

## Contents

### ARTICLES

Adrian BRICIU: On Context Shifters and Compositionality in Natural Languages .....	2
Marie DUŽÍ – Jakub MACEK: Analysis of Time References in Natural Languages by means of Transparent Intensional Logic .....	21
Pavel MATERNA: On Tichý’s Attempt to Explicate Sense in terms of Turing Machines .....	41
Joachim LIPSKI: Radical Rationalization Accommodates Rampant Irrationality .....	53
Richard VALLÉE: Fictional Names and Truth .....	74
Ondřej BÍBA – Jitka PAITLOVÁ: Grundprobleme, or Popper Meets Kant .....	100

### DISCUSSIONS

Miloš KOSTEREC: Some Problems of Glavaničová’s Approach to Fictional Names .....	120
--	-----

### BOOK REVIEWS

Tomáš KOLLÁRIK: F. Gahér, V. Marko, <i>Method, Problem, and Task</i> .....	126
--	-----

### REPORTS

Martin VACEK: Modal Metaphysics: Issues on the (Im)possible V .....	132
Daniela GLAVANIČOVÁ: Current Trends in Deontic Logic .....	134
Martin VACEK: Deflationism in Metaphysics .....	136

## On Context Shifters and Compositionality in Natural Languages

ADRIAN BRICIU<sup>1</sup>

**ABSTRACT:** My modest aim in this paper is to prove certain relations between some type of hyper-intensional operators, namely context shifting operators, and compositionality in natural languages. Various authors (e.g. von Fintel & Matthewson 2008; Stalnaker 2014) have argued that context-shifting operators are incompatible with compositionality. In fact, some of them understand Kaplan's (1989) famous ban on context-shifting operators as a constraint on compositionality. Others, (e.g. Rabern 2013) take context-shifting operators to be compatible with compositionality but, unfortunately, do not provide a proof, or an argument in favor of their position. The aim of this paper is to do precisely that. Additionally, I provide a new proof that compositionality for propositional content (intension) is a proper generalization of compositionality for character (hyper-intensions).

**KEYWORDS:** Compositionality – context-shifting operators – indexicals – natural language semantics.

---

<sup>1</sup> Received: 12 May 2017 / Accepted: 20 July 2017

✉ Adrian Briciu

LOGOS Research Group, Universitat de Barcelona  
C. Montalegre, 6-8, desp. 4009, Barcelona 08001, Spain

West University of Timisoara  
Bv. V. Parvan, 4, et.5, 300223 Timisoara, Romania

e-mail: adrian.briciu@gmail.com

## 1. Introduction

Formal semantics has recently witnessed a flurry of interest on whether natural languages contain a special type of hyper-intensional operators (i.e. context-shifting operators) or not (see Rabern & Ball 2017, Santorio 2012 and the many references therein). Ever since Kaplan (1989) there has been a certain resistance in accepting that some natural language expressions are best modeled as context-shifting operators. One reason for this resistance is that, apparently, such operators cannot be handled compositionally.

The hypothesis that natural languages have a compositional semantics is usually taken to play a part in explaining their productive feature; that is in explaining speakers' ability to produce and understand complex expressions that they have never encountered before. Compositionality is, thus, a fundamental tenet of formal semantics, one that formal semanticists are extremely reluctant to renounce. Though, of course, some philosophers of a Wittgensteinian and Austinian bent who doubt that formal semantics is a viable project (because it can't model the pervasive and radical forms of context sensitivity present in natural languages), are ready to deny that natural languages are compositional (see Travis 1996). The formulation of compositionality most commonly found in philosophical and linguistic literature is the following: a language is compositional if the meaning of each of its complex expressions is determined by their syntactic structure and the meanings of their respective constituents. Although there is no consensus on the precise interpretation this general formulation, most semanticists take it to mean that a language is compositional if the meaning of each of its complex expressions is *a function* of their syntactic structures and of the meaning of their respective constituents.<sup>2</sup> As Partee (1995, 153) points out

---

<sup>2</sup> There are dissenters, though. For example, Szabó (2000; and 2013) argues that the intuitive formulation of compositionality (the meaning of complexes is *determined* by the meaning of constituents and syntactic structure) is stronger than, and not captured well by the formulation of compositionality in terms of functions: the meaning of a complex expression is *a function* of the meaning of its constituents and its syntactic structure. This is correct, but the solution is to put constraints on what functions can be employed as composition functions in natural language semantics. The principle of compositionality defined in terms of functions claims only that the meaning of a complex expression is a function of the meanings of its constituents and its syntactic structure. The principle is silent with respect to what that function can

“the principle can be made precise only in conjunction with an explicit theory of meaning and of syntax, together with a fuller specification of what is required by the relation ‘is a function of.’” This is to say that different specifications of syntax, of what meaning is, and of what meaning rules (i.e. functions) are acceptable in natural language semantics can give rise to different interpretations of the principle, and these interpretations might settle different constraints on semantic theories. In other words, in order to give a formally precise formulation of compositionality two things are needed: an account of how complex expressions in natural languages are syntactically generated and an account of what meaning is. In what follows I’ll give only the minimal details of how syntax of natural languages and their meanings can be formally modeled; details that suffice for a precise formulation of compositionality.

The complex expressions of a language are generated by a syntax  $\Sigma$ , which can be represented as an algebra  $\Sigma = (E, A, F)$ , where  $E$  is the set of linguistic expressions of the language,  $A$  the set of simple expressions of the language (thus  $A \subset E$ ), and  $F$  the set of syntactic rules of the language. Members of  $F$  are functions defined over  $E$  and with values in  $E$ , and  $E$  is closed under operations in  $F$  (i.e. every member of  $E$  is either a member of the subset  $A$  or is the value of an operation of  $F$  on members of  $E$ ). Given that natural language expressions belong to different grammatical categories, and that each syntactic rule concerns only expressions of certain grammatical categories and not of others, each member of  $F$  is defined over expressions of certain grammatical categories and with values in expressions of certain grammatical categories. In other words, each syntactic rule is specified in terms of the grammatical categories of its arguments as well as the category of its value. One way to formally implement this insight is to take the set  $E$  of expressions to be an indexed

---

be. But, obviously not any function can be employed as a composition function in natural languages, for although there are an indefinite number of ways in which meanings can combine, some of them could not possibly be ways in which *meanings in natural languages* combine. In other words, there must be a restriction on what functions are acceptable composition functions (i.e. ways of combining meanings) for natural languages. This amounts to restricting the class of acceptable semantic rules in natural languages. For various constraints put on composition functions see Keenan & Stabler (1996), Szabó (2000), and Dever (2006).

family of sets:  $E = (E_g)_{g \in G}$  where  $G$  is the set of *sorts* which model grammatical categories, and for each  $g \in G$  there is a set  $E_g$  which is the set of expressions of category  $g$ . Then each syntactic rule  $\alpha \in F$  is defined only on certain sets  $E_g$ ; that is it yields a value only for sequences of expressions of  $E$  which belong to certain grammatical categories. For example, in modeling English there will be a set  $E_N$  (for nouns) and a set  $E_V$  (for verbs) and a syntactic rule which is a function defined on ordered pairs  $\langle e^N, e^V \rangle$  where  $e^N \in E_N$  and  $e^V \in E_V$  and with values in the set  $E_{VP}$  (for verb-phrases).<sup>3</sup>

A semantics is an interpretation function  $\mu$  that maps expressions on meanings: that is, a function defined over  $E$  and with values in a set  $M$  of meanings.<sup>4</sup> Then compositionality can be formulated in a precise form in the following way:

A semantics  $\mu$  is compositional iff for any syntactic rule  $\alpha \in F$  there is function  $f$  such that for any expressions  $e_1, \dots, e_n$  of  $E$ , if  $\alpha(e_1, \dots, e_n)$  is meaningful then:  $\mu(\alpha(e_1, \dots, e_n)) = f(\alpha, \mu(e_1), \dots, \mu(e_n))$ .

What kind of entities are the members of  $M$ ? What is meaning, in other words? It has been long noticed that when it comes to natural languages we should distinguish two types of meaning: what an expression means independently of any context of utterance, and what an expression means relative to a context of utterance. The first type of meaning roughly corresponds to what speakers know when they know the meaning of an expression but are completely oblivious to the details of the context in which the

---

<sup>3</sup> The rule given here for exemplification is, of course, very course-grained. A proper representation of English syntax must take into account other features like subject-verb agreement. Representing syntax as *many sorted algebras* is favored by Janssen (1983) and Hendriks (2001). But this is not the only way to model syntax. Hodges (2001) and Pagin & Westerståhl (2010) prefer to represent grammars as *partial algebras* where syntactic rules are represented as functions which are simply undefined on unwanted arguments. A note on notation: from now on, I'll ignore the superscripts for grammatical categories.

<sup>4</sup> If we believe that some expressions (e.g. “Green ideas sleep furiously”) are grammatical but meaningless we should take  $\mu$  to be a partial function: undefined for some members of  $E$ . If one believes that “Green ideas sleep furiously” is grammatical and meaningful, but false, and that there are no grammatical but meaningless expressions (as Magidor 2009 does), then one can let  $\mu$  be defined over all members of  $E$ .

expression was uttered. The second type of meaning roughly corresponds to what speakers know when they know what is said by an expression at a particular context of utterance.

Following Kaplan (1989), the established term for the latter type of meaning is *propositional content*. Propositional content can be thought of as the information that determines the extension of an expression at possible states of affairs, and thus can be modeled as an intension: a function from possible worlds to *extensions* (individuals for singular terms, sets for predicates, and truth-values for sentences). Given that expressions of natural languages are context-sensitive, they can be assigned propositional content only relative to a context of utterance. Notice that in the absence of a context the English sentence “I am hungry” fails to have express anything which is truth-evaluable, since in order to determine its truth-value we need to determine who uttered it when. Moreover, depending on who utters it, “I am hungry” expresses different things: if uttered by David Kaplan it expresses the content that David Kaplan is hungry, if uttered by Saul Kripke it expresses the content that Saul Kripke is hungry. Thus we say that “I am hungry” expresses different contents at different contexts of utterance. Then propositional content is a property of expressions at contexts, and a semantics that assigns propositional content will assign it not to expressions themselves but to expressions at contexts. Let  $M_I$  be the set of intensions, then for any meaningful  $e$  of  $E$  and context  $C$ ,  $\mu(e, C) \in M_I$  is such that  $\mu(e, C): W \rightarrow \text{Ext}$  (where  $W$  is the set of possible worlds, and  $\text{Ext}$  is the set of possible extensions). Each  $w \in W$  is an extension determining circumstance, in the sense that the extension of an expression is always given relative to a  $w$ .<sup>5</sup> Using the standard notation  $[[e]]^w$  for *the extension of  $e$  at  $w$* , we write the content of the expression as  $\lambda w. [[e]]^w$ . Notice that

---

<sup>5</sup> There is a long lasting debate in formal semantics on whether we can model natural languages with sparse circumstances of evaluation, consisting only of possible worlds, or whether we need richer ones. For example, Kaplan (1989) argues that we should take circumstances of evaluations to consist of world-time pairs. Others argue that circumstances of evaluation are even richer than that (see Kölbel 2008 for an overview). If we take circumstances of evaluation to be world-time pairs we should define intensions in the following way. Let  $W$  be the set of possible worlds,  $T$  the set of time moments, and  $W \times T$  their Cartesian product, then each  $\mu(e, C) \in M_I$  is such that  $\mu(e, C): W \times T \rightarrow \text{Ext}$ .

the propositional content of some natural language expressions (e.g. descriptions like “the president of France”) can be best modeled as non-constant functions from worlds to extensions, while the propositional content of other expressions (e.g. proper names like “David Kaplan”) can, arguably, be treated as a constant function from worlds to extensions.

Also following Kaplan (1989) the established term for the meaning that expressions have independent of context is *character*. The character of an expression is the convention associated with that expression and something like a rule of use: it tells what an expression can say when used in any arbitrary context. Characters are a property of expressions themselves, and can be modeled as functions from possible contexts of utterance to propositional contents; that is, as *hyperintensions*. Let  $M_H$  be the set of hyperintensions, then, for any meaningful expression  $e$  of  $E$ ,  $\mu^*(e) \in M_H$  is such that  $\mu^*(e) : K \rightarrow M_I$ , where  $K$  is the set of possible contexts. Contexts as formal objects of the theory have the job to represent the concrete situations in which language use takes place. Given that any use of an expression is performed by someone, at a place and time and in a possible world, then each  $C \in K$  is an ordered tuple of parameters consisting of an agent  $a$ , a world  $w$ , a time  $t$  and a location  $l$ :  $C = \langle a, w, t, l \rangle$ .<sup>6</sup> Notice that the character of some natural language expressions (e.g. indexicals: “I”, “here”, “now”, etc., descriptions: “the tallest man in town”, etc.) can be best modeled as a non-constant function from contexts to propositional content, while the characters of others (e.g. proper names, numerals) are best treated as constant functions.

If we distinguish between the meaning that an expression has independent of context of use, and the meaning that it has relative to a context of use, and we decide to model the first as a function from contexts to functions from possible worlds to extensions, then the extension of an expression is *double relative*. The extension of an expression is a function of the

---

<sup>6</sup> Some theorists (e.g. Stalnaker and his followers) prefer to represent contexts in a finer-grained manner, as the common ground between the participants in a conversational setting (i.e. the body of information commonly available to conversational partners) which can be modeled as the set of possible worlds compatible with the presumed common knowledge of the participants. These two notions of context can be complementary: a context as a common ground is determined by a context as a tuple of parameters: it is a fact about an *agent* at a *world*, *time* and *place* that she is presupposing certain propositions and that certain propositions are common ground in the conversation she is taking part. See Stalnaker (2014, 24-26) for elaborations.

context of utterance and of the circumstance of evaluation, and the two play different roles in determining the extension.<sup>7</sup> The character of an expression  $e$  determines, relative to any context  $C$ , the propositional content of  $e$  at  $C$ , which in its turn determines its extension, relative to any world  $w$ . Then I'll write  $[[e]]^{C,w}$  for the extension of  $e$  at  $C$  and  $w$ , and write the character of  $e$  as:  $\lambda c.\lambda w[[e]]^{C,w}$ .

Then, a theorist has two options when it comes to giving an account of natural languages. She can choose a semantics that assigns characters: a function defined over  $E$  and with values in the set of characters  $M_H$ . Alternatively, she can choose a semantics that assigns propositional content to expressions in context: a function defined over the Cartesian product of  $E$  and the set of contexts  $K$  with values in the set of propositional contents  $M_I$ . The first option is:

$\mu^*: E \rightarrow M_H$ , where each member of  $M_H$  is a function  $\mu^*(e): K \rightarrow M_I$

$\mu: E \times K \rightarrow M_I$ , where each member of  $M_I$  is a function  $\mu(e, C): W \rightarrow$   
Extension

Observe that  $\mu^*$  is the curried version of  $\mu$ .<sup>8</sup>

---

<sup>7</sup> To better see the need for double evaluation, consider the sentence: "Once, everyone now alive hadn't been born yet." Observe that in order to determine its truth-value we need to determine who is alive at the context of utterance; that is we need to look at the time of the context of utterance and determine the reference of the restrictor "alive now". But this is not enough. In order to determine its truth-value we also need to look in the past: to see whether it is true that there was a time in the past at which those alive at the time of the context of utterance were not born yet. In other words, "once" takes the propositional content expressed by the sentence embedded under it, and evaluates it to all time moments that precede the time of the context, such that it yields true if there is a time previous to the time of the context at which the propositional content is true. Notice that the distinct evaluations at time-moments play different roles: one is to fix the reference of "alive now", (reference which cannot be shifted by "once") and the other to evaluate for truth-value the propositional content expressed at the context of utterance by the sentence embedded under "once".

<sup>8</sup> Currying is a standard procedure by which the addicity of a function can be reduced. Given a function  $f$  of type  $f: (X \times Y) \rightarrow Z$ , currying it provides the function  $f^*: X \rightarrow (Y \rightarrow Z)$ . See Curry, Feys & Craig (1958).

## 2. Compositionality for Content and for Character

Depending on whether we define compositionality for propositional content or for character, formulations of different strength are obtained. This is how semanticists traditionally formulated compositionality for character and for propositional content respectively.<sup>9</sup>

*i. Compositionality of character:* The character of a complex expression is a function of the character of its constituents and of its syntactic structure. More precisely, a semantics  $\mu^*$  is *character compositional* iff for any syntactic rule  $\alpha$  there is a function  $f$  such that for any  $e_1, \dots, e_n$  if  $\alpha(e_1, \dots, e_n)$  is meaningful then:

$$\mu^*(\alpha(e_1, \dots, e_n)) = f(\alpha, \mu^*(e_1), \dots, \mu^*(e_n)).$$

A semantics fails to be character compositional if for some expressions  $e_i, e_j, e_k \in E$  and syntactic rule  $\alpha$ ,

$$\mu^*(e_j) = \mu^*(e_k) \text{ and } \mu^*(\alpha(e_i, e_j)) \neq \mu^*(\alpha(e_i, e_k)).$$

A semantics fails to be character compositional if substitution of character-equivalent expressions within a larger one does not preserve the character of the larger expression. In plain words, a semantics fails to be character compositional if substitution of synonyms is not character preserving in that semantics. To give an illustration, a semantics of English (or of a fragment of English) fails to be character compositional if it assigns the same characters to “attorneys” and to “lawyers”, but assigns different characters to “Attorneys are rich” and “Lawyers are rich”.

