tried to say more about this in Koťátko (1995).

us (cf. especially p. 55), Frege admits that in case of expressions with extremely complex senses, it would be unrealistic to expect that whenever we use them we operate with their complete senses (as something we grasp in its complexity). Then the expression serves as a "receptacle" for sense and we know that we can open it if necessary, by going through the expression's definition: owing to such a definition, "we can cram the sense into the receptacle and also take it out again" (Frege 1979, 209). Now, it requires just one more step to realize that there are no such definitions available for the expressions we use in everyday communication and any choice of a definition would be a matter of an arbitrary decision. So there is no general criterion for deciding what belongs into the receptacle "whale" and what is just a piece of extra-semantic knowledge we associate with the expression on the basis of what we have learned at school, read in popular zoological magazines, etc.[13] The actual use of the expression, which can be the only source of its meaning, simply does not provide us with any such criterion. Then, to follow the well-known line, if we find a creature which in all respects resembles to whales, as we know them, except that it is not a mammal, there is no chance to decide whether calling it "whale" should count as a continuation of or a departure from the existing usage, and hence as a preservation of its meaning or its replacement by a new one. Similarly, we cannot decide whether it directly follows from "Ann is a whale" that "Ann is a mammal", or we need an extra empirical premise "All whales are mammals" to justify such an inference. Hence it would be perhaps more realistic to say that the expression serves as a receptacle which (in any particular moment), contains all pieces of information which count (in that moment) as knowledge about whales – but then the content of the container corresponds rather to Putnam's "stereotype" than to Fregean sense. In Putnam's version (cf. esp. Putnam 1975), the core element of the meaning of "whale" is neither the Fregean sense, nor the stereotype, but an indexical identification of exemplars of animals for which (together with all other individuals sharing the essence with them) the term "whale" has been introduced.

Fredrik shows that Frege's intuitive (epistemological) criterion for identification of senses can function only provided that we replace the real sub-

---

[13]   This is of course one of the famous indeterminacies analyzed by Quine, namely the inextricability of meaning (cf. e.g. Quine 1960, 38).

jects with their idealised counterparts. This raises questions concerning the practical usefulness of the criterion but should not strike us too much provided that Fregean senses, as Fredrik also emphasizes, are not supposed to be of human origin (for instance are not bound to such human creations like language) and that's why are not construed so as to match human cognitive capacities. The question then, of course, is why the author insisting on this status of senses repeatedly appeals precisely to this criterion (cf. Stjernberg 2015, Sec. 5). Here is an analogy which immediately comes to my mind: in a bit more recent (though already also classical) literature we find at least one even more striking example of a similar discrepancy. Grice's well-known attempts to define the speaker's meaning acquired (under the pressure of more and more complex counterexamples) the form of an infinite series of conditions, specifying intentions the speaker must have in order to mean something by her utterance. After some (unsuccessful) attempts to stop the regress or to make it harmless Grice (1989) had to admit that the terrestrial speakers (as finite beings) are unable to mean something by their utterances in the full sense and that the speaker's meaning is an ideal limit to which we can only approximate. This should be much more striking than Frege's problem: we end with a psychologically unrealistic construct, while the original promise of the Gricean project was to bring the semantic notions down to earth by showing their psychological basis. Schiffer (1987), after trying to lock the (apparently unavoidable) infinite regress into the notion of "mutual knowledge", suitably located in the definition of speaker's meaning, had to admit that this attempt failed together with all others, the whole project of intentional semantics is in ruins and his conclusion was that "meaning is theory resistant" (since its basic level, speaker's meaning, proved to be undefinable). Fredrik does not seem to be equally sceptical about the prospects of semantics based on the Fregean notion of sense, as the suggestion discussed in the last part of his paper shows.

Kathrin Glüer – Åsa Wikforss[14]

Let me say in advance that I share Kathrin's and Åsa's position in their dispute with normativists and I have repeatedly had an opportunity to refer

---

[14] Reply to Glüer – Wikforss (2015). All the page references which appear without the authors' name refer to this paper.

to their arguments or apply a modified version of them. In particular, I fully agree that there is nothing intrinsically normative in the notion of conditions of correct application of an expression (similarly like in case of conditions of correct use of a washing machine, tommy gun etc.). The fact that an expression E has certain conditions of correct application assigned to it by conventions of English acquires normative consequences only within the context created by a norm which commits us to conform to conventions of English (like: "everybody employed in this office is obliged to speak correct English"). The obligation clearly comes from the outside, *in addition to* the fact that E has certain conditions of correct application, rather than being intrinsic to that fact.[15] How it comes out that expressions have certain meanings and hence certain conditions of correct application (and what makes the links between expressions and meanings in some cases relatively stable) can be explained without using normative terms. In case of sociolects like English, it can be done in terms of regularities fixed by certain complex of attitudes of the kind specified in Lewis' famous definition of convention.[16] In case of an idiolect of some speaker S, the fact that certain expression has certain conditions of correct use is simply a matter of S's communicative habits. And in case of particular utterances, the fact that an expression E has been used with certain conditions of correct application (whether or not these conditions have been fulfilled), is a matter of S's communicative intentions. They may be based on S's idiolect or on her (true or false) beliefs about the standard meaning of E, or on her (true or false) beliefs about the interpreter's idiolect, or on her (true or false) beliefs about the interpreter's beliefs about the standard meaning of E, etc. Finally, if we accept Davidsonian account of utterance meaning (in general or for the type of discourse or communicative situation in question), we will say that an expression E, as it appears in given utterance, has certain conditions of correct application, if and only if they were assigned to E (on that occasion) both in the speaker's communicative intention and in the audience's

---

[15] Generally speaking, Lewisian convention is "a regularity sustained by a special kind of a system of beliefs and desires", defined in non-normative terms (Lewis 1983, 179). In agreement with this, Lewis warns us not to confuse conventions in his sense with "our famously obscured friends, the rules of language, renamed" (ibid.). No wonder, Lewis' theory of convention has become a target of some normativists.

[16] A scheme of a scenario showing how such regularities and complex of attitudes may have developed can be borrowed from Schiffer (1972, 189), or Grice (1989).

interpretation. It makes no difference whether this match has been reached in a standard or a highly non-standard way: there is no general reason to allow conventions, rules (which should not be confused with the former – cf. above note 15) or other kind of standards to re-evaluate or retouch the result of the speaker's and audience's cooperative enterprise, which led to understanding. This counts, of course, for those communicative situations, in which the interest in mutual understanding weights more than the standards or our loyalty to them. Presumably, most of our everyday conversation (unlike the use of language in highly formalized or ritualized areas of discourse) is of this kind.

On the level of a sociolect (where the philosophers' insistence on the normativity of meaning can be expected to be most obsessive), the fact that E has such and such (conventionally fixed) conditions of correct application gives, quite straightforwardly, rise to hypothetical imperatives like: "If you want to conform to conventions of English, use E in such and such way!" It surely makes good sense to react to the introductory conditional clause by saying: "I don't! (or: I don't care!). I simply want to be understood." And it certainly would not be right to reply: "But that's, practically speaking, the same." While in many cases conformity to linguistic standards is the best way to understanding (and to fulfilling our practical aims which require understanding), in other cases it would generate misunderstanding. Obviously, the cases in which the audience makes systematic mistakes in connecting certain words with meanings or ascribes such mistakes to the speaker (and interprets her utterances in accordance with this assumption) or believes that the speaker ascribes such a mistake to her (and expects the speaker to base her choice of words on this assumption) etc. are of this kind.