*ii. Compositionality for content:* given that content is assigned to expression-context pairs, in order to formulate compositionality for content we need to take into account the role that context plays in the determination of the content of complexes. Standardly, compositionality for content is

---

<sup>9</sup> See Kaplan (1989, 507) where both varieties are given informally. For their formal rendering see Pagin & Westerståhl (2010, 259-260) and Dever (2006, 634).

formulated in the following way. The content of a complex expression relative to a context is a function of the content of its constituents at that context and of its syntactic structure. More precisely, a semantics  $\mu$  is *content compositional* iff for every syntactic rule  $\alpha$  there is a function  $f$  such that for any expressions  $e_1, \dots, e_n \in E$  and for any context  $C$ , if  $\alpha(e_1, \dots, e_n)$  is meaningful at  $C$  then:

$$\mu(\alpha(e_1, \dots, e_n), C) = f(\alpha, \mu(e_1, C), \dots, \mu(e_n, C))$$

A semantics fails to be content compositional if the substitution of co-intensional expressions within a larger expression does not guarantee to preserve the intension (content) of the larger expression. More precisely, a semantics fails to be content compositional if there is a syntactic rule  $\alpha$ , some expressions  $e_i, e_j, e_k$ , and context of utterance  $C$ , such that

$$\mu(e_j, C) = \mu(e_k, C) \text{ and } \mu(\alpha(e_i, e_j), C) \neq \mu(\alpha(e_i, e_k), C)$$

For example, a semantics that assigns to “I” relative to a context  $C$ , which has David as its agent parameter, the same content that it assigns to “David” at context  $C$ , but assigns different contents to “I am hungry” and “David is hungry” at  $C$  fails to be content compositional. If the failure condition obtains then the content of  $\alpha(e_i, e_j)$  at  $C$  and the content of  $\alpha(e_i, e_k)$  at  $C$  are not a function of the content of their constituents at  $C$ , for a function should return the same value given the same argument.<sup>10</sup>

---

<sup>10</sup> In order to avoid a potential retort that would side-track the discussion, let me point out that there is another way in which compositionality for content can fail: as a result of *context-shift failure*. That is, a semantics fails to be content compositional if a complex expression varies its content across contexts of utterance although its constituents have unvarying contents across the very same contexts of utterance. Formally, if there are some expressions  $e_i, e_j$ , syntactic rule  $\alpha$  and contexts  $C_1, C_2$  such that

$$\mu(e_i, C_1) = \mu(e_i, C_2) \text{ and } \mu(e_j, C_2) = \mu(e_j, C_1) \text{ and } \mu(\alpha(e_i, e_j), C_1) \neq \mu(\alpha(e_i, e_j), C_2)$$

obtain.

Nevertheless, the failure to preserve content under the substitution of co-intensional parts is more calamitous than context-shift failure, in the sense that any semantics that fails to preserve intension under the substitution of co-intensional parts also exhibits context-shift failure, but not the other way around. For proofs, see Pagin (2005, Appendix 1) and Westerstahl (2012). Relatedly, some authors (e.g., Pagin 2005; Lasersohn

Given the relation between  $\mu^*$  and  $\mu$ , if a semantics  $\mu$  satisfies or fails to satisfy content compositionality so does its curried version  $\mu^*$ , and vice-versa: if a semantics  $\mu^*$  satisfies or fails to satisfy character compositionality so does its un-curried version  $\mu$ .

### 3. Some context-shifting operators

The double-index framework introduced above allows us to treat some expressions as hyper-intensional operators, in the same manner in which it allows us to treat certain expressions as intensional operators. As it is well known, intensional operators take the intension of an expression and evaluate it at alternative circumstances of evaluation: in other words, they shift the circumstance of evaluation at which the intension they operate on is evaluated. Hyper-intensional operators work at the level character in the same fashion in which intensional operators work at the level of content. Whether natural languages contain expressions which are best treated as context-shifting operators, or whether such expressions can be added to a natural language is still a matter of debate.<sup>11</sup> Just as an intensional operator

---

2012) have proposed two readings for compositionality for content: a stronger and a weaker one, where (a) the strong one entails the weak one but not the other way around; (b) the strong version has as its negation condition, context-shift failure, and (c) the weak version has its negation condition the failure to preserve intension under substitution of co-intensional parts. If (a) is true so is its contrapositive: (d) failure to preserve intension under substitution of co-intensional parts entails context-shift failure.

<sup>11</sup> Kaplan (1989, 510-511) answers in negative to both questions, but Perry & Israel (1986) and Santorio (2012) argue that some epistemic modals of English are best treated as context-shifting operators. Since it is still a matter of debate whether there are English expressions that are best treated as context-shifting operators, I focus the discussion on the rather artificial construction “in some other context”. To get a flavor of how such operators might look like in English, Israel & Perry (1996, 311) suggest that we consider sentences containing epistemic modals evaluated under the veil of ignorance, as when the speaker of a context is ignorant about the values of the contextual parameters. Santorio (2012, 291) imagines precisely such a scenario: Rudolf and Carl, two amnesiacs, are kidnapped. Each of them knows that he is one of them but not which one. They are subjected to the following experiment, about which are informed: both are anesthetized and a coin is tossed. If the coin lands head, Rudolf will be killed and Carl released

shifts the circumstance at which an expression is evaluated, a hyper-intensional operator shifts the context at which the sentence is evaluated. In this sense, they are *context-shifting operators*. For illustration, consider the sentence

(1) I am hungry

and the sentence (2) obtained from (1) by prefixing it with the (rather artificial) context-shifting operator “In some other context”:

(2) In some other context I am hungry.

Where, by stipulation, (2) is true in the context of utterance just in case the agent of some other context is hungry at the time and world of that other context. In other words, (2) is true at the context of utterance if and only if there is some other context  $C^*$  such that (1) expresses a content that is true at circumstance determined by  $C^*$ . In other words, the operator “In some other context” operates on the character of (1): it takes the character of (1), evaluates it at alternative contexts, and yields true if and only if there is an alternative context  $C^*$  such that the content of (1) at  $C^*$  is true at the circumstances of  $C^*$ . We can introduce this operator (abbreviated as *Op*) in the following way:

$Op(S)$  is true at  $C$  iff there is a context  $C^*$  such that  $S$  is true at  $C^*$ .

In the course of evaluation of  $Op(S)$  at context  $C$  we evaluate  $S$  at context  $C^*$ : what the operator *Op* does is to look across contexts and yield true if there is a context  $C^*$  such that  $S$  is true at  $C^*$ . Before I move on, let me point one more thing. Notice that given that the character of some natural language expressions (e.g. proper names) is modeled as a constant function

---

on Harvard’s campus, if it lands tail, Carl will be killed and Rudolf released on Stanford’s Campus. Later on, one of them wakes up and utters: “I might be in Stanford” and “But I might well be in Harvard”. Intuitively both sentences are true. Then “I” refers to two different individuals, irrespective of the actual context of utterance. Schlenker (2002) argues that certain propositional attitude verbs in Amharic and English are best treated as context-shifting operators, but see Maier (2016) for a defense of Kaplanian ban on such operators.

from contexts to propositional content, if we embed such an expression under a context-shifting operator their content doesn't get shifted. Context-shifting operators shift only the content of expressions whose character is a non-constant function from context to propositional content. Compare with intensional operators: intensional operators cannot shift the extension of expressions whose propositional content is treated as a constant function (e.g. proper names, among others), but only of those whose propositional content is treated as a non-constant function (e.g. descriptions, among others).

#### 4. Compositionality and context-shifters

Context-shifting operators are not compatible with compositionality for content, but they are compatible with compositionality for character. Remember, a semantics  $\mu$  fails to be content compositional if the following obtains: for some expressions  $e_i, e_j, e_n$ , syntactic rule  $\alpha$  and context of utterance  $C$ :

$$\mu(e_j, C) = \mu(e_n, C) \text{ and } \mu(\alpha(e_i, e_j), C) \neq \mu(\alpha(e_i, e_n), C)$$

That is, if relative to a context  $C$  two expressions  $e_j$  and  $e_n$  are assigned the same content  $\mu(e_j, c) = \mu(e_n, c)$ , but two complex expressions  $\alpha(e_i, e_j)$  and  $\alpha(e_i, e_n)$  are assigned different contents relative to  $C$ , then such a semantics is not content compositional.

Given that  $\mu^*$  is the curried version of  $\mu$ , we can write the failure condition in the following way: for some expressions  $e_i, e_j, e_n$ , syntactic rule  $\alpha$  and context of utterance  $C$ :

$$\mu^*(e_j)(C) = \mu^*(e_n)(C) \text{ and } \mu^*(\alpha(e_i, e_j))(C) \neq \mu^*(\alpha(e_i, e_n))(C)$$

It can be easily shown that context-shifting operators are incompatible with content-compositionality. If  $\text{Op}(S)$  is a sentence that contains a context-shifting operator, the content of  $\text{Op}(S)$  at a context of utterance is not a function of the content of its constituents at that context of utterance. Rather, the content of  $\text{Op}(S)$  at a context  $C$  is a function of the content of its constituent  $S$  at a context  $C^*$ , where  $C \neq C^*$ . In more detail, if two co-

intensional expressions are embedded under context-shifting operators their substitution does not guarantee, anymore, preservation of neither content nor truth. Consider a language  $L$  (which is a fragment of English) for which a content compositional semantics  $\mu$  can be given. Assume that that for the two sentences of  $L$

( $\Phi$ ) I am hungry

and

( $\Psi$ ) Kaplan is hungry

and context  $C_1$  (where  $C_1$  is such that Kaplan is its agent)  $\Phi$  is true at  $C_1$  iff the agent of  $C_1$  is hungry at the time of  $C_1$  and  $\Psi$  is true at the same context iff Kaplan is hungry at the time of the context. Given that Kaplan is the agent of  $C_1$ ,  $\Psi$  and  $\Phi$  express the same content at  $C_1$ : a function from possible worlds  $w$  to truth-values which yields true if Kaplan is hungry at the time and world of  $C_1$  and yields false otherwise. That is:  $\mu(\Phi, C_1) = \mu(\Psi, C_1)$ . Since (by assumption)  $L$  has a semantics  $\mu$  which is content compositional the contents of  $\Phi$  and  $\Psi$  relative to  $C_1$  are a function of the contents of their respective constituents at  $C_1$  and of their respective syntactic structure.

Now consider an extension  $M$  of the initial language  $L$ .  $M$  contains all expressions of  $L$  plus the sentential operator  $\text{At}C_2$  and sentences constructed with the help of this operator. Thus, for any sentence  $S$  if  $S$  is a sentence of  $L$  then  $S$  is also a sentence of  $M$ . And for any sentence  $S$  of  $L$  there is a sentence of the form  $\text{At}C_2 S$  in  $M$ . We give a semantics  $\mu_M$  for  $M$  that assigns propositional content and preserves, or inherits the assignments of semantics  $\mu$  for  $L$ . For any sentence  $S$  of  $L$  the semantic function  $\mu_M$  of  $M$  agrees with the semantic function  $\mu$  of  $L$ :  $\mu(S) = \mu_M(S)$ . In other words, the semantics  $\mu_M$  of  $M$  is the same as semantics  $\mu$  of  $L$  for all sentences that do not contain the operator  $\text{At}C_2$ .

The operator  $\text{At}C_2$  is defined as follows. For any context  $C$ , the sentence  $(\text{At}C_2, S)$  is true relative to  $C$  iff  $S$  is true at  $C_2$ . In other words,  $\text{At}C_2$  is a function that takes the character of  $S$  and evaluates it at context  $C_2$  such that the complex sentence  $(\text{At}C_2, S)$  is true at  $C$  iff the evaluation of the character of  $S$  at  $C_2$  yields true. That is,  $\mu^*_M(\text{At}C_2, S)(C) = 1$  iff  $\mu^*_M(S)(C_2)$

= 1. Notice that by un-currying  $\mu^*$  we can define the operator for a semantics that assigns propositional content to expression-context pairs:  $\text{At}C_2$  is a function such that for any sentence  $S$  and any context  $C$ , it takes the content expressed by  $S$  at  $C$  (i.e.  $\mu_M(S, C)$ ) and yields the content expressed by  $S$  at  $C_2$  (i.e.  $\mu_M(S, C_2)$ ). In other words, for any context  $C$ ,  $\mu_M(\text{At}C_2, S), C) = \mu_M(S, C_2)$ .

It is easy to show that semantics  $\mu_M$  of the extended language  $M$  is not content compositional. Take the two sentences  $(\Phi)$  and  $(\Psi)$  and build two sentences with the help of the sentential operator  $\text{At}C_2$ . We get the following sentences in the extended language:

( $\Upsilon$ )  $\text{At}C_2$  I am hungry

( $\Omega$ )  $\text{At}C_2$  Kaplan is hungry

Suppose that  $C_1$  and  $C_2$  differ at most with respect to the agent of the context: Kaplan is the agent of  $C_1$  but not the agent of  $C_2$ . Then, ( $\Upsilon$ ) is true at  $C_1$  iff the speaker of  $C_2$  is hungry at the world of  $C_2$ . But notice that ( $\Omega$ ) is true at  $C_1$  iff Kaplan is hungry at the world of  $C_1$ . The content expressed by ( $\Upsilon$ ) at  $C_1$  is different than the content expressed by ( $\Omega$ ) at  $C_1$ :

$$\mu_M(\text{At}C_2, \Phi), C_1) \neq \mu_M(\text{At}C_2, \Psi), C_1)$$

Given that  $M$  is an extension of  $L$ , and its semantics  $\mu_M$  preserves the assignments of  $\mu$  for expressions of  $L$ , since  $\mu(\Phi, C_1) = \mu(\Psi, C_1)$  then  $\mu_M(\Phi, C_1) = \mu_M(\Psi, C_1)$ . Then,  $\mu_M$  is not content compositional because for some expressions  $\Phi$  and  $\Psi$ , syntactic structure and context  $C_1$ ,  $\mu_M(\Phi, C_1) = \mu_M(\Psi, C_1)$  and  $\mu_M(\text{At}C_2, \Phi), C_1) \neq \mu_M(\text{At}C_2, \Psi), C_1)$ . ■

The extended language satisfies, nevertheless, character compositionality. Remember, a semantics  $\mu^*$  fails to be character compositional if the following obtains: for some expressions,  $e_i, e_j, e_n$ , and syntactic rule  $\alpha$ :

$$\mu^*(e_j) = \mu^*(e_n) \text{ and } \mu^*(\alpha(e_i, e_j)) \neq \mu^*(\alpha(e_i, e_n))$$

By currying  $\mu_M$  we obtain a function  $\mu^*_M$  and it can be showed that  $\mu^*_M$  is character compositional. All we need to show is that the following holds for  $\mu^*_M$ : there are no expressions  $e_i, e_j, e_n$ , and syntactic rule  $\alpha$  such that  $\mu^*_M(e_j) = \mu^*_M(e_n)$  and  $\mu^*_M(\alpha(e_i, e_j)) \neq \mu^*_M(\alpha(e_i, e_n))$ . Given the definition

of  $\text{AtC}_2$  operator the following holds true of  $(\Upsilon)$ : for any context  $C$ ,  $\mu_M(\text{AtC}_2, \Phi)C = \mu_M(\Phi, C_2)$ . For the  $\mu_M^*$  semantics we write this in the following way:

$$(1) \quad \text{For any } C, \mu_M^*(\text{AtC}_2, \Phi)(C) = \mu_M^*(\Phi)(C_2)$$

and the following holds true of  $(\Omega)$

$$(2) \quad \text{For any } C, \mu_M^*(\text{AtC}_2, \Psi)(C) = \mu_M^*(\Psi)(C_2)$$

We can show *by reductio* that  $\mu_M^*$  character compositional. Assume that  $\mu_M^*$  for  $M$  is not character compositional. Then for some expressions  $\Phi$  and  $\Psi$  the following holds (which is the failure condition for character compositionality):

$$(3) \quad \mu_M^*(\text{AtC}_2, \Phi) \neq \mu_M^*(\text{AtC}_2, \Psi)$$

$$(4) \quad \mu_M^*(\Phi) = \mu_M^*(\Psi)$$

From (4) we get (by definition of  $\mu^*$ ):

$$(5) \quad \text{For any } C, \mu_M^*(\Phi)(C) = \mu_M^*(\Psi)(C)$$

From (5) we get (by universal instantiation):

$$(6) \quad \mu_M^*(\Phi)(C_2) = \mu_M^*(\Psi)(C_2)$$

From (1), (2), and (6) we get:

$$(7) \quad \text{For any } C, \mu_M^*(\text{AtC}_2, \Phi)(C) = \mu_M^*(\text{AtC}_2, \Psi)(C)$$

From (3) and the definition of  $\mu_M^*$  we get:

$$(8) \quad \text{There is a } C, \text{ such that } \mu_M^*(\text{AtC}_2, \Phi)(C) \neq \mu_M^*(\text{AtC}_2, \Psi)(C)$$

But (8) contradicts (7), therefore  $\mu_M^*$  is character compositional – i.e.  $\neg((3) \ \& \ (4))$ . ■

In brief, I showed that there are semantics that satisfy character compositionality but fail to satisfy content compositionality, namely those of languages that contain context-shifting operators. Interestingly, if we can show that any semantics which satisfies content compositionality also satisfies character compositionality, then together these results show that content compositionality is a proper generalization of character compositionality.<sup>12</sup>

I'll show, now, that if a semantics is content compositional then it is also character compositional. This is a *proof by contraposition*. It is easy to prove that if a semantics fails to be character compositional then it also fails to be content compositional which *is the contrapositive* of the claim that if a semantics satisfies content compositionality it also satisfies character compositionality.

A semantics fails to be character compositional if for some expressions  $e_i$ ,  $e_j$ , and  $e_n$  and syntactic rule  $\alpha$  the following assignments hold:

- (a)  $\mu^*(e_j) = \mu^*(e_n)$  and
- (b)  $\mu^*(\alpha(e_i, e_j)) \neq \mu^*(\alpha(e_i, e_n))$ .

Notice that by definition of  $\mu^*$

- (a) entails (a'):  $\forall C, \mu^*(e_j)(C) = \mu^*(e_n)(C)$  and
- (b) entails (b'):  $\exists C^*$  such that  $\mu^*(\alpha(e_i, e_j))(C^*) \neq \mu^*(\alpha(e_i, e_n))(C^*)$ .

The conjunction of (a') and (b') entails failure of content compositionality: there are some expressions  $e_i$ ,  $e_j$ ,  $e_n$ , syntactic rule  $\alpha$  and a context  $C^*$  such that  $\mu^*(e_j)(C^*) = \mu^*(e_n)(C^*)$  and  $\mu^*(\alpha(e_i, e_j))(C^*) \neq \mu^*(\alpha(e_i, e_n))(C^*)$ . ■

In conclusion, if a semantics satisfies content compositionality then it also satisfies character compositionality.

---

<sup>12</sup> This is an alternative proof to the one given in Westerståhl (2012). His proof, though, is significantly different than the one given here, and I believe that the proof given here is much simpler than the one he gives.

## 5. Conclusion

Whether English, or other natural languages, contain expressions which are best treated as context-shifting operators, or whether such operators can be added to natural languages is an open empirical question. Whatever the answer to the empirical question, the results of the paper show that one cannot ban context-shifting operators on the basis that they are incompatible with compositionality. Such operators are incompatible with content-compositionality but are compatible with character compositionality. If a theorist has good arguments to believe that natural languages contain context-shifting operators, but she desires to retain compositionality she must settle on character-compositionality. Only if someone has good independent arguments to believe that propositional content must be compositional and that compositionality is non-negotiable, can she give a principled argument against context-shifting operators. That is, only if one believes that content compositionality explains, but character compositionality fails to explain certain features of natural languages that we expect semantic theories to model is one entitled to ban context-shifting operators on the basis that they ruin the content-compositional machinery. In fact, Rabern & Ball (2017) convincingly argue that Kaplan's ban on context-shifting operators rests on his further theoretical commitments that tied him to the idea that propositional content must be compositional. More precisely, Kaplan is wedded to the idea that one and the same entity must play two roles: (a) be the object of natural language operators, such that these operators can be defined compositionally, and (b) the content of speech acts and the object of propositional attitudes. Only propositional contents, and not characters, can be the content of speech acts and the object of propositional attitudes.

### Acknowledgments

I would like to thank Peter Pagin and Dag Westerståhl for helpful discussions on the subject, and Max Kölbel, Zoltán Gendler Szabó and Dan Zeman for comments on a previous version of this work. I would like to thank anonymous reviewers of this journal for helpful comments on the paper. I acknowledge the financial support received from the Spanish Ministry of Economy and Competitiveness, EXCELENCIA programme, project no. FFI2016-80636-P: Foundations and Methods of Natural Language Semantics.

## References

- CURRY, H., FEYS, R. & CRAIG, W. (1958): *Combinatory Logic: Volume I*. 2<sup>nd</sup> ed. Amsterdam: North-Holland Publishing Company.
- DEVER, J. (2006): Compositionality. In: Lepore, E. & Smith, B. (eds.): *The Oxford Handbook of Philosophy of Language*, Oxford, Oxford University Press, 633-666.
- ISRAEL, D. & PERRY, J. (1996): Where Monsters Dwell. In: Seligman, J. & Westerståhl, D. (eds.): *Logic, Language and Computation*. Stanford: CSLI Publications, 303-316.
- JANSSEN, T. (1983): *Foundations and Applications of Montague Grammar*. Amsterdam: Mathematisch Centrum.
- HENDRIKS, H. (2001): Compositionality and Model-Theoretic Interpretation. *Journal of Logic, Language and Information* 10, No. 1, 29-48.
- HODGES, W. (2001): Formal Features of Compositionality. *Journal of Logic, Language and Information* 10, No. 1, 7-28.
- KAPLAN, D. (1989): *Demonstratives*. In Almog, J., Perry, J. & Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford University Press, 481-563.
- KEENAN, E. & STABLER, E. (1996): Abstract Syntax. In: Di Sciullo, A. (ed.): *Configurations*. Somerville: Cascadilla Press.
- KÖLBEL, M. (2008): The Evidence for Relativism. *Synthese* 166, No. 2, 375-395.
- LASERSOHN, P. (2012): Contextualism and Compositionality. *Linguistics and Philosophy* 35, No.2, 171-189.
- MAGIDOR, O. (2009): Category Mistakes are Meaningful. *Linguistics and Philosophy* 32, No. 6, 553-581.
- MAIER, E. (2016): A Plea against Monsters. *Grazer Philosophische Studien* 93, 363-395.
- PAGIN, P. (2005): Compositionality and Context. In: Preyer, G. & Peter, G. (eds.): *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Oxford University Press, 303-348.
- PAGIN, P. & WESTERSTÅHL, D. (2010): Compositionality I: Definitions and Variants. *Philosophy Compass* 5, No. 3, 250-264.
- PARTEE, B. (2004): Compositionality. In: *Compositionality in Formal Semantics: Selected Papers of Barbara Partee*. Oxford: Blackwell, 153-181.
- RABERN, B. (2013): Monsters in Kaplan's Logic of Demonstratives. *Philosophical Studies* 164, No. 2, 393-404.
- RABERN, B. & BALL, D. (2017): Monsters and the Theoretical Role of Context. *Philosophy and Phenomenological Research*. DOI: 10.1111/phpr.12449.
- SANTORIO, P. (2012): Modals Are Monsters: On Indexical Shift in English. *Proceedings of SALT 20*, 289-308.
- SCHLENKER, P. (2002): A Plea for Monsters. *Linguistics and Philosophy* 26, No. 1, 29-120.

- STALNAKER, R. (2014): *Context*. Oxford: Oxford University Press.
- SZABÓ, Z. (2000): Compositionality as Supervenience. *Linguistics and Philosophy* 23, No. 5, 475-505.
- SZABÓ, Z. (2013): Compositionality. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition). Available at: <https://plato.stanford.edu/archives/fall2013/entries/compositionality/>.
- TRAVIS, C. (1996): Meaning's Role in Truth. *Mind* 105, No. 419, 451-466.
- VON FINTEL, K. & MATTHEWSON, L. (2008): Universals in Semantics. *The Linguistic Review* 25, 139-201.
- WESTERSTÅHL, D. (2012): Compositionality in Kaplan Style Semantics. In: Werning, M., Hinzen, W. & Machery, E. (eds.): *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press, 192-219.

## Analysis of Time References in Natural Language by Means of Transparent Intensional Logic

MARIE DUŽÍ – JAKUB MACEK<sup>1</sup>

**ABSTRACT:** In this paper, we deal with sentences containing time references like ‘five years ago’, ‘three years older’, ‘in five seconds’. It turns out that such sentences are pragmatically incomplete, because there is an elliptic reference to a calendar that makes it possible to determine the length of the time interval associated with a time duration like a year, month, day, or to compute the time interval denoted by terms like ‘February 29, 2016’. Since Transparent Intensional Logic (TIL) takes into account two modal parameters, namely possible worlds of type  $\omega$  and times of type  $\tau$ , and this system is particularly apt for the analysis of natural language expressions, our background theory is TIL. Within this system, we define time intervals, calendar time durations, and last but not least a method for adding and multiplying time durations in a way that takes into account the leap days and leap seconds. As sample applications, we analyse two sentences, to wit, “A year has 365 days” and “Adam is 5 years older than Bill”.

**KEYWORDS:** Calendar – Gregorian calendar – Julian calendar – time duration – time interval – TIL – time point – time span – Transparent Intensional Logic – typed system – year.

---

<sup>1</sup> Received: 9 May 2017 / Accepted: 27 June 2017

✉ Marie Duží

VSB-Technical University of Ostrava, Department of Computer Science FEI  
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic  
e-mail: marie.duzi@vsb.cz

✉ Jakub Macek

VSB-Technical University of Ostrava, Department of Computer Science FEI  
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic  
e-mail: jakub.macek.0@gmail.com

## 0. Introduction

Terms specifying time-referring objects like ‘five years ago’, ‘next month’, ‘for three days’ and sentences containing such terms are part and parcel of our everyday vernacular. The goal of this paper is to present a logical analysis of natural-language terms specifying *time durations* (‘year’, ‘month’, ‘day’, etc.) and their mutual relations in different contexts. As an example, we are going to analyse two sample sentences containing such terms:

“A year has 365 days.”

“Adam is five years older than Bill.”

We believe that these two sample sentences characterise well the issues connected with the analysis of such sentences containing time duration and time references. The first sentence might appear as an analytic one; yet it is not so, as we are going to show below. The second sentence illustrates an ordinary relation-in-intension between two individuals; yet the term ‘five years’ is vague, as we are going to show as well.

Yet, to the best of our knowledge, the analysis of such terms and sentences has been rather neglected by logicians as well as philosophers of language. There are several temporal logics that deal with sentences in the present, past and future tenses. These formal systems are mostly viewed as a special case of modal logic interpreted by means of Kripkean possible-world semantics. The term temporal logic is broadly used to cover all approaches to the representation of the temporal dimension within a logical framework. More narrowly, it is also used to refer to a particular modal system of temporal propositional logic that Arthur Prior introduced in Prior (1957; 1962; and 1967) under the name ‘tense logic’. Despite the great applicability of particular variants of tense logic in the semantics of programming languages, the systems just mentioned suffer a drawback when applied to the semantics of natural language. The drawback is their inability to adequately analyse sentences indicating a point of reference referring to the interval when the sentence was or will be true. Such sentences come attached with a presupposition under which a sentence is true or false.

This issue has been properly analysed in TIL that is an expressive logic apt for the analysis of sentences with presuppositions, because in TIL we work with partial functions, in particular with propositions with truth-value gaps (see Tichý 1980; and also Duží 2010).

In computer science, rigorous analysis of terms specifying time-referring objects is crucial. For instance, Ohlbach (1998) presents *Calendar Logic*, a propositional temporal logic whose operators quantify over time intervals that are specified using the terms of common vernacular, such as ‘next week’ and ‘June 2000’. *Calendar Logic* uses two modal operators, ‘sometimes within  $T$ ’ and ‘always within  $T$ ’ where  $T$  may be one of the (finite) time intervals. This in effect allows *Calendar Logic* to retain the decidability of propositional logic, albeit at the expense of expressivity. A system for refinement of time intervals that captures the complexity of the Gregorian calendar is presented. For example, the following formula specifies the time interval denoted by the term ‘29<sup>th</sup> day of February 1998’ where the initial interval of 1998 is assigned to the variable  $x_{year}$  and further intervals are specified using the functions *February* and *day\_within\_month*:

$$[1998; year] : day\_within\_month(February(x_{year}); 29)$$

The decidability of the system is the primary focus, and individual time intervals are denoted by the formal language constructs that partition the timeline into a finite number of continuous intervals. For example, consider these two sentences:

“If the temperature was below 0°C on February 20<sup>th</sup>, 2000, it snowed on February 20<sup>th</sup>, 2000.”

“The temperature was below 0°C the entire February of 2000.”

There are three continuous intervals: February 1<sup>st</sup>, 2000 to February 19<sup>th</sup>, 2000 (I1), February 20<sup>th</sup>, 2000 (I2), and February 21<sup>st</sup>, 2000 to February 28<sup>th</sup>, 2000 (I3). The final (pure) propositional logic formula for the first sentence is “ $T_{I2} \rightarrow S_{I2}$ ” and the formula for the second sentence is “ $T_{I1} \wedge T_{I2} \wedge T_{I3}$ ”.

In Ohlbach & Gabbay (2004), the approach is extended to fuzzy time intervals, and the notion of time duration within this system is defined.

Hobbs & Pan (2004) proposes a similar system with the intention to represent time-based statements within OWL ontologies.<sup>2</sup>

The approach handles the property of leap days, but leap seconds are intentionally left out. To facilitate this, several predicates are defined, but the notion of duration is represented using predefined constants such as *\*Day\**. For example, the following formula states that if  $m$  is an interval that is the month of February of the interval  $y$  and the interval  $y$  is a leap year, then  $m$  has 29 days:<sup>3</sup>

$$\text{February}(m, y) \wedge \text{leapYear}(y) \rightarrow \text{Hath}(29, \text{*Day*}, m)$$

For comparison, we propose an approach that applies Tichý’s Transparent Intensional Logic (TIL) with procedural semantics based on ramified type hierarchy. In TIL we furnish the three different time objects (points, intervals, and durations) with types within a ramified hierarchy of types. In other words, we analyse the natural-language terms denoting them in a fine-grained way as any other terms of natural language. This applies, *inter alia*, to the objects of calendars (e.g. the *Gregorian calendar*) as well, and allows us to render the meaning of sentences like “Any year in the Gregorian calendar is longer than the same year in the Julian calendar.” For instance, the year 2017 in the Gregorian calendar is longer than the year 2017 in the Julian calendar.

The rest of this paper is organized as follows. In Section 1 we introduce a fragment of TIL that we need for the analysis of time references. Section 2 deals with two basic terms that are used in time modelling, namely ‘time point’ (referring to a single point in time) and ‘time interval’ (standing for, e.g., the year 2017). In TIL, time is modelled as a set of real numbers, and these two terms are defined accordingly; *time point* as a real number and *time interval* as an interval of real numbers. Section 3 presents the main *novel contribution*, which are the definitions of *time duration* and *calendar time duration*. The fact that various years have different lengths is taken into account as well as the different notions of a *year* according to the Gregorian and the Julian *calendar*. Here we also

---

<sup>2</sup> The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies.

<sup>3</sup> Citation is exactly as in the paper.

deal with the problem of leap days and leap seconds. Section 4 presents an analysis of the sentence “A year has 365 days”. Here we put forward several building blocks, notably the addition of *calendar time durations* (e.g. *a year and a day*), multiplication of *calendar time durations* (e.g. *356 days*) and a *modifier of calendar time duration*. In Section 5, we analyse the other sample sentence “Adam is five years older than Bill”. Two possible alternatives with slight technical differences are proposed, and the advantages and disadvantages of both are discussed. In the concluding Section 6, the proposed solutions are summarized and further research suggested. The latter includes, inter alia, an analysis of the sort(s) of calendars that are actually used by people in their everyday lives, such as the calendars that are implemented in cell phones and computers.

## 1. Fundamentals of TIL

As mentioned above, our background theory is TIL, namely the version presented in Duží et al. (2010) – see also Tichý (1998) and Tichý (2004). From a formal point of view, TIL is a partial, typed lambda calculus with a procedural semantics. This means that we explicate the meanings of expressions as abstract procedures encoded by the expressions. These procedures are rigorously defined as TIL *constructions*. All the entities of the stratified ontology of TIL receive a type. Thus, the core of TIL consists of the definition of the type hierarchy and the definition of constructions. For the sake of simplicity, we first define types of order 1 that include types of non-procedural objects, then four kinds of constructions, and finally the ramified hierarchy of types of order  $n$ .

### **Definition 1** (*types of order 1*)

Let  $B$  be a *base*, where a base is a collection of pair-wise disjoint, non-empty sets. Then:

- (i) Every member of  $B$  is an elementary *type of order 1 over  $B$* .
- (ii) Let  $\alpha, \beta_1, \dots, \beta_m$  ( $m > 0$ ) be types of order 1 over  $B$ . Then the collection  $(\alpha \beta_1 \dots \beta_m)$  of all  $m$ -ary partial mappings from  $\beta_1 \times \dots \times \beta_m$  into  $\alpha$  is a functional *type of order 1 over  $B$* .