According to Kathrin and Åsa, we should approach the term "correct", as it appears within the term "conditions of correct application", or to be precise, within some principle of the form

(C) w means $F \rightarrow \forall x$ (w applies correctly to $x \leftrightarrow x$ is f)[17]

as a placeholder for a suitable basic semantic concept (p. 69). The choice depends on our favourite semantic theory: Kathrin and Åsa mention *truth*

---

[17]   Here "w" is a word, „F" gives its meaning and "f" is that feature in virtue of which „F" applies. Kathrin and Åsa adopt (C) from D. Whiting for the purposes of their polemics with his normativist position.

and *warranted assertibility* as "the main contenders" (p. 66). In the final part of their paper, they react on the fact that their normativist opponent, Daniel Whiting, shares their view that the basic semantic concept (they focus on *truth*) is not normative. Then, in order to show that "correct" in (C) must be interpreted normatively, "he would have to argue that the anti-normativist fan of truth-conditional semantics, for instance, is missing something essential to meaning by interpreting 'correct' in (C) as a place-holder for 'true'..." (p. 72).

But, as we know, not everybody is ready to the concession concerning the notion of truth which Kathrin and Åsa appreciate in Whiting's case. Instead of reproducing the views of their opponents (like Hans Joachim Glock) insisting on the normativity of truth, let me turn to one of the greatest authorities of the recent past and recall the way in which Michael Dummett construed the relation between the notion of truth and the notion of correctness. Dummett famously argued that we can justify the meaning-theoretical relevance of the notion of truth[18] only by specifying the "linking principle" between that notion and our practice of making assertions. And the core of solving this task consists, according to his view, in deriving the notion of truth from the "more basic" notion of correctness of assertions. Cf. e.g.:

> Any workable account of assertion must recognize that assertion is judged by objective standards of correctness and that, in making an assertion, a speaker lays claim, rightly or wrongly, to have satisfied those standards. It is from these primitive conceptions of the correctness or incorrectness of assertion that the notions of truth and falsity take their origin. (Dummett 1976, 83; cf. also 1973, 289; 1991, 165f. et al.)

The first question of course is whether we should follow this way of relating the basic concepts we are dealing with here (*truth*, *correctness* and the *theory of meaning*). If it is worth doing, the second question is whether the notion of correctness from which the notion of truth is derived here is *normative* or not. The answer seems to be "yes", since Dummett speaks about the objective "standards of correctness" to which all speakers are committed – other-

---

[18] Since Dummett takes semantics as having by definition *truth* as its central notion, the same point can be put as a problem of justification of the presumption that a meaning theory must have a semantic basis.

wise all assertions would not be judged by these standards (cf. the quotation above) and it would not automatically belong to making an assertion that the speaker "lays claim… to have satisfied those standards". That looks like a challenge for anti-normativists. Although in (C) the term "correct" occurs in quite a different context than in the sentence quoted above from Dummett, if it serves in (C) as a placeholder for the notion of truth and that notion is derived from the normative notion of correctness of assertion, the occurrence of "correct" in (C) in the end serves as a channel through which normativity comes in. So, when defending their interpretation of (C), the anti-normativists have a good reason (the one presented here, not to mention others) to go into confrontation with Dummett.

This is compatible with the platitude that the conditions of correct use of expressions and the correctness conditions of assertions are two different issues, belonging to two different areas: debates about expression meanings and debates about speech acts and utterance meanings. While I share the anti-normativist position in the former respect, I believe that speech act types should be defined in terms of their normative consequences (namely the commitments they impose on speakers). And the same holds for utterance meanings, if we identify them with speech acts performed in utterances. This is not motivated by any of the arguments which appeared in the dispute between normativists and anti-normativists: rather, it is an outcome of my critical confrontation with the Gricean-type intentionalists. As I understand Kathrin and Åsa (cf. p. 67), their anti-normativist interpretation of the principle (C) does not commit them to go into polemics with me in this respect.

To summarize my views on this topic in the most general (and rather declarative) way, I suggest that from the following three claims about normativity of meaning we accept the first one and reject the remaining two:

a) One of the ways of identifying the utterance meaning is to specify the normative consequences of the utterance (of the kinds generally characterized in the definition of the relevant speech act type).
b) The utterance meanings must be determined with normative force.
c) The assignments between expression types and their meanings must be normatively fixed.

Stephen Barker[19]

In the first part of his rich and inspiring paper Stephen focuses on the interpretation of explicit performatives (like „Hereby I order you to leave").[20] He defends the classical Austinian account, according to which performative verbs, as they appear in these sentences, serve to specify the illocutionary force of utterances (in which these sentences are used), rather than to assert that the speaker performs an act with that force. As Stephen points out (p. 75f.), sentences in which these verbs occur are indeed used to *describe* performative acts of the speakers, and in doing so are supposed to match reality (to represent how things are), but not in a way which is assessable as true or false. The idea that there are ways of representing reality which do not admit truth-evaluation is by far not uncontroversial and a considerable part of Stephen's paper is devoted to its explication and defence. Let me proceed in the reverted order, i.e. start with this latter, more general issue, and return to explicit performatives in the end.

I should say in advance that I do not see any real opportunity for a polemic confrontation with Stephen's views: instead, I will try to shortly sketch the reasons, why they are so close to my way of thinking about these issues. We seem to share the view that one should not waste the importance of Austin's lesson (which is, unlike the theoretical elaboration of Austin's position, fully in Wittgensteinian spirit) that there is a variety of ways in which our utterances may succeed or fail, equally relevant like truth or falsity in case of assertions. Truth/falsity may seem to be the only (or the most) relevant dimension of evaluation of utterances only if we focus exclusively (or primarily) on assertions – which we should avoid doing if we are really interested in how the actual human communication works.[21] If

---

[19]   Reply to Barker (2015). All the page references which appear without the author's name refer to this paper.

[20]   In fact, at least two other authors participating in the *Organon F* volume referred to were involved in the debates about this topic, namely Peter Pagin and Manuel García-Carpintero: it would be exciting to see them discussing with Stephen at a colloquium on explicit performatives (in some attractive place, if possible).

[21]   Nevertheless, in his discussion about explicit performatives, Stephen distinguishes the question whether they should or should not be approached as assertions from the question whether they admit truth-evaluation: after he answers the first question negatively, he brings independent arguments for the negative answer to the second one.

this is right, we should not have any problem with taking propositional contents of non-assertive utterances as determining other kinds of *conditions of satisfaction* (to follow John Searle's terminology) than the truth-conditions. I find Searle's term quite suitable[22] – and will stick to it, although Stephen does not use it. Instead, he speaks about *representation conditions* (p. 91f.), which is perfectly compatible with my terminological preference, since the satisfaction conditions of an utterance specify the state of affairs *represented* by the utterance as that state of affairs which, if it obtains (if it is a *fact*) in the world we are speaking about, would make the speech act performed in the utterance *satisfied* (in a way corresponding to its force). For instance, it would make an order obeyed, promise fulfilled or assertion true.

To be sure, once we say this we are confronted with various kinds of objections and widely shared ways of thinking about truth, meaning and related issues. Let me mention some of these challenges, more or less straightforwardly linked with Stephen's arguments.

(1) One of the implications of this view is that the true/false assessability must be justified on the level of illocutionary acts, rather than automatically belonging to locutionary acts. I fully agree with the way Stephen puts this point (cf., e.g., p. 93) and with his corresponding abandonment of what he labels as the *principle RT*:

> *RT*:   A sentence *S* is true/false-assessable iff *S* has propositional content, it is a representation of how things might be. (p. 76)

One may insist that the true/false assessability cannot depend on the assertive force of the utterance, since, as Stephen points out, "being unasserted does not imply not being true/false assessable. For example, constituent sentences of a logical compound are not asserted, but they are true/false assessable" (p. 82). Similarly, Peter Geach (1972) has argued against the doctrine of the "Oxford philosophers of language" that a proposition can be evaluated as true or false only if it is introduced into discourse in a way which raises the claim for truth, as it happens in assertions. He objected that it makes a good sense to evaluate as true or false both dis-

---

[22]   Obviously, the applications of the term "satisfaction" to expressions, rather than to utterances, like: "x satisfies 'is tall' iff x is tall", should be then read in an equally generalized way, i.e. not as automatically replaceable by "'is tall' is true of x iff x is tall".

juncts in, for instance, "It's raining or it's snowing", though neither of them is asserted. One may feel inclined to interpret this as showing that the truth evaluation can be applied to bare propositions, in abstraction from the force with which they are introduced into communication. But this is certainly not what happens in our example. Here, the truth evaluation of the disjuncts is justified by the assertion of the disjunction as a whole: we ask whether at least one of the disjuncts is true in order to find out the truth value of *what is asserted*. Similarly, an order can have disjunctive conditions of satisfaction, like in case of "Wash the car or return the money!": we ask whether at least one of these conditions has been fulfilled in order to find out whether the order has been *obeyed*. Of course, to say that an order has such and such conditions of being obeyed, or to say that these conditions are fulfilled, is to make assertions, and the latter assertion is (in our case) true if the addressee of the order has washed the car or returned the money. But this platitude clearly does not justify the claim that the disjunctive *content of the order* in our example has such and such *truth-conditions*.

(2) Perhaps one can still insist that even the analysis of the way the propositions are introduced into discourse in non-assertive acts, like order or promise, cannot do without the notion of truth. For instance, to order to a person P to wash the car is to order to P to behave in such a way as to make the proposition "P has washed the car" true. The question is what the possibility of such a paraphrase is supposed to show. As philosophers should know better than anybody else, it is (usually) quite easy to introduce into the analysis the concepts we want to have there: but such a maneuver does not in itself amount to discovering a new parameter of the phenomenon analyzed. It is always useful to recall Wittgenstein's remark (from Wittgenstein 1953, § 22):

> We might also write every statement in the form of a question followed by a 'Yes'; for instance: 'Is it raining? Yes!' Would this show that every statement contained a question?

Examples of this kind can be produced *ad libitum*: for instance, the assertion "P washed the car" can be paraphrased by saying: "P behaved in such a way that if he were ordered to wash the car, the order would have been obeyed". Does this show that assertions are parasitic on orders? I think everybody will agree that what I have offered is just an artificial and bizarrely

monstrous construction. And this is how we should comment also the suggestion from the beginning of this paragraph: the paraphrase of an order in terms of making a proposition true.

(3) It is certainly true that an analysis of one speech act type can help us in understanding other speech act types; and analogically for the types of conditions of satisfaction. But even here, i.e. in the order of explanation, there is no *a priori* reason to assign the privileged role to the notions of assertion and truth. Dummett (1978) gave us an opposite example when he based his considerations concerning truth conditions of certain types of assertions on considerations concerning conditions of obeying analogically structured orders. Similarly, Austin's analysis focusing on linguistic acts establishing institutional facts (as paradigmatic examples of "doing things with words") helps us to understand better what is going on in linguistic communication in general, including acts of assertion (after the performative/constative opposition has been abandoned). And the same holds for Searle's paradigmatic analysis of promise.

(4) Insisting on the central position of the notion of truth may be motivated by our appreciation of the role it plays in some theory (regarded as) successful in important respects. A prominent example is, of course, Davidsonian theory of meaning construed on the basis of Tarski's theory of truth. Stephen (on p. 85) mentions and criticizes Davidson's well-known attempt to show how such a theory, while keeping truth as its central notion, can deal with non-indicative sentences and non-assertive utterances (cf. Davidson 1984b). Lepore and Ludwig, in their attempt to provide a general account of logical form on the basis of Davidsonian truth-theoretic semantics, have admitted that "non-declarative sentences ... present an especially interesting challenge to any conception of logical form grounded in truth-theoretic semantics, since uses of them are neither true nor false" (Lepore – Ludwig 2001, 130). They focus on imperatives and notice that their uses "admit of bivalent evaluation, though they are not truth-valued". To account for this, they introduce the "notion of fulfillment conditions which subsumes both compliance conditions for imperatives and truth-conditions for declaratives" (Lepore – Ludwig 2001, 131). Nevertheless, their truth-theoretical approach commits them to "exhibit compliance conditions as recursively specifiable in terms of truth-conditions". This requires that they construe imperative sentences as having a "declarative core", assuming that "the compliance conditions for im-

peratives will then be exhibited as given in terms of the truth conditions for their declarative cores" (Lepore – Ludwig 2001, 131). So, the notion of "fulfillment conditions", which they present as "a generalization of the truth-theoretical approach" (Lepore – Ludwig 2001, 131), plays a subsidiary (if not just decorative) role within the truth-theoretical framework: no "interpretive fulfillment theory" as a generalization of "interpretive truth-theory" (Lepore – Ludwig 2001, 113) is in fact proposed. Obviously, this framework does not provide any space for asking the question whether the attractions of Davidson's theory, like its ability to systematically account for the compositionality of meaning or its ability to link (within the broader framework of the theory of interpretation) in a productive way meanings with beliefs and other types of attitudes, can be preserved if the notion of *truth* gives up its central position in favor of the more general notion of *fulfillment*.[23]

Davidson's fundamental considerations (esp. in Davidson 1984a) about the form of the desired theory of meaning (a theory which would, on a finite basis, generate interpretative theorems for each of the infinite number of sentences of the object-language), famously resulted in accepting – as one of the possible solutions – the model of Tarski's truth theory. I believe that they could, with small modifications, equally well result in accepting an alternative model with theorems of the form "S is satisfied iff p".[24]

(5) Utterances can represent states of affairs (as their conditions of satisfaction) with a variety of communicative functions, corresponding to the variety of speech act types. The specification of various ways in which propositions can be introduced into discourse then very much depends on the central notions on which our speech act theory is based. I do not think that speech act types can be defined in terms of speaker's intentions: at least on this level (on the level of definitions of speech act types) I would not say, with Stephen, that to X (i.e. to make a promise, assertion etc.) is "to produce a sentence with such and such intentions. That's what consti-

---

[23] Not surprisingly, the correlative notion of the conditions of satisfaction (in Searle's sense), together with the notion of conditions of success, play the central role in an attempt to build the Searlian type of speech act theory on formal basis: cf. Searle – Vanderveken (2009).

[24] I have tried to show this in Koťátko (2001). The same applies to the way in which McDowell (1980) specified the general aspirations of Davidson's programme and their fulfillment in the Tarskian type theory.

tutes *X*ing" (p. 78). It's true that in speech acts we manifest certain intentions, beliefs and other attitudes (in case of assertion that p it is the belief that p, the intention to produce or activate the belief that p in the audience etc.).[25] I believe that speech acts can be defined in terms of these manifestations – or, correlatively, in terms of commitments imposed by these manifestations on the speaker. But one can certainly manifest the belief that p and the intention to create or activate that belief in the audience without actually having any such belief and intention – and the utterance will still have the assertive force. The possibility of such discrepancies opens space for various kinds of linguistic tricks which, in turn, can serve as evidence that the manifestations (on which these tricks are based) indeed belong to the act of asserting, independently on the actual speaker's beliefs and intentions. The way in which the propositional variable p is imbedded within the specification of manifestations (or commitments) constituting an act of asserting that p is, within this approach, supposed to characterize the way in which a proposition is introduced into communication in assertion – and analogically for other types of speech acts. Within this framework, there is no space for assigning any privileged position to expressing propositions with assertive force, which opens space for truth-evaluation.