- (iii) Nothing is a *type of order 1 over B* unless it so follows from (i) and (ii).  $\square$

For the purposes of natural-language analysis, we are currently assuming the following base of *atomic types*, which form part of the ontological commitments of TIL:

- o: the type of truth-values = {T, F}
- i: the type of individuals (the universe of discourse)
- $\tau$ : the type of real numbers (doubling as time points)
- $\omega$ : the type of logically possible worlds (the logical space)

As mentioned above, in TIL we have two mutually independent modal parameters, namely possible worlds and times. Thus, unlike Montague's IL logic, we can apply explicit intensionalisation and temporalisation, which we need for the analysis of empirical sentences containing time references.<sup>4</sup>

**Definition 2** (*construction*)

- (i) A *variable*  $x$  is the *construction* that constructs an object  $X$  of the respective type assigned to  $x$  as the range of  $x$  dependently on a valuation  $v$ ;  $x$  *v-constructs*  $X$ .
- (ii) Where  $X$  is an object whatsoever, *Trivialization* is the construction  ${}^0X$ .  ${}^0X$  constructs  $X$  without any change of  $X$ .
- (iii) Let  $X, Y_1, \dots, Y_m$  be constructions. Then *Composition*  $[X Y_1 \dots Y_m]$  is the following *construction*. If  $X$  *v-constructs* a function  $g$  of a type  $(\alpha\beta_1 \dots \beta_m)$ , and  $Y_1, \dots, Y_m$  *v-construct* entities  $B_1, \dots, B_m$  of types  $\beta_1, \dots, \beta_m$ , respectively, then the *Composition*  $[X Y_1 \dots Y_m]$  *v-constructs* the value (an entity, if any, of type  $\alpha$ ) of  $g$  on the tuple argument  $\langle B_1, \dots, B_m \rangle$ . Otherwise the *Composition*  $[X Y_1 \dots Y_m]$  does not *v-construct* anything and so is *v-improper*.
- (iv) The *Closure*  $[\lambda x_1 \dots x_m Y]$  is the following *construction*. Let  $x_1, x_2, \dots, x_m$  be pair-wise distinct variables *v-constructing* entities

---

<sup>4</sup> See Duží et.al (2010, § 2.4.3) for criticism of Montague's implicit intensionalisation.

of types  $\beta_1, \dots, \beta_m$  and  $Y$  a *construction* typed to  $v$ -construct an  $\alpha$ -entity. Then  $[\lambda x_1 \dots x_m Y]$  is the *construction  $\lambda$ -Closure*. It  $v$ -constructs the following function  $f$  of the type  $(\alpha\beta_1 \dots \beta_m)$ . Let  $v(B_1/x_1, \dots, B_m/x_m)$  be a valuation identical with  $v$  at least up to assigning objects  $B_1/\beta_1, \dots, B_m/\beta_m$  to variables  $x_1, \dots, x_m$ . If  $Y$  is  $v(B_1/x_1, \dots, B_m/x_m)$ -improper (see iii), then  $f$  is undefined at  $\langle B_1, \dots, B_m \rangle$ . Otherwise the value of  $f$  at  $\langle B_1, \dots, B_m \rangle$  is the  $\alpha$ -entity  $v(B_1/x_1, \dots, B_m/x_m)$ -constructed by  $Y$ .

- (v) Nothing is a *construction*, unless it so follows from (i) through (iv).  $\square$

*Remark.* Definition 2 leaves out constructions *Single* and *Double Execution*,  ${}^1X$  and  ${}^2X$ , which we do not need for the present study.

**Definition 3** (*ramified hierarchy of types*)

$T_1$  (*types of order 1*). See Def. 1.

$C_n$  (*constructions of order  $n$* )

- (i) Let  $x$  be a variable ranging over a type of order  $n$ . Then  $x$  is a *construction of order  $n$  over  $B$* .
- (ii) Let  $X$  be a member of a type of order  $n$ . Then  ${}^0X$  is a *construction of order  $n$  over  $B$* .
- (iii) Let  $X, X_1, \dots, X_m$  ( $m > 0$ ) be *constructions of order  $n$  over  $B$* . Then  $[X X_1 \dots X_m]$  is a *construction of order  $n$  over  $B$* .
- (iv) Let  $x_1, \dots, x_m, X$  ( $m > 0$ ) be *constructions of order  $n$  over  $B$* . Then  $[\lambda x_1 \dots x_m X]$  is a *construction of order  $n$  over  $B$* .
- (v) Nothing is a *construction of order  $n$  over  $B$*  unless it so follows from  $C_n$  (i)-(iv).

$T_{n+1}$  (*types of order  $n + 1$* ).

Let  $*_n$  be the collection of all constructions of order  $n$  over  $B$ . Then

- (i)  $*_n$  and every type of order  $n$  are *types of order  $n + 1$* .
- (ii) If  $m > 0$  and  $\alpha, \beta_1, \dots, \beta_m$  are types of order  $n + 1$  over  $B$ , then  $(\alpha \beta_1 \dots \beta_m)$  (see  $T_1$  ii)) is a *type of order  $n + 1$  over  $B$* .
- (iii) Nothing is a *type of order  $n + 1$  over  $B$*  unless it so follows from  $T_{n+1}$  (i) and (ii).  $\square$

*Remark.* As a notational convention, ‘ $a/\alpha$ ’ means that the object  $a$  is of type  $\alpha$ , while ‘ $C \rightarrow \alpha$ ’ means that the construction  $C$  is typed to  $\nu$ -construct objects of type  $\alpha$ . Where  $C$  is a construction, the frequently used Composition  $[[C w] t]$  will be abbreviated as  $C_{wt}$ .

*Empirical expressions* and sentences denote so-called *PWS-intensions*, which are functions of type  $((\alpha\tau)\omega)$ , abbreviated as ‘ $\alpha_{\tau\omega}$ ’, that is, mappings from possible worlds to *chronologies* of objects of type  $\alpha$ . Note that in TIL we have two independent modal parameters at our disposal, namely possible worlds of type  $\omega$  and times of type  $\tau$ , which is another reason we recommend TIL as a theory apt for analysis of empirical expressions with time references. Throughout this paper we use variables  $w$  and  $t$  ranging over  $\omega$  and  $\tau$ , respectively. Intensions, being functions of  $\alpha_{\tau\omega}$ , are  $\nu$ -constructed by Closures of the form  $\lambda w \lambda t C$ , where  $C \rightarrow \alpha$ .<sup>5</sup>

## 2. Time point and time interval

In TIL, time is modelled by the type  $\tau$ , the set of real numbers.<sup>6</sup> Therefore, any *time point* is modelled as a real number. Thus, the binary relation of equality between two *time points* is defined as the identity relation  $=/(\sigma\tau\tau)$ . The binary relation of precedence between two *time points* is defined as the less-than relation  $</(\sigma\tau\tau)$ . Furthermore, the binary relation  $\leq$  is defined as  $t_1 \leq t_2$  iff  $t_1 = t_2$  or  $t_1 < t_2$ . For the sake of simplicity, when applying these relations, we use infix notation (and without Trivialization) as is common in mathematics.

Now we are going to deal with time intervals. Unfortunately, the term ‘time interval’ is commonly used with two different meanings. It either denotes a *time duration*, such as *20 seconds* as in the sentence “The light changes colour every 20 seconds”, or a particular *interval* of time points, i.e. the set of real numbers/time points with the property that any number that is in between two numbers in the set is also included in the set. Time

---

<sup>5</sup> Jespersen talks about this logical form characteristic of empirical expressions as *explicit intensionalization and temporalization*; see Jespersen (2005).

<sup>6</sup> In practice, time can be modelled in a different way. For example, for the purposes of programming, discretization of time is necessary; in such cases, time can be modelled as a set of integer numbers.

duration will be dealt with in the next section. To avoid confusion, from now on we will terminologically distinguish ‘time duration’ and ‘time interval’. The latter notion is mathematically defined as follows.

#### Definition 4

Let  $t_1, t_2$  be time points such that  $t_1 < t_2$ . Then a *time interval between the time points*  $t_1$  and  $t_2$  is a bounded half-open interval  $[t_1, t_2)$  excluding the point  $t_2$ . Hence, it is a set of real numbers constructed by  $\lambda t [t_1 \leq t < t_2]$ .

The reason we define *time interval* as a half-closed mathematical interval excluding the last end-point is this. We assume that *time is linear and continuous*.<sup>7</sup> If we went for discrete time, then the interval could be closed. Yet on the assumption of a time continuum, the alternatives would cause severe problems. The first alternative, an open interval, for instance (the first moment of 2017, the last moment of 2017) obviously excludes the first and the last moments of 2017. The second alternative, a closed interval [first moment of 2017, last moment of 2017], presumes the existence of a last moment of 2017. If there were such a *time point*  $x$ , there would be infinitely many time points between  $x$  and the first moment of 2018. Therefore,  $x$  would not be the last moment of 2017. The third alternative would be the half-open interval excluding the first end-point. Yet, since it is natural to deal with time as flowing forward, we choose the half-open interval excluding the last end-point.<sup>8</sup> Next, we define the *length of a time interval*:

---

<sup>7</sup> We do not deal here with ‘branching time’ theories; see, for instance, Placek (2012). These theories have many useful applications in computer science in the research on parallel and concurrent processes; see Nain & Vardi (2007).

<sup>8</sup> Another reason for this approach is explained in detail in Hobbs & Pan (2004, 76). The authors suggest that “we get a cleaner treatment if, for example, all times of the form 12:xx a.m., including 12:00 a.m., are part of the same hour and day, and all times of the form 10:15:xx, including 10:15:00, are part of the same minute” and support this claim by practical examples.

### Definition 5

The *length*  $L$  of time interval  $[t_1, t_2)$  is a (non-negative) real number  $L = t_2 - t_1$ .

Depending on the calendar, there are time intervals that play a special role in our everyday lives, like the one denoted by ‘the day of September 11, 2001’. In general, many time intervals receive a name in our everyday vernaculars, like ‘the year 2017’, the ‘month January of 2017’, etc. For example, let  $t_1$  be the first moment of January 1<sup>st</sup>, 2017, and  $t_2$  the first moment of January 1<sup>st</sup>, 2018. Then the interval  $[t_1, t_2)$  is the *year 2017*. Yet, as mentioned above, the length of particular time intervals, such as the year 2017, depends on calendars.

## 3. Calendars and time duration

*Time durations*<sup>9</sup> are objects that are denoted by expressions such as ‘year’, ‘month’, ‘day’, ‘hour’, ‘minute’, ‘second’, ‘5 years’, ‘5 years ago’, ‘a year and a month’ and ‘15 hours 30 minutes’. These objects are dependent on a particular calendar for their duration. There have been many *calendars* in use around the world. Some of them, such as the *Gregorian calendar* and *Julian calendar*, have relatively minor differences, others, such as the fiscal calendar for accounting and budget purposes, may define different rules for various *time durations*, or the duration of a year according to a solar or a lunar calendar also differ.

The Cambridge Dictionary<sup>10</sup> defines *calendar* as “the system used to measure and arrange the days, weeks, months, and special events of the year according to a belief system or tradition” and the *Gregorian calendar* as a “system used in many parts of the world to divide the 365 days of the year into weeks and months, and to number the years”.<sup>11</sup>

---

<sup>9</sup> Sometimes the alternative term ‘time span’ is used.

<sup>10</sup> See <http://dictionary.cambridge.org/dictionary/english/calendar>.

<sup>11</sup> While in this paper we discuss ordinary, recently used calendars that operate with objects such as *months* and *minutes*, historically there have been different peculiar calendars like the one operating with the “sinking-bowl” of water for measuring intervals of time in India; see Plofker (2011). As far as we know, the Ancient Britons, probably

In what follows we take a *calendar method* as being an object of type  $*_n$ , i.e., a construction. This is a simplification, for sure, yet for our purposes this simplification is harmless. From the practical point of view, it is more important to analyse the structure of a calendar time durations, because the reasonable definitions of calendar methods computing these durations are rare.<sup>12</sup> A *calendar* (e.g. the *Gregorian calendar*) is then an empirical function of type  $((*_n\tau)\omega)$ , or ‘ $*_{\tau\omega}$ ’ for short, that yields a *calendar method* for a given world and time. We define a calendar in this way, because calendar methods are based on empirical observations (such as the solar cycle or the lunar cycle), and they can be adjusted from time to time.

There are time durations that differ in different calendars. Yet even within one and the same calendar these time durations are not of the same length. This is due to leap seconds and leap days. Thus, we define:

### Definition 6

A *time duration* is a function of type  $(\tau\tau)$ . A *time interval*  $[t_1, t_2)$  has a *time duration*  $d$  iff  $[d t_1] = t_2$ . A *calendar time duration* is a function from *calendar methods* to time durations. Hence a *calendar time duration* is a function of type  $((\tau\tau)*_n)$ .

For example, a *year* is a *calendar time duration* that for a given *calendar method*  $c$  associates any *time point*  $t_1$  with the *time point*  $t_2$  that comes one year after  $t_1$  (according to a given calendar method). Note the difference between the *length* of an *interval* and *time duration* of an *interval*. The *length* of a given *interval* is an exact real number, whereas its *time duration*, for instance a *year in the Gregorian calendar*, does not determine a definite number. It can be 365 or 366 *days in the Gregorian calendar*, the lengths of particular *days* can also differ due to leap seconds, etc. Only

---

under the influence of the Druids, used similar bowls for measuring intervals of time. The bowls had a small hole in the bottom, and in use it was placed on the surface of water, which slowly leaked into it until, after a certain interval of time, the bowl sank. The interval was the unit of time; in the case of the bowl found in County Antrim, Northern Ireland, it was approximately one hour.

<sup>12</sup> This topic would be a subject of further research that is out of the scope of the present paper.

when obtaining additional pieces of information, like the exact point of the beginning of the *interval* and a *calendar*, is one able to compute rigorously the actual *calendar time duration* of a given *interval*.

For some applications, it may be feasible to define *time duration* in a simpler way, for instance as a time difference in seconds. However, this is not acceptable when analysing natural language. As mentioned above, due to the existence of leap days and leap seconds, various *time intervals* that have the same *time duration*, for instance a *minute*, may have different lengths. This is particularly obvious of the *calendar time duration month*. Saying that this or that lasted a month one is not conveying much information. It can be 28, 29, 30 or 31 days.

Things are even more complicated with leap days, even within the *Gregorian calendar*. The most complicated problem of *calendar time durations* is the question what day follows exactly *one year after the 29<sup>th</sup> of February* of a leap year, for instance 2016.<sup>13</sup> Intuitively, one would say that it must be a regular day in the calendar, not a virtual one. For sure, because this question is important, for instance for legal purposes, to compute the age of criminal responsibility. In England children under 10 cannot be arrested or charged with a crime. If a child was born on the 29<sup>th</sup> of February 2000 and commits a criminal act on the 28<sup>th</sup> of February 2010, are they responsible? Unfortunately, there is no consensus on the solution of this problem.<sup>14</sup>

For the moment, let us assume that<sup>15 16</sup>

---

<sup>13</sup> Note that the same problem applies to leap seconds, for example the December 31<sup>st</sup>, 2005, 18:59:60 leap second.

<sup>14</sup> To illustrate, we have tested several programming platforms; the .NET framework class `DateTime` gives February 28<sup>th</sup>, 2017 as *one year after February 29<sup>th</sup>, 2016*, the same as the Java class `GregorianCalendar` (Java forces the programmers to choose explicitly the calendar they want to use). The PHP class `DateTime`, however, yields March 1<sup>st</sup>, 2017.

<sup>15</sup> The technical details of the addition of *time* are explicated in TIL in Section 5. However, in accordance with our intuition it should hold that the addition of *time duration*  $d$  to a *time interval*  $[t_1, t_2)$  yields a *time interval*  $[d(t_1), d(t_2))$ .

<sup>16</sup> In what follows, we assume the *Gregorian calendar* for the purpose of obtaining *time durations* from *calendar time durations*.

*February 29<sup>th</sup> 2016 + 1 year = February 28<sup>th</sup> 2017*

The following additions are less problematic:

*February 28<sup>th</sup> 2017 + 1 year = February 28<sup>th</sup> 2018*

*February 28<sup>th</sup> 2018 + 1 year = February 28<sup>th</sup> 2019*

*February 28<sup>th</sup> 2019 + 1 year = February 28<sup>th</sup> 2020*

The troubling part is immediately apparent, because one would also assume that

*February 29<sup>th</sup> 2016 + 4 years = February 29<sup>th</sup> 2020*

From this, it consequently follows that:

*((February 29<sup>th</sup> 2016 + 1 year) + 1 year) + 1 year) + 1 year  
 ≠ February 29<sup>th</sup> 2016 + 4 years*

Thus, when we define the operation of adding *calendar time durations*, the following holds:

*((1 year + 1 year) + 1 year) + 1 year ≠ 4 years*

In general, we cannot define the operation of multiplication of a *calendar time duration* as a series of additions of that *calendar time duration*. Moreover, the same problem arises with *negative calendar time durations*. We have

*February 28<sup>th</sup> 2017 + (-1 year) = February 28<sup>th</sup> 2016*

But

*(February 29<sup>th</sup> 2016 + 1 year) + (-1 year) ≠ February 29<sup>th</sup> 2016*

Intuitively, it should be clear what we mean by ‘+ 1 year’ or ‘-1 year’; yet the rigorous definition is needed. Thus we define.