(6) As to the case of explicit performatives, I have always regarded the original Austinian version, defended by Stephen, as most intuitively plausible, found it natural that John Searle preserved it in the original version of his speech acts theory[26] (which exposed him to forceful attacks, cf. e.g. Stampe 1975) and has been surprised by Searle's conversion (in Searle 1989).[27] As Stephen points out, it is undeniable that in "I hereby order you to leave" the directive force is specified in another way than in "Leave!" In the former case the directive act is being *described* (p. 75) – and to admit this allows us to preserve the presumption that the meaning of the compo-

---

[25]   These are just two of the manifestations constituting the act of asserting. The second one is, moreover, simplified (the precise formulation must account for the plurality of possible communicative functions of assertions). A complete definition of assertion in terms of manifestations has been proposed e.g. in Koťátko (1998).

[26]   There, performative verbs are classified as "force indicating devices" precisely like "word order, stress, intonation contour, punctuation, the mood of the verbs"; cf. Searle (1965, 6).

[27]   I have had an opportunity to express that surprise already in 1990, when Searle gave a talk on this issue in Prague.

nents of the sentence contribute to the meaning of the whole in the standard compositional way (they jointly participate in the complex representation of how things are). But there is no reason to insist that the descriptive specification of the directive force is somehow *less direct* or straightforward than the indication of that force via the imperative mood. In particular, we should reject the presumption that what the descriptive specification in our case *directly* serves to must be an act of assertion – and then we will not face the task of construing a mechanism which would bring us from assertion (as a directly performed act) to an order (as an indirectly performed act).

An analogy with the general theory of descriptions might be illuminating. Some orthodox Russellians, when confronted with the actual communicative function of sentences like "The present French president is a socialist" (in particular with the typical ways we react to them and report about them), feel obliged to admit that descriptions can indeed serve to refer to individuals, but in an indirect way – on the level of implicatures (cf. Neale 1990). The right reply is, I think, that we need not invoke anything like the Gricean mechanism of implicatures (and ascribe to the speaker the corresponding communicative strategy, directed at generating quite a complex kind of reasoning in the audience) in order to justify the following report of John's utterance of our sentence: "John said about Holland that he is a socialist." And quite similarly, if John says to Paul "I hereby order you to leave", nothing like Gricean mechanism (and no correspondingly complex speaker's strategy) is needed to justify the claim that John ordered Paul to leave. On the contrary, reports like "John asserted that he ordered Paul to leave" would be, in the ordinary communication, most probably evaluated as bizarre – as totally missing what happened (what move has been made) in conversation. This argument concerns the way we report about performative utterances, but Stephen (among other things) draws our attention also to the speaker's communicative intentions: "Why would speakers want the extra illocutionary act to get in the way of *X*ing [e.g. ordering in our case or cursing in Stephen's example], since *X*ing is the main point" (p. 78).

Stephen refers to Bach and Harnisch as proponents of (one version of) the implicature-approach to explicit performatives and I fully share his critical attitude. In general, saying "I hereby order you to leave" is perhaps a more ceremonial or pompous, but not at all a less direct way of ordering

somebody to leave than saying "Leave!".[28] So, I agree with Stephen that "performatives are not indirect speech acts but primarily performances of orderings, commandings, adjournings, that is of *X*ings, and not acts that are constituted by assertions allowing us to infer that a performance of *X*ing is going on" (p. 79).

I also agree with Stephen (p. 81) that uttering "You ought to leave the room", unlike "Leave the room!" or "I order you to leave the room" can function as an assertion, but I would not describe it as an "utterance in which a speaker defends the state of desiring that the audience leaving the room". As far as I can see, uttering such a sentence with the affirmative force does not include manifestation of such a desire on the part of the speaker nor manifestation of an intention to produce such a desire in the audience. My evidence for this is that it is quite coherent to say: "You ought to leave the room, but I will be happy if you stay." Claiming that one ought to do something (according to some standards applying to her) and expressing the wish that she does not do so does not establish straightforward inconsistency, nor a communicative defect of the type of Moore's paradox.

Marián Zouhar[29]

An important part of Marián's work in the philosophy of language concerns rigid designation – and the present paper continues in this line in an innovative and most inspiring way. Following the tradition of our polemical

---

[28]   As to specific form of explicit performatives, my underdeveloped linguistic intuition in English does not enable me to appreciate all of Stephen's confrontations of cases in which the same verb (specifying some kind of illocutionary act) once has and once lacks performative function (cf. pp. 77-78 and 80). As far as I can see, no standard translation of these examples into Czech could serve to demonstrate that contrast. I cannot avoid mentioning an opposite case – one in which my Czech linguistic intuition enabled me to identify a mistake in an English article on speech acts. Searle (1975) claims that unlike the English sentence "Can you pass me the salt?", the Czech sentence „Můžete mi podat sůl?" (which is a straightforwrd translation of the former) cannot serve to perform an indirect speech act of a request. I have been pleased to demonstrate to him (sitting at the table of one of the Prague pubs) that it can.

[29]   Reply to Zouhar (2015). All the page references which appear without the author's name refer to this paper.

exchanges, I will abstract from all the points of agreement and focus on the rest, rather marginal in the context of Marián's inquiries. Let me start with Marián's quotation of Kripke's remark (see Kripke 1980, 21, footnote) explaining *de iure* rigidity as a case "where the reference of a designator is *stipulated* to be a single object, whether we are speaking of the actual world or of a counterfactual situation". I read this as saying that it is part of introducing an expression of this kind into our vocabulary that it is designed to play the referential role mentioned – in virtue of which rigidity belongs to its basic semantic status. "Introducing an expression into our vocabulary", as I understand it, can mean both making it part of the conventionally fixed reservoir of some sociolect, or using it for fulfilling an *ad hoc* communicative function – provided that this use is successful (i.e. that the communicative intention behind it is recognized by the audience – and on this basis the audience assigns the intended meaning to the utterance). When admitting both possibilities I no more follow Saul Kripke and I clearly depart from Marián's course of reasoning, as we will see shortly.[30]

Marián also goes beyond Kripke's explicit remarks on this issue when he asks "how it happens that proper names are rigid *de iure*?" (p. 104). Nevertheless, his reply is based on the Kripkean idea that proper names typically acquire their referential functions in the acts of baptism and on the interpretation of baptism as a convention establishing act. Granted this, "names designate their objects because of the linguistic conventions introduced during the baptismal acts" (p. 105). This, according to Marián, means that "an object needs not satisfy any descriptive condition to be designated by a given proper name. ... So, the name designates the object irrespectively of

---

[30]   When Kripke in (1977, 256) defines semantic referent as "given by a general intention of the speaker to refer to certain object whenever the designator is used", he adds (in footnote 20): "If the views about proper names I advocated in 'Naming and Necessity' are correct..., the conventions regarding names in an idiolect usually involve the fact that the idiolect is no mere idiolect, but part of a common language, where the reference can be passed from link to link." The way in which the term "convention", is used here, is certainly incompatible with its Lewisian (by far most influential) definition as a regularity prevailing in certain community, fixed by a complex of shared beliefs and preferences reflected in "common knowledge". But more importantly in our context, Kripke admits cases, corresponding to referential use of descriptions in Donnelan's sense, in which a proper name is (successfully) used to refer to an individual which is not its semantic referent (not to speak about referent assigned to it by conventions of some sociolect). I will refer to Kripke's well known example later.

virtually any property the object exemplified or might have exemplified" (p. 104; cf. also p. 115). When put so straightforwardly, the claim raises certain doubts, since a few lines above Marián described the property an object *must have* to become (in the typical way) the bearer of a name: namely to be linked with it in an act of baptism. In general, one can certainly say that an object can be a referent of a singular term, be it definite description, name or demonstrative (as used on particular occasion) provided that it meets certain empirical condition: it uniquely satisfies the description in question or has been assigned to the name in question as its bearer at the beginning of the chain to which the given utterance of the name belongs, or is the contextually most plausible candidate for the referent of the given use of a demonstrative. This, I believe, needs to be added to the contrast between descriptions and names drawn by Marián in Section 4: both linguistic conventions connecting names with objects and satisfaction relations between descriptions and objects are established by empirical facts (consisting in the objects' bearing certain empirical properties).