### Definition 7

A *calendar time duration*  $d$  is *positive* according to a *calendar (method)*  $c$  iff for all *time points*  $t$  holds that  $t < [[d\ c]\ t]$ . A *calendar time duration*  $d$  is *negative* according to a *calendar (method)*  $c$  iff for all *time points*  $t$  holds that  $t > [[d\ c]\ t]$ . Finally, a *calendar time duration*  $d$  is *zero* according to a *calendar (method)*  $c$  iff for all *time points*  $t$  holds that  $t = [[d\ c]\ t]$ .

For example, *15 hours 30 minutes* is a *positive calendar time duration* (like in “The cricket match lasted 15 hours and 30 minutes”) and *5 years ago* is a *negative calendar time duration* (like in “I last saw him 5 years ago.”) according to the *Gregorian* as well *Julian calendar*.

A *zero-calendar time duration* is used in our common vernacular to communicate that one event followed another without delay, that is, *immediately* as in “After being mixed, the liquid turned red immediately.” For comparison, consider a similar sentence, “After being mixed, the liquid turned red 2 minutes later” with the *positive calendar time duration* *2 minutes*.

## 4. “A year has 365 days”

Since the sentence does not mention any calendar, its meaning is pragmatically incomplete; it means that the construction encoded by the sentence is typed to  $v$ -construct a proposition, but it is an open construction with a free variable  $c \rightarrow *_{\tau_0}$  ranging over calendars. From the linguistic point of view, this is a case of *ellipsis*. Only when obtaining a piece of information about a pragmatic context, the sentence can be completed so that its meaning would be a closed construction denoting a proposition the truth-condition of which can be evaluated like, for instance, is the case of the sentence “A year has 365 days according to the Julian calendar”.

Both expressions ‘a year’ and ‘365 days’ denote a *calendar time duration*. Thus, the meaning of the sentence “a year has 365 days” is this open construction:

$$\lambda w \lambda t \ [^0_{=(\tau\tau)} \ [^0 Year\ c_{w\tau}] \ [^0 365 Days\ c_{w\tau}]]$$

*Year, 365Days*/ $((\tau\tau)*_n)$ ,  $=_{(\tau\tau)}/(o(\tau\tau)(\tau\tau))$ ,  $c \rightarrow *_{\tau_0}$ ,  $w \rightarrow \omega$ ,  $t \rightarrow \tau$ .

Clearly, the construction of the function denoted by the term ‘365 days’ can be further refined. To facilitate this, we are going to define the operation of *adding calendar time durations*, the entity *modifier of calendar time duration* and the operation of *multiplying a calendar time duration* by a number.

### Definition 8

The *addition of calendar time durations*, *AddTD*, is defined as follows:

$${}^0\text{AddTD} = \lambda d_1 \lambda d_2 \lambda e \lambda t \ [[d_2 e] \ [[d_1 e] t]] \\ d_1, d_2 \rightarrow ((\tau\tau)*_n), e \rightarrow *_n, t \rightarrow \tau, \text{AddTD} / (((\tau\tau)*_n)((\tau\tau)*_n))((\tau\tau)*_n)$$

### Definition 9

A *modifier of calendar time duration* is a function from *calendar time durations* to *calendar time durations*.<sup>17</sup> The TIL type of a *modifier of calendar time durations* is thus  $((\tau\tau)*_n)((\tau\tau)*_n)$ .

Note that applying the function *AddTD* to a *calendar time duration*  $d_1$  yields a *modifier of calendar time durations*. If this modifier is applied to a *calendar time duration*  $d_2$ , it yields the *calendar time duration* that is the sum of  $d_1$  and  $d_2$ .

As a consequence of the above observation regarding *February 29<sup>th</sup>, 2016*, we cannot define the multiplication of *calendar time durations* simply as adding time durations, since the details are different for different calendars and their irregularities (such as leap days).

### Definition 10

The operation *MultTD* of *multiplying calendar time duration* is a function of type  $((((\tau\tau)*_n)((\tau\tau)*_n))\tau)$  that associates a real number  $x$  with a *modifier of a calendar time duration*  $M$  such that  $M$  applied to a *calendar time duration*  $d$  yields as its value a *calendar time duration* that is  $x$  times longer than  $d$ .

---

<sup>17</sup> Note that there is no requirement for a *modifier of a calendar time duration* to be a total function.

*Example:* For the Gregorian calendar (*GrC*) it holds that

$$[[[{}^0\text{MulTD } 0.5] {}^0\text{Year}] {}^0\text{GrC}_{wr}] {}^0_{=(\tau\tau)} [[[{}^0\text{MulTD } 0.6] {}^0\text{Month}] {}^0\text{GrC}_{wr}].$$

Also, since *year* is a *positive time duration* according to the Gregorian calendar,  $[[[{}^0\text{MulTD } 0.-1] {}^0\text{Year}] {}^0\text{GrC}_{wr}]$  *v*-constructs the negative time duration a *year ago* (again according to the Gregorian calendar).

The resulting *modifier of a calendar time duration* may yield for some *calendar time durations* that are defined in a certain *calendar* a *calendar time duration* that is undefined in this *calendar*, depending on the multiplication number. For example, a week is well-defined in the Gregorian calendar (as 7 days), and any integer multiplication, such as 2 weeks, 3 weeks, and so on are well-defined as well. However, a quarter of a week is not. Therefore, the modifier *multiply by 0.25* applied to *week* yields a *calendar time duration* that is undefined in the Gregorian calendar (but may very well be defined in other calendars).

Thus the more detailed analysis of the sentence “A year has 365 days” comes down to this construction.

$$\lambda w \lambda t [{}^0_{=(\tau\tau)} [{}^0\text{Year } c_{wr}] [[[]^0\text{MulTD } 0.365] {}^0\text{Day}] c_{wr}]$$

*Year, Day/((\tau\tau)\*n), 365/\tau, =(\tau\tau)/(o(\tau\tau)(\tau\tau)), c \rightarrow \*\_{\tau\omega}*

In the interest of better readability, we may improve this analysis by defining the shorthand function *Days*:

$${}^0\text{Days} = \lambda x [[{}^0\text{MulTD } x] {}^0\text{Day}]$$

The analysis of “A year has 365 days” is then this construction:

$$\lambda w \lambda t [{}^0_{=(\tau\tau)} [{}^0\text{Year } c_{wr}] [{}^0\text{Days } 0.365] c_{wr}]$$

## 5. “Adam is 5 years older than Bill”

The sentence “Adam is 5 years older than Bill” does not mention any specific calendar, hence its meaning is again pragmatically incomplete. Traditionally, we consider the age of a person to be a particular number,

e.g. “Adam is 27” (years old). It is, however, the often-unspoken part with the word “years” that raises the question “*according to which calendar*”? And the same applies to the difference in age of two people.

For the sake of simplicity, we introduce two additional pieces of shorthand for *calendar time duration*:

$${}^0\textit{Years} = \lambda x \text{ [[}{}^0\textit{MulTD } x \text{]} \text{ } {}^0\textit{Year}]$$

$${}^0\textit{Months} = \lambda x \text{ [[}{}^0\textit{MulTD } x \text{]} \text{ } {}^0\textit{Month}]$$

As always, we start with the type analysis. Both “Adam” and “Bill” denote individuals and the term “5 years” denotes a *calendar time duration*. The expression “is 5 years older than” denotes a relation-in-intension between individuals, i.e.<sup>18</sup>

$$5\_Years\_Older/(ou)_{\tau\omega}$$

Thus a coarse-grained analysis of our sentence is simply this construction:

$$\lambda w \lambda t \text{ [}{}^0\textit{5\_Years\_Older}_{wt} \text{ } {}^0\textit{Adam } {}^0\textit{Bill}]$$

However, refinement of this analysis is rather complicated. First, there is an ambiguity. Either ‘five years older’ means exactly five years older, or approximately five years older.

The first option is the simpler one and allows us to define the relation of being *five years older* by means of the entity *Older* of type  $(ou(\tau\tau))_{\tau\omega}$ : the ternary relation-in-intension between two individuals and a *time duration*.

$${}^0\textit{5\_Years\_Older} = \lambda w \lambda t \lambda x y \text{ [}{}^0\textit{Older}_{wt} \text{ } x \text{ } y \text{ [}{}^0\textit{Years } {}^0\textit{5}] } c_{wt}]$$

Types:  $x, y \rightarrow i$ ;  $[{}^0\textit{Years } {}^0\textit{5}] \text{ } c_{wt} \rightarrow (\tau\tau)$ ;  $c \rightarrow *_{\tau\omega}$ : a calendar;  $=/((ou)_{\tau\omega})(ou)_{\tau\omega}$ : the identity of binary relations-in-intension.

---

<sup>18</sup> In general, a TIL type of an  $n$ -ary relation-in-intension is  $(((\alpha_1\alpha_2\dots\alpha_n)\tau)\omega)$ , or ‘ $(\alpha_1\alpha_2\dots\alpha_n)_{\tau\omega}$ ’ for short, where  $\alpha_i$  may be any TIL type.

The refined analysis of the first option is thus:

$$\lambda w \lambda t [{}^0\text{Older}_{wt} {}^0\text{Adam} {}^0\text{Bill} [{}^0\text{Years} {}^05] c_{wt}]$$

To analyse the second option, we must introduce a measure of tolerance. Intuitively, the term “60-months-in-the-Gregorian-calendar” seems more specific and therefore would allow for less tolerance. An inaccuracy of 1 month might be negligible for 5 years, significant for 60 months and too much for 1826 days (a best guess for the number of days in 5 years).<sup>19</sup>

This different level of tolerance is, however, lost in the first analysis. Let us again assume that the used *calendar c* is the *Gregorian calendar*. According to this calendar the *time duration* of 5 years is the same as the *time duration* of 60 months. From this it follows that the proposition  $\nu$ -constructed by the following construction is the same as the one  $\nu$ -constructed by the previous construction.

$$\lambda w \lambda t [{}^0\text{Older}_{wt} {}^0\text{Adam} {}^0\text{Bill} [{}^0\text{Months} {}^060] c_{wt}]$$

In other words, these two constructions are equivalent by producing one and the same proposition:

$$\lambda w \lambda t [{}^0\text{Older}_{wt} {}^0\text{Adam} {}^0\text{Bill} [{}^0\text{Years} {}^05] {}^0\text{GrC}_{wt}] = \\ \lambda w \lambda t [{}^0\text{Older}_{wt} {}^0\text{Adam} {}^0\text{Bill} [{}^0\text{Months} {}^060] {}^0\text{GrC}_{wt}]$$

To allow for different tolerance for different units of time (years, months, ...) we apply the relation-in-intension *OlderCal* of type  $(ou((\tau\tau)*_n)*_n)_{\tau\omega}$ ; a relation-in-intension between two individuals, a *calendar time duration* and a *calendar method*. The analysis of the second reading of the sentence is then:

$$\lambda w \lambda t [{}^0\text{OlderCal}_{wt} {}^0\text{Adam} {}^0\text{Bill} [{}^0\text{Years} {}^05] c_{wt}]$$

---

<sup>19</sup> The difference is important in a multi-valued logic where the increasing tolerance results in a lower degree of truthfulness.

## 6. Conclusion

In this paper, we have defined several basic notions needed for the analysis of sentences involving time references, of which the most important are *time duration* and *calendar time duration*, including *year*, *month*, *day*. Moreover, we proposed a method for dealing with *adding* and *multiplying calendar time durations* so as to be able to present a fine-grained analysis of terms like ‘15 hours and 30 minutes’ or ‘5 years ago’ respecting leap days and seconds. To this end we defined a *modifier of calendar time duration*.

There are two interconnected avenues of *further research* that we believe will result in significant contribution to the topic. First, the connection between *time intervals* and *calendar time durations*; we should be able to compute the *time interval* denoted by a *calendar time duration* like for instance that denoted by the term ‘year 2017’. Furthermore, there is a *calendar* object involved in the specification of time points denoted by expressions such as ‘January 1<sup>st</sup>, 2017, 15:30’. This requires further investigation both of the structure of a calendar *date* (i.e. the object denoted by ‘January 1<sup>st</sup>, 2017, 15:30’ before any particular calendar is taken into consideration) and specialties of individual common *calendars*. These include phenomena such as time zones and daylight savings time, and also the fact that the official reference points for many calendars (e.g. the birth of Jesus Christ or the creation of the world) are imprecise at best and made up at worst. It is therefore unreasonable to claim that the time in any computer is computed on the basis of these reference points, because in order that the computer could compute time, it must have some constant time point to start with. To this end is usually used some external impulse, for instance, synchronization with Internet time. Thus the reference point is not settled at zero; rather, the computer takes as the starting time-point, for instance the time “it is now the time 2017-06-01, 14:36:00” and computes time from this reference point.

### Acknowledgements

This research has been supported by the Grant Agency of the Czech Republic, project No. GA15-13277S, *Hyperintensional logic for natural language analysis*, and by

the internal grant agency of VSB-TU Ostrava, project SGS No. SP2017/133, “Knowledge modeling and its applications in software engineering III”.

## References

- DUŽÍ, M., JESPERSEN, B. & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic*. Berlin: Springer, series Logic, Epistemology, and the Unity of Science, vol. 17.
- DUŽÍ, M. (2010): Tenses and Truth-Conditions: a Plea for *If-then-else*. In: Peliš, M. (ed.): *Logica Yearbook 2009*. London: College Publications, 63-80.
- HOBBS, J. R. & PAN, F. (2004): An Ontology of Time for the Semantic Web. *ACM Transactions on Asian Language Information Processing* 3, 66-85.
- JESPERSEN, B. (2005): Explicit Intensionalization, Anti-Actualism, and How Smith’s Murderer Might not Have Murdered Smith. *Dialectica* 59, No. 3, 285-314.
- NAIN, S. & VARDI, M. Y. (2007): Branching vs. Linear Time: Semantical Perspective. In: Namjoshi, K. S., Yoneda, T., Higashino, T. & Okamura, Y. (eds.): *Automated Technology for Verification and Analysis. ATVA 2007. Lecture Notes in Computer Science*. Vol. 4762, 19-34.
- OHLBACH, H. J. (2004): Calendrical Calculations with Time Partitionings and Fuzzy Time Intervals. *Lecture Notes in Computer Science* 3208, 118-133.
- OHLBACH, H. J. & GABBAY, D. (1998): Calendar Logic. *Journal of Applied Non-Classical Logics* 8, No. 4, 291-323.
- PLACEK, T. (2012): On Individuals in Branching Histories. *Synthese* 188, No. 1, 23-39.
- PLOFKER, K. (2011): The Mathematics of Measuring Time: Astronomical Timekeeping and the Sinking-Bowl Water-Clock in India. In: Jardine, D. & Shell-Gellasch, A. (eds.): *Mathematical Time Capsules: Historical Modules for the Mathematics Classroom* 77, 55-62.
- PRIOR, A. N. (1957): *Time and Modality*. Oxford: Oxford University Press.
- PRIOR, A. N. (1962): Tense Logic and the Continuity of Time. *Studia Logica* 13, 133-148.
- PRIOR, A. N. (1967): *Past, Present and Future*. Oxford: Clarendon Press.
- TICHÝ, P. (1980): The Logic of Temporal Discourse. *Linguistics and Philosophy* 3, 343-369. Reprinted in: Tichý (2004), 373-369.
- TICHÝ, P. (1988): *The Foundations of Frege’s Logic*. Berlin & New York: deGruyter.
- TICHÝ, P. (2004): *Collected Papers in Logic and Philosophy*. Svoboda, V, Jespersen, B. & Cheyne, C. (eds.), Prague: Filosofia, Czech Academy of Sciences & Dune-din: University of Otago Press.

## On Tichý's Attempt to Explicate Sense in Terms of Turing Machines

PAVEL MATERNA<sup>1</sup>

ABSTRACT: In Tichý (1969), it is shown that semantics of natural language can be pursued procedurally. Tichý supported his argument by defining elementary functions of logic (truth functions, quantifiers) using Turing machines and attempting to define the sense of empirical expressions using a simple semantic version of oracle. From the way how Turing machines and later constructions are defined it follows that even the sense of empirical expressions can be successfully handled but that the sense and denotation can be in principle effectively obtained while the actual value at the actual world can be, of course, never computed. The present paper comments on this attempt and compares the Turing machines argument with the possibilities given by TIL constructions. Turing machines guarantee the effective character of computing while the constructions do not, but expressive power of constructions is incomparably stronger, not only because Tichý's possible worlds from 1969 are atemporal: they define essentially 1<sup>st</sup> order operations and can be reinterpreted as one possible world enjoying (discrete) temporal changes. Both the TM conception and the "constructivist" one know that the question "which possible world is the actual one" cannot be ever answered by effective (computational) methods and their analyses of empirical expressions are therefore compatible.

KEYWORDS: Oracle – possible worlds – procedure – sense – Turing machine.