One might insist that these two kinds of facts (and correlatively two kinds of properties) should not be treated at the same level. Nevertheless, Marián treats them so in his exposition of the puzzle which is the central topic of his paper. The second of the two lines of reasoning which Marián introduces as generating the puzzle (p. 110) results in the statement that "there is at least one property P and at least one possible world w such that [an object] o exemplifies P with respect to w even though o fails to exist in w" (this is Marián's paraphrase of the conclusion from his Abstract). The property presented (at the same level with all other properties of objects) as justifying this claim is the semantic property of being named by (some name) α.[31]

But let me return to Marián's explanation of the *de iure* rigidity of proper names from their being introduced via conventions. My basic objections can be summarized as follows:

---

[31]   I have to add that the way out of the paradox, which Marián apparently finds most promising, includes resignation on construing this property (together with other language-dependent properties of objects) in the standard way, i.e. as an intension (function from possible worlds to extensions). His defense of this move (p. 115) is rather problematic for those, who approach language, including semantic relations between names and their bearers and linguistic conventions fixing these relations, as part of our world, and hence depending on how things are in this world.

(1) As I have pointed out, the link between a name and the individual designated by it needs not be established in a conventional way. For instance, some expression can effectively function as a name *ad hoc* – and even then it designates rigidly and its rigidity is *de iure*, in virtue of its semantic status of a name. For instance, I can say "Plato is coming!" or "Jacket is coming!" in a context in which it is clear to you as my audience that I mean our colleague John, although I have never used these expressions to refer to John before.[32] Then the terms "Plato" and "Jacket" function as proper names of John and *eo ipso* as rigid designators, without any convention governing our future linguistic behaviour being established. You may find this way of referring to John unacceptable, there need not be any reason to expect that I will continue in this practice on any future occasion, nevertheless, if my referential intention has been recognized by you and you have interpreted my utterance on this basis, the expressions "Plato" or "Jacket" functioned as names of John – and they did so successfully, i.e. in a way which led to understanding.

(2) The opposition between the referential function's being established via convention and it's being established in virtue of the fact that certain individual satisfies certain descriptively specified conditions[33] is, from my point of view, difficult to accept, not only for the reason mentioned above. The convention introducing the name N for the person P can precisely make P being the referent of N in virtue of the fact that P satisfies certain description conventionally linked with the name as fixing its referent. This is the case of the name "Jack the Ripper" in Kripke's well-known example: it has been introduced for the person who, in the actual world, committed those and those murders (but has never been revealed). Then it refers with respect to any possible world to that individual who committed those murders in the actual world.[34] Similarly, the description "the person who

---

[32]   Or, to slightly modify Kripke's scenario (from Kripke 1977): I say "The old Smith is coming" – in a context in which it is clear that I intended to speak about Jones but mistook the names.

[33]   Cf., e.g., p. 113 f.: "Anyway, we cannot retain both the idea of conventionally established name-bearer relations and the idea of name-bearer relations being determined such that the bearer of a name satisfied some kind of condition" (p. 114).

[34]   I find this incompatible with the way in which Marián presents the contrast between "the idea of conventionally established name-bearer relations" and "the idea of satisfactionally established name-bearer relations" in his footnote 28 (p. 114).

committed those and those murders" can in some uses function rigidly, i.e.
to refer, with respect to any possible world, to the person who uniquely sa-
tisfies it in the actual world, as it is in the utterance of "The person who
committed those and those murders need not have killed anybody – if he
properly studied Kant's second *Critique*". In both cases the term refers to
certain person with respect to all possible worlds due to the fact that it sa-
tisfies certain description in the actual world (plus, in the former case, in
virtue of the fact that the referent of the name "Jack the Ripper" has been,
in the actual world, fixed by means of that description). Nevertheless, in
the former case, the term is rigid *de iure*, due to its semantic status of
a proper name, while in the latter case it is rigid "merely" *de facto*, due to its
status of description (a kind of expression which does not include rigidity
as part of its general semantic characteristics).

(3) I agree with Marián, David Kaplan and others that we should ap-
proach names as obstinately (rather than persistently) rigid, i.e. that they
"designate the same thing with respect to every possible world, whether
that object exists there or not" (as Marián quotes from Salmon) and I share
Marián's view that this is "the other side" of their rigidity *de iure* (cf. p.
106).[35] But I cannot join Marián when he claims that this correlation stems
from the fact that "both features can be explained in terms of the conven-
tionally determined link between proper names and their bearers" (p. 106),
due to the objections presented in (1). Marián develops his point in consid-
erations concerning the order in which proper names are introduced into
language and possible worlds into our apparatus (cf. his argument on p. 107
f.). I find these things hardly comparable: obviously, language with its
proper names and their referential functions, with its devices for describing
actual as well as counterfactual states of affairs etc. comes first – and the
terms like "proper names" or "possible worlds" have been introduced as in-
struments for the description and analysis of these phenomena. But once
we have them, they enable us to say that proper names, independently of
the development of linguistics or logical semantics, always served to refer to
individuals with respect to the actual world as well as to other possible
worlds and that in fulfilling this task they always functioned as rigid desig-

---

[35] Cf. also Kripke (1980, 21, footnote): "Since names are rigid *de iure*... I say that
a proper name designates its referent even when we speak of counterfactual situations
where that referent would not have existed."

nators. What we attempt to capture here by means of the term "possible worlds" was always part of the communicative potential of our language – if it is true that it always served (among other things) to speaking about counterfactual states of affairs. I find it difficult to imagine the position from which I could say, with Marián, that "the name-bearer relation is not dependent on any possible world whatsoever". Here, the speaker manifestly exploits the possible world apparatus: then he should be able to say that the name-bearer relation is based on things which happened in the actual world (typically, but not necessarily, the act of baptism) and that the relation is (in virtue of its being naming relation) such that the name designates the same object with respect to all possible worlds (including those in which that object does not exist). This clearly does not explain how it comes out that names are obstinately rigid designators: what I want to say is just that it is not quite easy for me to accept Marián's explanation.

As to the presentation of Marián's puzzle, I have a problem, mentioned in another context above, with the second step in the second line of reasoning, leading to one of the incompatible claims involved in the constitution of the puzzle. I mean the following assumption: "2. For any linguistic convention c and any possible world w it holds that c is in force regardless of how things are in w" (p. 109). The question is how could a convention c be in force in some world w without certain things happening in w – things constituting and preserving the convention in some community. David Lewis' famous definition provides one possible list of such things (conditions which must be fulfilled in the community in question): if you accept this definition, you have to admit that once certain beliefs and preferences (of the kinds specified in the definition) cease to prevail in the community of citizens of the Czech republic, Czech language will cease to be the language conventionally fixed as the language of that community. Similarly for any alternative list of conditions you might adopt. So, "Praha" designates Prague in Czech in virtue of certain things being the case in the actual world. It refers to Prague also with respect to any other possible world regardless of how things are there (even the existence of Prague there is not required and, obviously, linguistic conventions which are in force there do not matter at all). But if the relevant things change in the actual world, "Praha" will cease referring to Prague in Czech with respect to the actual as well as any other possible world. Marián insists that "everyone who believes in rigidity *de iure*" should accept 2 "without much ado".