---

<sup>1</sup> Received: 2 March 2017 / Accepted: 19 June 2017

✉ Pavel Materna

Institute of Philosophy, The Czech Academy of Sciences

Jiřská 1, 110 00 Praha 1, Czech Republic

e-mail: materna@flu.cas.cz

## 1. Tichý's explication of Frege's *Sinn* in 1968

Frege's ingenious idea from 1892 of mediating the denotation (*reference*) by *sense* (as "the mode of presentation") was baptized "sense" but we know that Frege never *defined* the sense. The idea itself was however guessed to be sound. Nowadays it should be clear that to solve *semantic* problems of a natural language without using some notion essentially similar to Frege's *sense* means to replace semantic criteria by *pragmatics*. This is what Quine did in his famous (1953). His influence was strong, and soon one could read formulations like "Quine teaches us that..." or "Quine refuted the claim that ..." instead of more sober phrasings like "According to Quine..." or "Quine refused the claim that...". Quine's argumentation was interesting and we have to quote a key point of it to appreciate Tichý's *semantic turn*.

In Quine (1953), he says:

Once the theory of meaning is sharply separated from the theory of reference, it is a short step to recognizing as the primary business of the theory of meaning simply *the synonymy* of linguistic forms and *the analyticity* of statements; *meanings* themselves, as obscure intermediary entities, may well be abandoned. (Quine 1953, 22; italics mine, P.M.)

*Terminological remark:* It frequently happens that *meaning* is used instead of *sense*. As for *denotation*, Frege has *Bedeutung*, Church *denotation*, Geach and Black *reference*. Here we will use *denotation* and we will show that *reference* differs from *denotation* when empirical expressions are analyzed. Further, Quine identifies *meaning* with *intension* (cf. Quine 1953, 21).

From the very beginning Quine believed that the only logic deserving this name is just *extensional* (we could say "denotational") logic and even speaking about intensions was suspect for him. Therefore the term *meaning* or *sense* as an indication of something, which could threaten the idea of the extensional monopoly, was characterized as an "obscure entity". Thus we should, according to Quine, try to logically explain some phenomena that are allegedly defined by means of *meaning* and to show that such explanations necessarily fail.

The way Quine argues to show that synonymy and analyticity cannot fulfil their role is well-known. Don't forget however that Quine does not accept the possibility to use *meaning* as to solve the problems that make the task of satisfactorily defining the critical notions not feasible (see Materna 2007).

Now we can compare Quine's criticism with Tichý's (1968/2004). Tichý then did not know Quine's work (if he had known it he would surely have quoted the respective article. The world logical literature could not be sold in Communist Czechoslovakia then), but the text I will now quote looks like a comment to Quine:

In current logic there is a strong tendency to define the sense by means of synonymy or analytical identity of expressions. It stems from the assumption that the relation of synonymy or analytical identity is definable without the notion of sense. "... *It follows a description of the way it can be done.*" This approach is formally correct, but from the semantic-content point of view we can object that this method of defining is quite opposite to our intuition. "*It follows again a description of that way and the respective critique.*" Hence in both cases defining the sense by means of the relation of analytical identity is either to turn over the natural logical sequens of these notions, or to fall into a circular definition.<sup>2</sup> (Tichý 1968/2004, 81)

Tichý's conviction had to be supported by such a notion of meaning (here also denoted by *intension*) that would be compatible with the way the term *meaning/sense* is used in the Natural Language and possesses the (evidently Fregean) property

FP    *Two senses may concern one and the same object*

or in the more terminologically precise variant

FP'    *Two expressions that express distinct senses may denote one and the same denotation.*

---

<sup>2</sup> Here the pagination concerns the translation of Tichý (1968) in Tichý (2004).

Tichý has emphasized that the *classical* conception of sense (in particular the Aristotelian theory), although inadequate from the viewpoint of the contemporary logic, was closer to the truth that “the notion of intension<sup>3</sup> logically precedes the notions of truth, analyticity and synonymy, and not vice versa” than the contemporary attempts, so that “it is just necessary to replace the oversimplified classical family of qualities by a construction that would satisfy the requirements of modern logical analysis” (Tichý 1968/2004, *ibidem*).

The key notion which led Tichý to his solution was mentioned at the outset and in the title of the Czech paper from 1968: *Sense and procedure*. Here he writes:

[i]t is easy to see that, taken in an abstract way, the relation between sentences and procedures is of a semantic nature; for sentences are used to record the results of performing particular procedures. (Tichý 1968/2004, 80)

In the following text, which was intended as a popular Czech version of Tichý’s theory, the author introduces basic concepts of this theory, viz. *universe, intensional base, empirical system, Turing machines, procedures, possible worlds, languages as unities of primitive symbols and interpretation, and synonymity and analyticity*, so that *logical truth* and *analytical truth* can be defined and such “obscure entities” like *sense* or *meaning* get exact definitions.

## 2. The exact formulation in (1969)

In *Studia Logica* paper (see Tichý 1969/2004), Tichý has published a precise version of what has been articulated in Tichý (1968/2004). One basic term is characterized at the very outset, the term, which Quine characterized as an “obscure one”: the term *sense*. Tichý says: “The sense of an expression is an entity linking the expression with its denotation” (Tichý 1969/2004, 96). A most important refinement of this formulation follows: to know the sense of an expression “does not mean actually to know its

---

<sup>3</sup> Here Tichý means by *intension* a family of features, cf. Aristotelian definition.

denotation but to know how the denotation can be found,” so that “it seems to be natural to conceive of concepts as procedures” (Tichý 1969/2004, 97).<sup>4</sup> Applying a procedure to an input data consists in performing various transformations of this data according to a fixed program. Atomic units of these transformations are *steps*, either *autonomous* (the result depends just on the outcome of the foregoing steps) or *empirical* (the result depends on the state of the external world). So an *empirical system* which underlies each language consists of the external world, i.e. (finite) *universe of discourse* containing just bare individuals, and its state that is given by applying the members of *intensional basis* consisting of elementary tests that can be applicable to individuals or couples of individuals etc. The assumption is that all the members of the intensional basis are intensionally independent, and that the number of the available elementary tests is finite, which guarantees that as a conceivable state of the universe can be regarded any combinatorial possibility of the results of applying all the tests (from the intensional basis) to all the individuals (couples of individuals etc.). These possibilities can be called *possible worlds (with respect to) the empirical system*.

An empirical system is according to these definitions in a good sense *finite*. The finite character (to be argued for in the next paragraph and on pages 97, 98 of Tichý's 1969/2004 paper) makes it possible to apply Turing machines and, thereafter, to compare this finite version with the contemporary version of TIL.

### 3. The finite character of the 'empirical system' from 1969

Rewording what Tichý says in (1969/2004, 97) we get

- K* the cardinality of the Universe of discourse
- S* the greatest number *s* such that there are *s*-ary members of the intensional basis

---

<sup>4</sup> Here we can state that the future procedural theory of concepts is suggested, see, e.g., Duží, Jespersen & Materna (2010).

$P_0, P_1, \dots, P_S$  the number of 0-ary, ..., S-ary tests in the intensional basis  
(the sum of the particular numbers)

The sequence  $K S P_0 P_1 \dots P_S$  of natural numbers is called the *dimension* of the system. It makes it possible to define the length of the sequence (W) that corresponds to one realization of applying the tests from the intensional basis. We define this sequence (W) as follows (cf. Tichý 1969/2004, 98):

*$W_r^{qp}$  is 0 (1) if the outcome of the  $q$ -ary  $p$ -ary test from the intensional basis applied to the  $r$ -th ordered  $p$ -tuple of individuals from the universe is positive (negative) in the possible world  $W$ .*

Clearly, the length  $N$  of (W) equals the

Sum ( $\Sigma$ ) of  $P_i K_i$  from  $i = 1$  to  $S$

so that the *number of possible worlds defined in this way equals  $2^N$* . Interesting news: Having an empirical system we have got, true, many possible worlds but *finitely many of them*.

We will see, however, that some important properties of an empirical system have to be sacrificed to get a finite number of possible worlds. Anyway, an easy way how to represent an empirical system by a set of natural numbers less than  $N$  has been shown.

Now we will first comment on some important consequences of Tichý's conception of possible worlds in the years 1968 and 1969.

#### 4. The actual world

Tichý states (see Tichý 1969/2004, 98) that there is a possible world having a privileged status: the so-called *actual world*. The actual world will be denoted by  $W^*$ . It is of the form (W), where instead of  $W_r^{qp}$  is 0 (1) if the outcome of the  $q$ -ary  $p$ -ary test from the intensional basis applied ... we have *basis actually applied...* (emphasis mine, P.M.). Here a very important claim follows: "It goes without saying that within semantics we have no way of telling which of the possible worlds with respect

to a given empirical system is the actual one, since this is an (or rather *the*) empirical question" (Tichý 1969/2004, 98).

This statement is of key importance. It can be argued for in more details and we will therefore incorporate into the present paper a brief *intermezzo*.

*Intermezzo: TIL on the semantics of empirical expressions.* Tichý's Transparent Intensional Logic whose contemporary version can be found in Duží, Jespersen & Materna (2010) developed from Tichý's early intuitions and his papers from 1968 and 1969. Here we would like to show some details of arguing for the foregoing claim concerning the actual world. We will apply the contemporary TIL to the way Tichý talks about possible worlds in 1968, 1969. Thus what does TIL say about semantics of empirical expressions:

The sense of an empirical expression is always a construction that constructs a non-trivial intension. The type of an intension is  $((\alpha\tau)\omega)$ ,  $\alpha$  any type,  $\tau$  for time moment,  $\omega$  for possible worlds;  $\alpha_{\tau\omega}$  is used as an abbreviation. Thus an empirical expression denotes a *function* (*viz. of possible worlds and times*), which means that *it never denotes an actual object*. The expression *the Pope*, for example, denotes the function  $\iota_{\tau\omega}$  (an individual role/office), never a concrete individual that happens to be the Pope. Applying this function to  $W^*$  we would get the actual Pope, e.g. John Paul 2th in the year 1980 *if the  $W^*$  were a concrete possible world, but the type of  $W^*$  actually is  $(\omega\omega)$ , since the  $w$ , where  $W^*$  is actual, is, of course,  $w$* . Thus  $W^*$  is again a function, and we cannot know which its value is: the logic itself does not give us a criterion – as a semantic question it is beyond deciding.

The vain effort to 'construct' the transition from the denotation to *reference* can be explained just by this theory of the role and character of the *actual world*. The possibility to distinguish denotation and reference has been exploited only by TIL, where *denotation* is determined by the sense and is thus an analytic part of Frege's scheme whereas *reference* is defined as the (contingent!) value (if any) of the denotation of an empirical expression in the actual world. This distinction holds true only when empirical expressions are analyzed. (See Duží, Jespersen & Materna 2010, 176-187, 13-14.) *End of Intermezzo*.

## 5. The empirical system from 1969 is atemporal

The greatest distinction between the 1969 paper and the TIL version from 1988 and 2010 consists in the fact that the former unlike the later is *atemporal*. Thus the length of the ‘dimension’ can be always a natural number. It means however that Tichý’s empirical system corresponds at most to one temporal unit (moment) and represents the answer to the question “*what can happen during one moment when the given intensional basis is applied to the members of the given Universe*”. Our intuition has it however that one and the same possible world as a consistent combination of the outcomes of ‘tests’ can enjoy temporal changes so that it is rational to respect not only the procedures that lead to transitions to a distinct possible world but also changes that occur within one and the same possible world, namely temporal changes. The atemporal conception has been abandoned by Tichý in his (1988; see especially p. 186), and in the following pages the temporal conception is analyzed and argued for. Observe further Tichý (1980/2004, 373), where Tichý already works with temporal analyses.

We can see that even in very simple cases the analysis (for example of the sentence “It rains in Prague” would be counterintuitive if the type of propositions were  $(o\omega)$ . Clearly, this atemporal decision is incompatible with our most basic intuition: in one and the same possible world we can state once T, once F, our system is inconsistent. As soon as the type is  $((o\tau)\omega)$ , or  $o_{\tau\omega}$  for short, our intuition is restored, every possible world leads to a function from time to  $\alpha$  (here  $o$ ). So *give me the possible world and the time and I will say whether it rains in Prague*.

As soon as *time* is accepted the idea of a finite empirical system has to be abandoned, at least when time is continuum, which is the case of TIL, where  $\tau$  are *real numbers*. The papers from 1968 and 1969 remain to be finite in the sense explained above. The fact that  $N$  has been defined in that sense means that what has been actually defined is surely different from our intuitive notion of a class of possible worlds: we have already guessed that those definitions determine rather a class of temporal changes within one possible world. Let any such class be called  $Nw_i$  where  $i > 0$  and  $N$  is Sum  $(\Sigma)$  of  $P_i K_i$  from  $i = 1$  to  $S$ , where  $K$  is a constant universe and the values of  $S$  and  $P_i$  are dependent on  $w_i$ . Adding the temporal dimension means that a class of possible worlds arises as a class the members of which are  $Nw_i$  as interpreted in this way.

## 6. Procedures

Tichý's *empirical system* has been type-theoretically defined: we have got *universe* with *individuals* as its members, truth-values (0, 1), so that 1<sup>st</sup> order relations are definable. Besides, possible worlds are defined as consistent combinatorial possibilities w.r.t. outcomes of 'tests'. Before a 'formal language' is introduced a kind of Turing machines is described whose last quadruplet represents the state of the external world ("oracle"), i.e. the world *W*. (See the chapter 4.) The actual world w.r.t. the given empirical system is just the world that is the outcome of an actual Turing computation). An illusion may arise that, therefore, Turing machines could compute the value of the denotation of an empirical expression, which would contradict to the important claim that within semantics this is impossible (ch. 4). Simply: Which world is actual *in the world W*? Clearly, the world *W*. The *denotation* of the empirical expression "the actual world" is the function of the type  $(\omega\omega)$ . A *reference* is the contingent value of this denotation in the actual world. This reference can be never reached by a computation *ex definitione*: "the actual world" is an *empirical* expression, i.e., it is not a semantic expression.

Anyway, Tichý introduced in 1969 Turing machines to show that the *sense* of an expression is a *procedure* and he demonstrated the possibility of an exact explication of this notion. The 'mature' stage of TIL has instead defined the notion of *construction* (see Tichý 1988; Duží, Jespersen & Materna 2010). Constructions are no more atemporal and make it possible to procedurally define hyperintensionality (unlike Montague). In Tichý (1969/2004) procedures have been explicated in terms of Turing machines and the expressive power of this explication has been very weak in comparison with constructions. The choice of Turing machines meant that the *effectivity* of the respective computation was guaranteed. The transition to *constructions* lost this guarantee. As Tichý says in (1986/2004):

But not every construction is an algorithmic computation. An algorithmic computation is a sequence of *effective* steps, steps which consist in subjecting a manageable object (usually a symbol or a finite string of symbols) to a feasible operation. A construction, on the other hand, may involve steps which are not of this sort. An application of any function to any argument, for example, counts as a legitimate

constructional step; it is not required that the argument be finite or the function effective. Neither is it required that the function constructed by a closure have a finite domain or be effective. As distinct from an algorithmic computation, a construction is an *ideal* procedure, not necessarily a mechanical routine for a clerk or a computing machine. (Tichý 1986/2004, 613)

Let us return to Tichý's *empirical system* from 1969. Here all relevant notions are defined in terms of which the semantics of the syntactically simple languages is definable, namely languages with a fixed syntactic basis  $B$ , where  $I$  is *interpretation* that assigns the primitive symbols of  $B$  with their *senses* given by the empirical system  $E$ . Then the *value* of a well-formed formula or a term is given for a given *valuation* and *truth* of a well-formed formula, logical truth and analytic truth are safely defined.

(A linguistic remark: When Tichý *quotes* Frege's paper he speaks about *Sinn and Bedeutung* but he translates *Bedeutung* as *meaning* (which corresponds to a translation from German). We should not forget however that in the contemporary TIL literature Frege's idea is better understood if *Sinn* is what Frege would call *meaning* and *Bedeutung* would be interpreted as *denotation*.)

*Constructions* are much more complicated and are not limited by 1<sup>st</sup> order, they are not atemporal and are able to solve more important problems due to greater expressivity but effectivity is not guaranteed (see above). All the same some essentially new approach to analysis of natural language can be stated in the 1968 and 1969 papers.

Summarizing: Unlike Quine and most logicians (not only) then Tichý does not believe that the only proper logic is a set-theoretical logic, and from the very beginning he applies his logic to analyzing natural language on the assumption that the empirical expressions do not denote actual objects but offer the possible ways how to express their *senses*. The senses of empirical expressions are given by *tests* and registered in an *intensional basis*. The *interpretation* of the primitive symbols assigns *senses* rather than denotations to them.

A comparison with *constructions* in TIL in this respect is interesting. The *sense* of an expression (in the given language) is always the *construction* which is the result of logical analysis. Thus intensions as functions from possible worlds are what is *denoted* by the respective (empirical)

expression i.e. what is *constructed* by the respective sense. In the 1968, 1969 papers it seems that senses are very similar to this ‘mature’ conception. Indeed, first we read: “*The sense of an expression is an entity linking the expression with its denotation*” (Tichý 1969/2004, 96). This is just the sound idea underlying Frege’s *Sinn*. Now Tichý refines this idea as follows: he asks, what it means to know the sense, for example of the word “taller”, and says: “*to know a method or procedure by means of which the relation (here “taller”) can be identified.*” Atomic units of the respective transformations are *steps* and we can distinguish *autonomous* and *empirical steps*: the former are as their outcomes dependent just on the preceding steps, the latter are dependent on the state of the universe. Thus if the outcome of a procedure depends on the state of the universe then the respective procedure corresponds to an empirical expression. Thus the expression denotes what the outcome of the respective procedure is. So we can have a (rough) translation of the language of 1969 to the language of TIL:

1969	1988
Concept is a procedure	Concept is a procedure
Procedure is a Turing machine	Procedure is a <i>construction</i>
Outcome of (T, <i>n</i> )	The entity (if any) constructed by the construction

In particular: Outcome of (T, *n*) for empirical transformations is a non-trivial *Intension*: every outcome of a (T, *n*) is a (maybe nullary) *function*.

## 7. Conclusion

1. Tichý’s work from 1968 to 1988 shows a nice example of the development of a basic intuition to its fully explicated form.
2. The main idea is explicating the intuition concerning the *meaning/sense* of NL expressions as a *procedure* (1968 a Turing machine, 1988 a construction).

3. Instead of simply accepting the standard conception of formal language (syntax + axioms and rules + interpretation) Tichý exploits Turing machines as modeling the sense (together with an intuitive conception of NL).

4. What is added or what modifies the 1968/1969 theory to get the mature stage of TIL is, essentially, temporality and higher-order types. This makes it possible to get a *procedural theory of hyperintensionality*.

### Acknowledgement

Work on this paper has been supported by Research Grant No. 17-156455 of the Czech Science Foundation.

### References

- DUŽÍ, M., JESPERSEN, B. & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic*. Berlin: Springer.
- FREGE, G. (1892): Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25-50.
- MATERNA, P. (2007): Once mMore on Analytic vs. Synthetic. *Logic and Logical Philosophy* 16, 3-44.
- QUINE, W. V. O. (1953): *From a Logical Point of View*. New York & Evanston: Harper & Row.
- TICHÝ, P. (1968/2004): Smysl a procedura. *Filosofický časopis* 16, 222-232; translated as 'Sense and Procedure' in: Tichý (2004), 77-92.
- TICHÝ, P. (1969/2004): Intensions in terms of Turing Machines. *Studia Logica* 26, 7-25; reprinted in: Tichý (2004), 93-109.
- TICHÝ, P. (1980/2004): The Logic of Temporal Discourse. *Linguistics and Philosophy* 3, 343-369; reprinted in: Tichý (2004), 373-401.
- TICHÝ, P. (1986/2004): Constructions. *Philosophy of Science* 53, 514-534; reprinted in: Tichý (2004), 601-621.
- TICHÝ, P. (2004): *Collected Papers in Logic and Philosophy*. Svoboda, V., Jespersen, B. & Cheyne, C. (eds.), Prague: Filosofia, Czech Academy of Sciences, & Dune-din: University of Otago Press.

# Radical Rationalization Accommodates Rampant Irrationality

JOACHIM LIPSKI<sup>1</sup>

**ABSTRACT:** According to a classic position in analytic philosophy of mind, we must interpret agents as largely rational in order to be able to attribute intentional mental states to them. However, adopting this position requires clarifying in what way and by which criteria agents can still be irrational. In this paper I will offer one such criterion. More specifically, I argue that the kind of rationality methodologically required by intentional interpretation is to be specified in terms of psychological efficacy. Thereby, this notion can be distinguished from a more commonly used notion of rationality and hence cannot be shown to be undermined by the potential prevalence of a corresponding kind of irrationality.

**KEYWORDS:** Cognitive biases – holism – intentionality – interpretation – rationality – reason-explanation.

## 1. Introduction

Some fields, with intentional psychology and economy chiefly among them, methodologically require the assumption that human beings are “rational animals”. At the same time, claims that irrationality runs rampant in

---

<sup>1</sup> Received: 24 March 2017 / Accepted: 2 August 2017

✉ Joachim Lipski

Research Center for Neurophilosophy and Ethics of Neuroscience  
Ludwig-Maximilians-University Munich  
Schellingstr. 10/III, Room J 308, 80799 Munich, Germany  
e-mail: joachim.lipski@gmx.de

the general population are not hard to find. For example, Bertrand Russell lamented:

Man is a rational animal – so at least I have been told. Throughout a long life, I have looked diligently for evidence in favour of this statement, but so far I have not had the good fortune to come across it, though I have searched in many countries spread over three continents. (Russell 1950, 71)

In a more timely treatment of the topic in current cognitive science, Stanovich echoes this sentiment when he remarks that due to irrationality,

[P]hysicians choose less effective medical treatments; people fail to accurately assess risks in their environment; information is misused in legal proceedings; millions of dollars are spent on unneeded projects by government and private industry; parents fail to vaccinate their children; unnecessary surgery is performed; animals are hunted to extinction; billions of dollars are wasted on quack medical remedies; and costly financial misjudgments are made. (Stanovich 2003, 293; also cf. Stanovich 2009, 197 f.)

In this paper, I will explore whether said potentially rampant irrationality can conflict with the intentional (or “folk-psychological”) method of mental explanation. *Prima facie*, the assumption that there is such a conflict is motivated by a classic position in analytic philosophy, which holds that, in order to be able to explain a person’s intentional mental states and actions and to be able to attribute the former and responsibility for the latter to her, we need to interpret her as by and large rational. In this paper, I will argue that mental state attribution by way of intentional interpretation and the position we might call rationality-skepticism (as just expressed in Russell’s and Stanovich’s quotes) track two distinct notions of rationality, and I will offer a criterion for distinguishing them. Specifically, the kind of rationality methodologically required by intentional explanation is tied to psychological efficacy relative to an agent’s mindset, whereas the other is not.

To be sure, the possibility of irrationality is a classic and often-discussed topic. What I wish to add to the debate, which for present purposes

we can think of as having started with Davidson's famous entry (see Davidson 1980, 21-42; see my later fn. 10), is an explicit reconciliation of methodological rationalization and a common-sense theory of irrationality.<sup>2</sup> What we can find to dominate the literature are positions which take sides in this conflict by trying to demolish one of the opponents, either by denying the reality of irrationality (e.g. Bratman 1979; Buss 1997; Arpaly 2000) or by denying that explanation of minds require rationalization in the relevant sense (a denial of rationality; e.g. Mele 1987, 37). However, I believe there is room for acknowledging both; hence, I will argue for their reconciliation and the dissolution of their purported conflict. If anything, my strategy is similar to Holton's (1999), insofar as I think of irrationality as not touching methodological rationalization – but that is where the similarities between Holton's and my account already end.

## 2. Intentional explanation and rationalization

One way of explaining an agent's behaviour is by attributing intentional mental states to her (cf. Cummins 2000, 127 ff.; Fodor 1989, chap. 1). Following Davidson, a methodological prerequisite for intentional explanation is to construe the agent by and large – and as far as possible – as rational:

[I]f we are intelligibly to attribute attitudes and beliefs, or usefully to describe motions as behaviour, then we are committed to finding, in the pattern of behaviour, belief and desire, a large degree of rationality and consistency. (Davidson 1980, 237)<sup>3</sup>

In order to understand an agent, “we will try for a theory that finds him consistent, a believer of truths, and a lover of the good (all by our own

---

<sup>2</sup> In recent discussions of cognitive biases, the term “rationalization” has been popularly used as referring to the act of trying to make one's actions appear rational after the fact (so-called “post-hoc rationalization”; cf. Sie & Wouters 2010). However, I only use “rationalization” in its Davidsonian sense, which simply means the act of interpreting agents as by and large rational. According to Davidson, it is methodologically required when attributing mental states (see my section 2).

<sup>3</sup> Regarding this point, see also Davidson (1980, 221 f.); Davidson (2001a, 196-200); Lewis (1983, 113); and Dennett (1987, 19, fn. 1).

lights, it goes without saying)” (Davidson 1980, 222). According to this view, we must attempt to assign beliefs and desires which, by rational standards, jointly produce the actions we can observe an agent to carry out. To say that they jointly produce the action by rational standards is to say that, when taking means-ends beliefs and desires as premises (or “primary reasons”; Davidson 1980, 3-18) in a practical syllogism (Broadie 1986), the conclusion, whose content is the respective action, is logically derivable from them.

According to Davidson (whose writings, along with Quine’s, form the *loci classici* for methodological assumptions of rationality), the reason for why rationalisation is necessary is that in interpreting an agent we need to untangle her observable behaviour – which tracks what the agent holds true – into its two aspects, namely belief and meaning (Davidson 2001a, 148, 195). The only way to untangle this vector is by maximizing the truth of an agent’s beliefs (or their *correspondence* with actual facts; see Davidson 2001a, 196) on the one hand and the *coherence* of her intentional states on the other (see Davidson 1980, 237). That is, in order to be ascribable to an agent, intentional states have to stand in rational relations to her other mental states, to her actions, and to the world. Hence, irrationality is limited:

Coherence here includes the idea of rationality both in the sense that the action to be explained must be reasonable in the light of the assigned desires and beliefs, but also in the sense that the assigned desires and beliefs must fit with one another. The methodological assumption of rationality does not make it impossible to attribute irrational thoughts and actions to an agent, but it does impose a burden on such attributions. We weaken the intelligibility of attributions of thoughts of any kind to the extent that we fail to uncover a consistent pattern of beliefs, and, finally, of actions, for it is only against a background of such a pattern that we can identify thoughts (Davidson 2001a, 159).

As pointed out by Searle (2000, 106), intentional explanations work the way they do because their explananda are both rationally and causally derivable from the ascribed mental states: That is, if someone just drank a glass of water, then knowing that she was thirsty and that she believed she could quench her thirst by drinking a glass of water explains her drinking.

Invoking causal relations is necessary insofar as we cannot merely rely on rational relations to intentionally explain actions: For instance, the logical relations inherent in the syllogism “Drinking a glass of water is a way to quench thirst; she is thirsty; thus it would be reasonable for her to drink a glass of water” do not, by themselves, establish the required causal-psychological relation; i.e., the premise-conclusion-relations do not by themselves establish that the conclusion expresses a psychological motivation caused by the truth of the first and the instantiation of the second premise. Because even if it was in this sense rational for someone to drink a glass of water, two things may keep her from drinking: further reasons speaking against drinking or any type of external (i.e. non-mental) obstacle. We distinguish these two cases by saying that we either *decided* against drinking or that we were *kept* from drinking. The first implies rational control of the agent, the other a non-mental obstacle (which may be a brick wall just as much as a mental disease – meaning this obstacle does not have to be external to the body, but rather beyond agential control). It is the first case with which we are presently concerned, since it says something about the agent’s mindset in regard to rationality whereas the latter does not. That is, even if a thirsty person eventually decides against drinking, the fact that she had a reason for drinking is not to be disregarded in an account of her mental state. Rather, we say that the agent had conflicting reasons, and if we wish to continue thinking of her as rational, we should say that the stronger reason won out and caused her not to drink.<sup>4</sup> This way, the form of a causal and rational explanation is maintained, even if some reasons (such as her thirst) ultimately proved not to be causally effective for her action.<sup>5</sup>

Reflecting these two components, causality and logicity, this kind of explanation is both nomological and normative: Its explanatory form, the practical syllogism, is formally analogous to deductive-nomological (DN)

---

<sup>4</sup> Sometimes, weaker reasons may win out, in which case we speak of akrasia or weakness of the will (cf. Davidson 1980, 21-42). I will not explore the specific issue of akrasia here.

<sup>5</sup> I tacitly join Searle (1979, 85-87) in assuming that reasons and causes need not be mutually exclusive. Further, I assume that intentional explanation is a quasi-scientific endeavour (cf. Sellars 1997, 90-117; Fodor 1989, 7; Davidson 1980, 221; Dennett 1991, 28 f.).

explanation (see Hempel & Oppenheim 1948), which also combines causality and logicity. For this reason, and for ease of exposition, I will henceforth call explanatory intentional psychological generalisations “psychological laws”.<sup>6</sup> One reason for why psychological laws have explanatory value is that they can be found to be widely applicable. This means, firstly, that their explanatory categories have been and continue to be instantiated across many different individuals and events and, secondly, that their instantiations are generalizable (i.e. that they can be viewed *as instantiations* of a general law) due to commonly known, if usually implicit, criteria for what counts as having the kind of psychological attitude that is attributed (e.g. it is generally known that someone’s raising their voice can provide evidence for attributing anger to them) (Cummins 2000, 127). Which is to say that reasons are psychologically efficacious: People act for reasons, their actions can be influenced by providing them with reasons, anyone’s actions can be predicted on the basis of knowing her reasons for acting, and so on (for a defence of this view see Fodor 1989, chap. 1). Additionally, their being rational generalisations means that the explanatory value of psychological laws also depends on their gaining their general applicability not just by expressing causal relations, but by expressing rational ones: by stating what is rational, and by thereby either being normatively binding themselves or by descriptively referencing such norms of rationality.

A caveat is in order. Clearly, there are psychological laws stating general effects which are not rational, such as in the case of active implicit biases (see Tversky & Kahnemann 1974; Sunstein 2005; Gigerenzer 2008; Sie & Wouters 2010, 126-128.).<sup>7</sup> That the causes of irrational cognition

---

<sup>6</sup> By which I neither mean to imply that psychological explanation is as strict, or as general as other kinds of explanation usually associated with the DN account, nor that we can generally conflate the two notions. (For whether and in what way explanation in psychology is “lawlike” in a stronger sense see Cummins 2000 and Bechtel & Wright 2009.)

<sup>7</sup> It should be noted that there is a tendency in recent moral psychology to operationalize reason as a process of conscious deliberation which should immediately precede decisions, judgments or, generally, actions. However, since what studies often find to precede actions appears as an intuitive, automatic process or impulse, it has been claimed that reason is not a cause of actions at all, that humans are not reasonable beings, or that no such thing as reason, understood as a process of deliberation, actually

and/or behaviour in such cases are “implicit” means that they are typically not ascribable agentially: People are usually not aware of them and will even provide “post-hoc” reasons to justify their biased conclusions (compare fn. 2). Since our behaviour is thereby revealed to be at least partially explained by causes which are not reasons, it seems that we might have to deny, firstly, that actions are caused by reasons, and, secondly, that psychological laws rely on reasons as causes.

Now, the second point should be partially conceded, leading to a clarification of the notion of “psychological explanation”: When they are not concerned with the explanation of behaviour which is under agential control and explainable by the ascription of intentional attitudes, psychological laws need indeed not invoke rational relations: They simply need to capture how thoughts and/or behaviour depend on internal and/or external conditions, and since agents may well behave systematically (i.e. in a specifiable and generalizable way) irrationally in the sense implied by research about implicit biases, there can turn out to be psychological laws of irrationality (cf. Ariely 2010). Of course it is simply wrong to require psychological explanation tout court to have to rely on reason explanations; that much should already be clear from even superficially browsing the current psychological literature. However, it is worth noting that the stated psychological effects are only describable as irrational when contrasted with an appropriate rational norm. Therefore, not only does rationalization constitute a methodological requirement for mental state attribution, it also supplies a foil for singling out biases as such, namely as deviations from what is normatively required or desired (cf. Davidson 2004, 180).

The first point, however – that actions are caused by reasons –, is not falsified merely by the existence of biases and similar confounders of rational cognition and behaviour. Their mere existence does not show that reasons are not generally causally effective and that reasons do not make for valuable predictors of behaviour. Even establishing that biases have the potential to override rational reasoning processes is not enough for inferring that reasons are generally ineffective and that reason-explanations are invaluable (cf. Triskiel 2016, 88 f.). Indeed, it would be irresponsible to

---

exists. Obviously, nothing in this paper called reason needs to be operationalized in this way (cf. Sauer 2012; 2017).

disregard reasons as commonplace determinants of behaviour: For example, knowing that Smith believes that boarding for his flight is about to start at gate 7 enables you to predict, *ceteris paribus*, where Smith is going to be next. Insights about biases, as valuable as they are for adjusting and correcting biased cognition and behaviour, cannot possibly undermine this kind of explanation in general (compare Fodor 1989, chap. 1). Rather, what research about biases shows is that they make for *additional* psychological causes beside reasons. And whenever intentional explanation remains valuable, methodological rationalization needs to be applicable. Hence, insights about biases can be seen as restricting the present domain of inquiry: They reveal the conditions under which identifying psychological causes does not require rationalization in Davidson's sense.

Now, it is still not entirely clear what the kind of rationalization required for intentional explanation exactly amounts to. While the claim that rationality must be ascribed methodologically indeed suggests a conflict with the possibility of irrationality running rampant, it is unclear how severe that conflict must be. For instance, just how irrational can people be without becoming intentionally unexplainable? Davidson clearly thought that there is a trade-off between the ascription of irrationality and intentional explainability (see Davidson 2001a, 159, as quoted at length above), but he would not draw a line (indeed, given his and Quine's stance on the indeterminacy of translation and interpretation, we should assume that, according to this position, there is no specific line, but that there are indefinitely many potential lines). But there are many further pertinent questions: If for every agent at any given time there is a maximally rational intentional description of her thoughts and actions available, under what conditions should we stray from it in our actual description of her thoughts and actions? How far could or should we stray from it? Should all of her mental states and actions exert the same "rationalization pressure" on this description, or are there "tentpoles" or minimal requirements (cf. Cherniak 1981) – i.e. restricted sets of mental states and/or actions whose consistency takes precedence over those excluded from these sets? If so, which are these, and to what degree(s) do they take precedence? And so on.