I firmly believe in rigidity *de iure* but don't know how to get rid of the ado just voiced.

For the reasons mentioned above in the remark (1), I have some problems also with the first step in the same argument, namely with adopting the claim "1. For any expression *e* it holds that if *e* is a proper name then *e* is a *de iure* rigid designator and there is an object *x* and a linguistic convention *c* such that *e* designates *x* on the basis of *c*" (p. 110). But one can still obtain the conclusion of this argument (and hence the puzzle) if one starts with a modified version of the step 3, namely with adopting the claim: "3′. If there is a linguistic convention *c* and a name α such that α rigidly designates *o* on the basis of *c*, then for any possible world *w* it holds that α rigidly designates *o* with respect to *w* regardless of *o*'s existence or non-existence in *w*." This is based on the assumption that proper names are obstinately rigid designators but does not rest on the stronger assumption 2. Second, 3′ mentions a conventionally established relation between a name and an object designated by it, without committing us to the assumption that the naming relation must be based on a convention, as the assumption 1 did.

The line of reasoning then can continue in the way suggested by Marián to the resulting claim incompatible with the conclusion of the first Marián's argument – and so we get the puzzle without being involved into the controversies mentioned above. This opens way to Marián's considerations about possible ways of solving the paradox and their costs, which I find brilliant and illuminating.

References

BARKER, S. (2015): Performatives and the Role of Truth in Semantics. *Organon F* 22, Suppl. Issue, 74-95.

DAVIDSON, D. (1984a): Truth and Meaning. In: Davidson, D.: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

DAVIDSON, D. (1984b): Moods and Performances. In: Davidson, D.: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

DAVIDSON, D. (1994): The Social Aspect of Language. In: McGuiness, B. – Oliveri, G. (eds.): *The Philosophy of Michael Dummett*. Dordrecht – Boston – London: Kluwer Academic Publishers.

DUMMETT, M. (1973): *Frege: Philosophy of Language*. London: Duckworth.

DUMMETT, M. (1976): What Is a Theory of Meaning? In: Evans, G. – McDowell, J. (eds.): *Truth and Meaning*. Oxford: Oxford University Press.

DUMMETT, M. (1978): Truth. In: Dummett, M.: *Truth and Other Enigmas*. London: Duckworth.

DUMMETT, M. (1991): *The Logical Basis of Metaphysics*. Cambridge (Mass.): Harvard University Press.

FREGE, G. (1979): Logik in Mathematics. In: Frege, G.: *Posthumous Writings*. Oxford: Blackwell, 203-250.

FREGE, G. (1980): *Gottlob Frege's Briefwechsel mit D. Hilbert, E. Husserl, B. Russell sowie ausgewählte Einzelsbriefe Freges*. Hamburg: Felix Meiner Verlag.

GEACH, P. (1972): Assertion. In: *Logic Matters*. Oxford: Oxford University Press.

GLÜER, K. – WIKFORSS, Å. (2015): Meaning Normativism: Against the Simple Arguments. *Organon F* 22, Suppl. Issue, 63-73.

GRICE, P. (1989): Meaning Revisited. In: *Studies in the Ways of Words*. Cambridge (Mass.): Harvard University Press.

KOŤÁTKO, P. (1995): Frege's Sinn and Utterance Meaning. In: Max, I. – Stelzner, W. (eds.): *Logik und Mathematik. Frege-Kolloquium Jena 1993*. Berlin – New York: Walter de Gruyter, 303-313.

KOŤÁTKO, P. (1998): Two Notions of Utterance Meaning. *Proceedings of the Aristotelian Society* 98, 225-239.

KOŤÁTKO, P. (2000): Mutual Beliefs and Communicative Success. *Theoria* 15, No. 39, 421-433.

KOŤÁTKO, P. (2001): Ke starým (i novějším) nárokům pojmu pravdy v teorii významu a interpretace. In: Nosek, J. (ed.): *Úvahy o pravdivosti*. Praha: Filosofia, 154-164.

KOŤÁTKO, P. (2006): Having a Concept. In: Marvan, T. (ed.): *What Determines Content? The Internalism/Externalism Dispute*. Cambridge: Cambridge Scholars Press, 55-68.

KOŤÁTKO, P. (2012): Searle's Defence of Internalism. *Organon F* 19, Suppl. Issue 2, 93-106.

KRIPKE, S. (1980): *Naming and Necessity*. Oxford: Blackwell.

KRIPKE, S. (1977): Speaker's Reference and Semantic Reference. In: French, P. A. – Uehling, T. E. – Wettstein, H. (eds.): *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesotta Press.

LEPORE, E. – LUDWIG, K. (2001): What Is Logical Form. In: Koťátko, P. – Pagin, P. – Segal, G. (eds.): *Interpreting Davidson*. Stanford: CSLI.

LEWIS, D. (1983): Languages and Language. In: *Philosophical Papers*. Vol. 1. Oxford: Oxford University Press.

McDOWELL, J. (1980): Meaning, Communication and Knowledge. In: Van Straaten, Z. (ed.): *Philosophical Subjects: Essays Presented to P. F. Strawson*. Oxford: Oxford University Press.

NEALE, S. (1990): Descriptions. Cambridge (Mass.): The MIT Press.

PAGIN, P. (2015): Intending to be Misinterpreted. *Organon F* 22, Suppl. Issue, 5-18.

PUTNAM, H. (1975): The Meaning of „Meaning". In: *Mind, Language and Reality*. Cambridge: Cambridge University Press.

QUINE, W. V. O. (1960): *Word and Object*. Cambridge (Mass.): The MIT Press.

SEARLE, J. (1965): What Is a Speech Act? In: Black, M. (ed.): *Philosophy in America*. London: Allen and Unwinn.

SEARLE, J. (1975): Indirect Speech Acts. In: Cole P. – Morgan, J. L. (eds): *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press, 59-82.

SEARLE, J. (1989): How Performatives Work. *Linguistics and Philosophy* 12, 535-558.

SEARLE, J. – VANDERVEKEN, D. (2009): *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

SCHIFFER, S. (1972): *Meaning*. Oxford: Oxford University Press.

SCHIFFER, S. (1987): *Remnants of Meaning*. Cambridge (Mass.): The MIT Press.

STAMPE, D. W. (1975): Meaning and Truth in the Theory of Speech Acts. In: Cole, P. – Morgan, J. L.(eds.): *Speech Acts* (*Syntax and Semantics* vol. III). New York – San Francisco – London, 1-39.

STJERNBERG, F. (2015): Frege on Knowing and Individuating Senses. *Organon F* 22, Suppl. Issue, 39-62.

VOLTOLINI, A. (2015): Heidegger's Logico-Semantic Strikeback. *Organon F* 22, Suppl. Issue, 19-38.

WITTGENSTEIN, L. (1953): *Philosophical Investigations*. Anscombe G. E. M. – Rhees, R. (eds.), Oxford: Blackwell.

ZOUHAR, M. (2015): A Puzzle about Rigid Designation. *Organon F* 22, Suppl. Issue, 96-117.

John Etchemendy: *O pojmu logického vyplývání*
Preložil Petr Hromek. Ostrava: Ostravská univerzita 2014, 231 strán

Recenzovaná kniha je českým prekladom známej práce Johna Etchemendy-ho *The Concept of Logical Consequence* z roku 1999.[1] Knihu preložil z anglického originálu Petr Hromek. Etchemendy v tejto monografii predkladá pozoruhodnú kritiku Tarského (sémantickej) analýzy logickej pravdy a následne aj logického vyplývania. Tarského analýza sa štandardne prijíma, preto je Etchemendyho pozícia pomerne kontroverzná. Čitateľ, ktorý zrejme v súlade s hlavným prúdom sympatizuje s Tarského koncepciou logického vyplývania, tak stojí pred úlohou vyrovnať sa s Etchemendyho argumentáciou.

Knižka pozostáva z dvanástich kapitol, ktoré, ako píše Etchemendy, tvoria vlastne jeden dlhý argument. Po nich nasleduje stručný, ale výstižný doslov, ktorý napísal prekladateľ knihy. Doslov neobsahuje iba zhrnutie hlavných téz a výsledkov knihy, ale aj krátky exkurz do dejín logiky, zaradenie knihy do kontextu modernej logiky, ďalší vývoj Etchemendyho pozície, či reakcie na jeho dielo.

Etchemendyho argumentácia proti štandardnej Tarského analýze je skutočne spletitá: Etchemendy najprv zavádza reprezentačnú sémantiku, ktorú neskôr odlišuje od interpretačnej sémantiky. Vysvetľuje tiež Bolzanovu substitučnú analýzu logickej pravdy, ktorú následne odlišuje od Tarského definície logickej pravdy pomocou pojmu splňovania a usiluje sa ukázať, že Tarského návrh je úspešnejší ako ten Bolzanov. Komparácia reprezentačnej a interpretačnej sémantiky či Bolzanovej a Tarského analýzy nie je hlavným cieľom knihy, preto sa pokúsim prejsť priamo k jadru Etchemendyho argumentácie.

Etchemendy tvrdí, že logicky pravdivý výrok je pravdivý výlučne vďaka významu logických termínov, ktoré obsahuje (s. 135). Následne obhajuje tézu, že Tarského analýza takýto pojem logickej pravdy nezachytáva (najmä v kap. 7, 8 a 9), pretože stotožňuje logickú pravdivosť nejakého výroku S s obyčajnou pravdivosťou uzáveru $\forall v_1...\forall v_n[S(e/v)]$ (zjednodušene povedané, tento uzáver vznikol nahradením všetkých výrazov v S, ktoré nie sú nemenné, premennými, ktoré boli následne viazané všeobecnými kvantifikátormi). Treba upozorniť, že je vždy potrebné najprv zvoliť určitú množinu nemenných termínov. Ak budeme za nemenné považovať štandardné logické konštanty, analýza povedie k lepším výsledkom. Kľúčový problém však spočíva v tom, že tento uzáver nemusí

---

byť pravdivý len vďaka významu logických termínov. Jeho pravdivosť môžu zabezpečovať mimologické fakty rôzneho druhu: „historická pravda, nějaký obskurní aritmetický nebo množinový teorém, dokonce to může být pravda jen
čistě náhodou" (s. 136). Z týchto dôvodov vedie Tarského analýza k nesprávnym výsledkom.

Uvediem jednoduchý príklad, ktorý síce možno poľahky kritizovať, dobre
však osvetlí spôsob Etchemendyho kritiky Tarského analýzy. Ak sú napríklad
aktuálne všetci senátori muži a logická pravdivosť výroku

(V)   Ak je Leslie senátor, tak je Leslie muž

sa stotožní s obyčajnou pravdivosťou

(V′) $\forall x$[Ak je x senátor, tak je x muž],

tak (V′) bude pravdivý. Potom však budeme musieť (podľa Tarského analýzy)
uznať výrok (V) za logickú pravdu. (V) však očividne nie je logická pravda, pretože by sa veci *mohli* mať inak. Keby existovali senátorky, už by nebolo pravdivé
(V′) ani (V). Tarského analýza preto v tomto prípade poskytla nesprávny výsledok. Ak by aj existovali senátorky a Tarského analýza správne tvrdila, že (V) nie
je logicky pravdivý, stalo by sa tak vďaka náhodnému empirickému faktu, nešlo
by o záruku poskytnutú Tarského analýzou.

Mimologické vplyvy, ktoré uvádza Etchemendy, sú podľa môjho názoru
troch zásadne odlišných druhov: (i) predpoklady analýzy, (ii) empirické fakty
a (iii) matematické či logické fakty, ktoré nie sú predpokladmi analýzy. Domnievam sa, že proti Etchemendyho argumentácii v prípade (i) a (ii) možno
vzniesť vážne námietky a brániť tak štandardnú analýzu. Naopak, Etchemendyho kritika je v prípade (iii) zrejme opodstatnená a poukazuje na pozoruhodnú
črtu logickej pravdy (a logického vyplývania).

Etchemendy očividne považuje za „mimologické" aj axiómy teórie množín či
významové postuláty, o čom svedčí napríklad tento citát: „standardní analýzu
nutí odvolat se na axiom nekonečna, tj. evidentně na předpoklad, který nijak
nesouvisí s logikou" (s. 158). Toto tvrdenie sa môže zdať čitateľom absurdné:
Sotva možno plauzibilne tvrdiť, že axiómy a významové postuláty sa nijako netýkajú logiky či významu. Ako teda môže Etchemendy niečo také tvrdiť? V pozadí tohto názoru je pravdepodobne hon za intuitívnym (predteoretickým) pojmom logickej pravdy (či logického vyplývania). Prekladateľ knihy v doslove
vhodne položil otázku, či má byť Tarského analýza teóriou, explikáciou alebo
modelom tohto pojmu. Treba však dodať, že žiadne teoretické uchopenie
predteoretického pojmu nemôže byť s týmto predteoretickým pojmom identické. Navyše, každé teoretické uchopenie pojmu logického vyplývania už predpo-

kladá nejakú logiku či sémantiku. Často voleným rámcom je Zermelova-Fraenkelova teória množín s axiómou výberu (ZFC). Ak predpokladáme (pri štandardnej voľbe logických konštánt) ZFC, musíme sa zmieriť s tým, že výrok „Existuje aspoň 27 entít" bude logickou pravdou (vďaka vyššie spomínanej axióme nekonečna). Ak predpokladáme klasickú logiku, musíme sa zmieriť s tým, že „Prší alebo neprší" bude logickou pravdou. Ak pracujeme s nejakým jazykom prvého rádu, ešte pred konkrétnymi analýzami musíme špecifikovať syntax a sémantiku. Vymedzenie univerza diskurzu a obvykle aj sémantiky kvantifikátorov je záležitosť, ktorú musíme vybaviť skôr, ako sa budeme pýtať na logické pravdy. Etchemendymu sa pritom nepozdáva ani taká samozrejmá požiadavka, že vlastné meno môže „pomenovávať libovolné individuum – za předpokladu, že toto individuum *je prvkem* oboru diskurzu" (s. 99). Indivíduom sa predsa štandardne rozumie práve prvok univerza diskurzu (minimálne v prvorádovej logike). Podobne nepresvedčivá je Etchemendyho kritika významových postulátov a požiadavky obmedzenej substituovateľnosti či jeho diskusia o kvantifikátoroch. Tam, kde Etchemendy vidí bludný kruh, vidí väčšina logikov predpoklady, bez ktorých ani nemožno začať analýzu. Nakoniec, už sám termín „logická pravda" predsa napovedá, že má ísť o akési pravdy logiky či pravdy založené na logike. Ak aspoň implicitne nepredpokladáme nejakú logiku, len sotva možno rozhodnúť, či nejaký výrok je alebo nie je logickou pravdou. V tomto prípade preto považujem Etchemendyho kritiku za neuspokojivú.

Druhým druhom sú mimologické fakty *par excellence*, totiž rôzne empirické (historické, kontingentné...) fakty. Sem spadá aj vyššie diskutovaný výrok (V) (*Ak je Leslie senátor, tak je Leslie muž*). V tomto prípade nemožno nesúhlasiť s Etchemendym, keď tvrdí, že (V) nie je logickou pravdou. Nesúhlasím však s Etchemendyho tvrdením, že „Tarského analýza (...) *nedokáže adekvátně zachytit žádné modální*, epistemické či sémantické vlastnosti našich původních pojmů logické pravdy a logického vyplývání" (s. 157, kurzíva D. G.). Pokúsim sa tento nesúhlas odôvodniť na príklade výroku (V). Etchemendy veľmi správne postrehol, že logická pravdivosť (či logické vyplývanie) má určitú modálnu črtu. Vo všeobecnosti sa nám nepozdáva, ak naša teória vyhodnotí ako logickú pravdu výrok, ktorý nie je pravdivý nevyhnutne, ale iba náhodou. Ak však chceme použiť Tarského analýzu, a zároveň reprezentovať modálny aspekt jazyka, nemôžeme výrokom poskytnúť analýzy, ktoré nie sú vôbec modálne. Vetu prirodzeného jazyka „Ak je Leslie senátor, tak je Leslie muž", by sme preto mali korektnejšie zapísať

(VW)   Ak je Leslie senátor vo svete W, tak je Leslie muž vo svete W.

Logickú pravdivosť (VW) by sme potom mohli stotožniť s obyčajnou pravdivosťou

(VW′) $\forall x \forall w$[Ak je x senátor vo w, tak je x muž vo w].

(VW′) by však nebola pravda ani za predpokladu, že aktuálne sú všetci senátori mužmi, a preto výrok (VW) naša upravená analýza nevyhodnotí ako logickú pravdu. Z týchto dôvodov tvrdím, že sa Etchemendy vo vyššie uvedenom citáte mýli. Spôsobom, ktorý som navrhla, možno vyriešiť problém so všetkými empirickými „hrozbami" pre Tarského analýzu. Treba však upozorniť, že Etchemendy sám zvažoval návrh pracovať s nevyhnutnou (nie obyčajnou) pravdivosťou Tarského uzáverov. Takéto riešenie však považoval za natoľko výraznú zmenu oproti Tarského pôvodnej analýze, že ju už nemožno považovať za vylepšenie Tarského analýzy. V prípade Etchemendyho návrhu to tak skutočne je. Moje riešenie je však v zásadnom ohľade odlišné: Tarského analýzu ponecháva bezo zmeny, iba navrhuje, aby sme (empirickým!) vetám prirodzeného jazyka poskytli adekvátnejšiu analýzu, t. j. takú, ktorá rešpektuje ich modálnu variabilitosť a neponechává ich v pazúroch extenzionalistickej sémantiky.

Zatiaľ sa teda zdá, že možno obhájiť štandardnú analýzu voči Etchemendyho útokom. Vec sa však komplikuje v prípade (iii). Bolo by absurdné tvrdiť, že nejaký komplikovaný, sofistikovane dokázaný dôsledok prijatého teoretického rámca je predpokladom analýzy (napríklad v ZFC Cantorova veta, Cantorova-Schröderova-Bernsteinova teoréma, Princíp dobrého usporiadania, atď.). Teorémy preto predstavujú samostatnú, tretiu skupinu. Podľa Tarského analýzy logického vyplývania každý takýto dôsledok danej teórie vyplýva z akéhokoľvek iného výroku (s. 66). Ak prijmeme Tarského analýzu, zaväzujeme sa napríklad k tomu, že Cantorova veta logicky vyplýva z axiómy dvojice. Ďalší vhodný príklad formuluje Etchemendy v tejto výstižnej pasáži:

> hypotéza kontinua, i kdyby tento metateorém teorie množin byl pravdivý, nemůže být logickým důsledkem axiomu neuspořádané dvojice. A přesto, je-li hypotéza kontinua pravdivá, pak zřejmě pravdivá nutně, a tudíž je nemožné, aby axiom neuspořádané dvojice byl pravdivý a hypotéza kontinua nepravdivá. Tento postřeh naznačuje, že modalita, o niž se nám jedná, má spíše epistemickou povahu. (s. 106-107)

Tieto problémy sú analogické paradoxom materiálnej implikácie. Zdá sa totiž, že premisy logicky platného úsudku by mali obsahovať všetko, čo zaistí pravdivosť záveru, alebo, ako písal Aristoteles v pasáži, ktorú citoval Etchemendy i prekladateľ v doslove, „aby ono jiné nutně nastalo, nepotřebujeme kromě těch předpokladů nic dalšího" (s. 106, s. 205) – azda s výnimkou najzákladnejších pravidiel pre pravdivostné spojky a kvantifikátory. Netvrdím, že takú-

to argumentáciu nemožno kritizovať – domnievam sa však, že ide o najpresvedčivejšiu časť Etchemendyho kritiky.

Teraz upozorním na pár drobností týkajúcich sa prekladu. Tabuľka pre negáciu na s. 26 prekladu je chybná, v origináli na s. 16 preklep nie je. Očividných preklepov je celkovo málo, napríklad „výrků" na s. 32, z logického hľadiska nevhodná spojka *a* na s. 50 prekladu (v origináli na s. 36 nájdeme spojku *or*) či chýbajúca bodka v poznámke na s. 209. Bežnejšia notácia s vektormi na s. 104 originálu by čitateľovi bola zrejme zrozumiteľnejšia ako použitie hrubého fontu na s. 136 prekladu. Treba poznamenať, že preklad je často voľnejší, v niektorých prípadoch zbytočne (ako napríklad v prípade názvu kapitoly *Logic from the Metatheory* preloženého ako *Logika Tarského metateorie* či pridanie krstných mien na s. 13 a 15, ktoré v origináli na s. 5 a 7 prítomné nie sú a pod.), inokedy však v prospech väčšej zrozumiteľnosti a lepšej štylistiky (ako napríklad v prípade výrazu *cross-term restriction* preloženého ako *požiadavka obmedzenej substituovateľnosti* či rozdeľovania dlhších súvetí na kratšie vety). Prekladateľ sa vyhol „počešťovaniu" veľkej časti príkladov, mohol tak však urobiť jednotne (Leslie či prezidenti USA ostali v pôvodnej verzii, no napríklad z Freda sa stal Honza). Koncové poznámky boli nahradené poznámkami pod čiarou, čo poctivému čitateľovi výrazne uľahčí čítanie. Prehľadnejšie je aj prekladateľovo členenie obsahu, ktoré obsahuje aj podkapitoly. Zároveň bolo doplnených mnoho vysvetľujúcich poznámok, ktoré nepochybne pomôžu najmä čitateľovi so slabším logickým a matematickým vzdelaním.

Nemožno nesúhlasiť s prekladateľom, že Etchemendyho knižka je zaujímavým a provokatívnym čítaním. Jej preklad do češtiny preto možno len uvítať. Knižka predpokladá základnú znalosť logiky a zároveň dostatočnú trpezlivosť pri filozofických odbočkách. Najviac preto poteší tých, ktorí sa zaujímajú o filozofickú logiku.

Daniela Glavaničová
dada.baudelaire@gmail.com

Literatúra

Etchemendy, J. (1999): *The Concept of Logical Consequence*. Stanford: CSLI Publications.

Contents · Obsah

DISCUSSIONS ▪ DISKUSIE

BOOK REVIEWS ▪ RECENZIE

REPORTS ▪ SPRÁVY

MISCELLANEA ▪ MISCELANEÁ