Given the possibility of answering these questions differently, and the added possibility of combining different answers, the possible routes to what can be viewed as an adequate kind of rationalization methodology multiply and diverge rapidly. Since I cannot consider all of these views, I

will fashion my following argument so as to address the strictest possible rationalization methodology. That is, what I will point out in the following will serve as an argument for establishing that even the strongest possible form of rationalization does not conflict with the possibility of irrationality running rampant. If this argument succeeds, then it can be ruled out that *any* methodological requirements of intentional explanation conflict with irrationality running rampant. This strongest possible form of rationalization is the following: If a person A thinks P or does Q, where P is an intentional mental state and Q is an intentional action, then it follows (just from the methodology employed in order to determine that A thinks or does P) that A's mental states are consistent, that holding P is rational in relation to A's other mental states, and that the best explanation of Q is a logical implication of P's pertinent instrumental belief(s) and desire(s). Call this form "radical rationalization". It is radical because it makes irrationality, as far as the concept pertains to the relations between A's mental states, as well as between her mental states and actions, conceptually impossible: Whatever A thinks or does is always the most reasonable thing to think or do from her point of view.

### 3. Two kinds of rationality

While it does not coincide with the distinction between the causal and rational aspects of intentional explanation, the dichotomy we are about to explore is rooted in it. As pointed out in the previous section, intentional explanations have to pick out a reason that is or was efficacious in the agent's mind – one that could cause her action by being part of her mindset (or "mindware"; Perkins 1995; Stanovich 2009). This notion of efficacy does not require an immediate awareness or subjective transparency of one's reasons, but marks a kind of reason which is an efficacious psychological cause. Any kind of unconscious or implicit cause can qualify as well. This notion is merely to be distinguished from reasons which are not present in an accurate description of an agent's mind, and/or which are psychologically inefficacious (relative to the individual and the moment which the description refers to).

I will define psychological efficacy as follows: A reason R is psychologically efficacious iff (1) R belongs to a person P's mind, (2) R has (direct

or indirect) causal powers in regard to P's actions, and (3) R's potential, anticipated or actual causal effects are, under a relevant description, logically consistent with R. Some clarifying remarks: (1) effectively distinguishes psychologically efficacious reasons from mind-external reasons by which an agent could also be judged irrational, thereby providing two distinct grounds for attributions of irrationality; (2) establishes causality between reasons and actions, and (3) establishes the rationality relation between reasons and actions. The descriptions mentioned in (3) are relevant if they manage to plausibly relate the respective effect to the agent's intentions (if only by construing the effect as a deviation from what was intended). (2) and (3) jointly allow psychologically efficacious reasons to serve their characteristic role in intentional explanation (that is, (2) and (3) explicate what it means to figure in a practical syllogism).

Psychological efficacy, so construed, pertains only to "internal components", if you will, of a given subject's mind. As mentioned, this construal takes a hint from Stanovich's (2009, 129) notion of *mindware*, which he defines as the totality of the memory-stored entities guiding decision-making and problem-solving. Like my notion of psychological efficacy, his construal marks things which (by way of being mentally stored and retrievable) belong to individual agents' minds and can causally affect their actions and/or their other mental states. Accordingly, psychologically efficacious reasons contrast with reasons which lack such efficacy because they either do not belong to P's mind (such as reasons P is oblivious to; cf. Stanovich 2009), or because they are dismissed by P (e.g. due to their being judged invalid and/or irrelevant).<sup>8</sup> Since, trivially, intentional interpretation is only concerned with ascribing those states to a person which belong to her own mind, it is reconcilable with the kind of irrationality that stems from acting against reasons which are rational but either external to her mind or which she dismisses.

There are, of course, other reasons apart from those stored in an agent's mind which can causally affect her actions and/or her other mental states. For instance, those given by other minds can do so. If I am told to do something for a reason I had not been aware of, but whose justification I agree

---

<sup>8</sup> Accordingly, reasons become psychologically efficacious by (1) the agent being made aware of them and/or (2) the agent making them part of the respective reasoning process (e.g. by judging them to be relevant for the reasoning process at hand).

with, it can causally affect my mind and/or behaviour, without the content of what I am being told needing to be recalled from my own memory. Accordingly, my use of the word “internal” simply means to mark reasons which mentally belong to the agent in question. This “internalism” is not to be understood as contradicting philosophical externalisms, such as semantic externalism (cf. Burge 1979) or the “extended mind” view (cf. Clark & Chalmers 1998). That is, mental content may well be individuated externally but still be had by individual agents; and bodily external but cognitively seamlessly accessible storage devices (such as notebooks; cf. *ibid.*) can qualify as belonging to individual minds – thereby “extending” them – and hence as in this sense “internal” as well. What is mentally external in the presently relevant sense is what belongs to other minds, to other persons. Hence, the distinction between psychologically efficacious and inefficacious reasons is based on that between the mental content of a subject’s mind and the mental content external to this subject’s mind. The former is (potentially) directly psychologically efficacious, the latter can only be indirectly efficacious by way of communication, transmission, or access to other minds. While all reasons which are acted upon are psychologically efficacious, not every reason by which we assess a given person to be rational or irrational is a psychologically efficacious reason. And since only psychologically efficacious reasons are relevant for intentional interpretation, there is room for a kind of irrationality which does not touch interpretability.

Two immediate ways in which reasons can be inefficacious is when an agent is oblivious to them or when she cannot grasp them. In such cases, reasons cannot even subconsciously or implicitly provide the kind of generalizable explanations which is characteristic for them (which is not to say they cannot have *any* effect on an agent who fails to grasp them; being aware of not grasping them can still have a frustrating effect, or the like). Since by far not all reasons are available to everyone, their being rational alone cannot make them agentially explanatory. So, an agent’s acting on reasons can only ever be rational if they are available to her. If they are not, then her not acting upon them cannot be irrational, and neither need any of her other mental states, which would contradict her potentially doing so, be irrational, since they can only be irrational in relation to cognitively available reasons. (Notably, her ignorance itself constitutes an intentional state which explains why the respective reason

fails to be efficacious – just as anyone’s holding two inconsistent beliefs can be explained by their being unaware of the way they contradict each other. For example, most of us can readily both believe that Elton John sang “Crocodile Rock” and doubt that Reginald Kenneth Dwight sang “Crocodile Rock”, namely when being ignorant of the fact that both names designate the same person.)

When it comes to the methodological requirements of intentional explanation, an agent’s being rational or irrational are not a matter of acting upon or considering everything that is generally rational, everything specified by good reasons (cf. McNaughton & Rawling 2004, 126). Hence, the notion of “all things considered” (cf. Davidson 1980, 21-42), which is notorious in debates revolving around irrationality, does, of course, not mean that *all* things are considered, only that all cognitively available things are. The kind of rationality which matters for intentional explanation is that which is attributed relative to an agent’s mindset and actions. Explanation by intentional states is concerned with psychological efficacy; and what is psychologically efficacious does not coincide with what is rational tout court. As long as what we mean by “reason” is an explanatorily valuable psychological cause, it is implied that this reason is an available, efficacious part of an agent’s mindset.

When I introduced the notion of radical rationalization in the previous section, one might have objected that, given our cognitive limitations, the empirical knowledge we have about human psychology, as well as the practical constraints of everyday requirements, no such picture should assume that intentional explanation requires a maximally rational interpretation of a maximal set of pertinent evidence at any given moment. Even if it were possible for cognitively limited agents like us to perform such interpretations, it would be highly inefficient. Rather, we undoubtedly assume that non-pathological agents use and understand standard senses of the terms belonging to the language they are speaking, that their understanding is somewhat consistent over time, that they are generally interpretable using the same (or similar enough) function mapping actions and contexts to intentional states, and so on.

Consequently, even if radical rationalization should rule out irrationality, failures of rationality may creep back in, namely as markers of diachronic deviations in meaning, deviations from norms of word use, deviations from what actions are standardly performed in a given context, and

so on. But, given what I pointed out just now, these forms of irrationality can also be sharply distinguished from failures of psychological efficacy. This is because, when in doubt about an agent's psychological causes, methodological rationalization requires us to go ahead and assign deviant understandings or deviant mental states as psychological causes. This deviance is a deviance from an external norm (and often merely from a pragmatic one) and has no bearing on that person's intentional explainability. For example, we may very well go ahead and stipulate that a given person means "yaw1" by the expression "ketch" in order to arrive at a sufficiently rational interpretation of her mental states and behaviour (cf. Davidson 2001a, 196). Here, she deviates from the external norm that yawls should be referred to by the word "yaw1". If her behaviour regarding yawls is irrational, it is only to the extent that she violates this external norm, but certainly not in the sense that it exhibits a failure of psychological efficacy. Psychological efficacy is satisfied insofar as her describing a yawl as a "ketch" is psychologically motivated by the perception of what she may very well believe to be a yawl, but *also* mistakenly believes to be referred to by the word "ketch". It is this interpretational ascription of mistakenly violating an external norm, and hence the ascription of a deviant mental state – one which deviates from an external norm – which renders her sufficiently rational.

This brings us to our second sense of the term "reason", namely that which refers to what is expressed by a psychologically external norm. Unlike our first kind, this kind of reason is external to an agent's mindset. This is not to say that such reasons cannot be psychologically efficacious, only that they are not in the case of the agent who is in this second sense irrational. As examples, consider cases in which an agent does something she might be persuaded not to do if she were made aware of its negative consequences. Strikingly, this second kind of reason can as well be internal *and* psychologically inefficacious, as in the case of a smoker's acknowledging that reason demands that ("most reasonably" or "all things considered") one should stop smoking, without at the same time taking this demand as a reason (i.e. psychological cause) for herself to stop smoking. To clearly bring out the difference between these two notions of "reason", note that there is generally no contradiction in specifying an unreasonable (i.e. normatively or logically unsound) desire as being the reason (i.e. the psychological cause) for someone's action. And the only

way we can take this real possibility to not express a contradiction is by acknowledging that these two uses of the word “reason” express two different meanings.

Even though their attribution may also have normative aspects (i.e. agents should follow *some* norms of rationality in order to be able to have reasons at all), psychological causes must, in any robustly empirical psychological theory, be descriptive notions: They must not be attributable on the grounds of logical (“a priori”) criteria alone, but on descriptive ones also, such as by whether they are part of the actual psychological make-up of an agent.<sup>9</sup>

Again, this is not to say that what is rational in a second sense – in the sense that goes beyond what descriptively persists in an agent’s mind – cannot be psychologically efficacious, but that it can be so only by becoming part of an agent’s mindset. That is, reasons beyond our own mindset (from now on referred to as “reasons<sub>2</sub>” which are “rational<sub>2</sub>”) potentially shape our minds and our actions insofar as we have the means and the reasons in our first sense (from now on referred to as “reasons<sub>1</sub>” which are “rational<sub>1</sub>”) to act in accordance with these reasons<sub>2</sub>. That is, in order for mind-external reasons<sub>2</sub> to become mind-internal reasons<sub>1</sub>, they need to be both accessible to our minds and there need to be some reasons<sub>1</sub> motivating their incorporation. For example, perhaps there is someone who desires apples more than oranges, and oranges more than bananas, but bananas more than apples. Any such preference ordering  $A > B > C > A$  is irrational in the sense that it makes us exploitable: on a behavioural interpretation, it means that we are willing to trade A and some sum for B, then to trade B and some sum for C, then C and some sum for A, ad infinitum, thus losing everything while never gaining anything – anything but the satisfaction of our irrational desire, perhaps (cf. Ramsey 1931, 156-198; Davidson et al. 1955). Yet, while they are in this sense irrational, it may be true of anyone

---

<sup>9</sup> Although I am willing to consider that intentional explanation has both normative and descriptive aspects, I will not take a stance on whether it could not rather be *purely* descriptive or empirical (insofar as there might be purely descriptive explication of what it means to follow the kinds of norms of rationality necessary for having reasons). I only mean to say that it must at least be *also* empirical in order to be (quasi-)scientific. This robustness criterion is inspired by Piccinini (2007).

that they have these desires, and in such a case these causally explain their actions.

Firstly, such cases illustrate that irrational<sub>2</sub> mental states can constitute psychological causes (because these only need to be rational<sub>1</sub>). Accordingly, Davidson did not take believing in astrology, flying saucers or witches, intending to climb Mount Everest without oxygen, or trying to square the circle as necessarily irrational (Davidson 2004, 170).<sup>10</sup> (Of course, we should add that they *can* also be irrational<sub>1</sub>, depending on whether they contradict the agent's *other* relevant mental states.) Yet, anyone who has these irrational desires may wise up as a consequence of being made aware of their exploitability and cease acting on such desires. Thereby, such cases secondly illustrate that norms of rationality<sub>2</sub> can become reasons<sub>1</sub> by becoming part of an agent's mindset. Such norms do not *directly* enter into predictions or explanations of intentional psychology, but only by way of their psychological efficacy. Hence, even if we both assume that intentional explanation requires radical rationalization and that irrationality<sub>2</sub> may run rampant, people can still be explainable intentionally.

I wish to point out one last kind of cases when the two kinds of irrationality can come apart, one that is perhaps more interesting than ignorance and failure to grasp the rational validity or content of a norm. In the latter cases, norms cannot be psychologically efficacious because they are external to an agent's mindset. But, akrasia aside (see fn. 4), how can it be the case that reasons are cognitively available to an agent, yet still fail to be efficacious? One way we can construe this possibility is to consider a kind

---

<sup>10</sup> However, we can also find Davidson expressing that he does "not think we can clearly say what should convince us that a man at a given time (or without any change of mind) preferred a to b, b to c, and c to a" (Davidson 1980, 237). This might be read in the following way: Perhaps the principle of transitivity is so fundamental that attributing its application to an agent's mental states constitutes a *conditio sine qua non*. This, in turn, would suggest that Davidson's holism is restricted: If there are mental states *sine qua non*, then not every mental state attribution depends on other mental states. I will not pursue this point further, but only suggest that holism might be defended by considering cases in which, other things being equal, attributing this principle's violation to an agent makes her appear more rational and is thereby justified. (For the question whether there are principles which constitute conditions *sine qua non*, see Cherniak 1981.)

of partiality on behalf of the agent. That is, agents can fail to abide by norms whose abiding by is, *ceteris paribus*, generally found to be rational if they are partial to not abiding by them. Since there can be no general impartiality (and since it need not be generally rational to be partial to a certain cause), a given norm may be judged perfectly rational, while abiding by it is not necessary for a specific agent to be rational. In such cases, we should assume that the agent in question has a set of psychologically efficacious reasons which motivate her to dismiss the external norm by which she might be judged irrational.

Returning to our previous example, someone's action of smoking could be judged irrational if she has sufficient reason to establish the general principle that one shouldn't smoke. (One need not be aware of medical details regarding the consequences of smoking here; vaguely knowing that smoking is unhealthy may already suffice to establish this principle.) In other words, if one has good reason for believing this principle to be rational, but still smokes, then one is irrational<sub>2</sub>. Why does this constitute irrationality<sub>2</sub> rather than irrationality<sub>1</sub>? Why are the respective reasons not psychologically efficacious, even though they are part of the agent's mindset? It is because the psychological inefficacy of a generally rational principle can be accommodated by intentional explanations by way of attributing partiality to an opposing principle to the agent (such as an overriding desire to smoke), thereby establishing rational<sub>1</sub> consistency. Of course, any such partiality (e.g. any desire of this kind) must not be assumed in the face of sufficient evidence against it; but any such evidence must in turn be weighed against the evidence of the agent's smoking in the face of realizing that, generally, one shouldn't smoke.<sup>11</sup>

---

<sup>11</sup> This point may be likened to Davidson's, that reasons of the "all things considered"-type (abbreviated as ATC) may still fail to make for reasons to act, or all-out reasons (abbreviated as AO; cf. Davidson 1980, pp. 21-42). However, I take my point to be more clearly committed to construing the eventual AO as the superior reason to the ATC. According to radical rationalization, we are committed to thinking of the ATC as an external norm: One which the agent thinks should "rationally<sub>2</sub>" be imposed on her, but which she rejects because of her partiality to the AO. Taken literally, there can be no such thing as overriding an internal ATC, since, if all (available) reasons have truly been considered, then the strongest of them must be an AO. According to radical rationalization, it is impossible to decide against one's best reason. One may, of course, still deceive oneself by thinking that what one did was not for

Finally, I would like to address a potential objection, namely that the distinction between rationality<sub>1</sub> and rationality<sub>2</sub> is so vague as not to be helpful. This objection is rooted in the already briefly mentioned worry that, given Davidsonian holism (cf. Davidson 1980, 256 f.; Davidson 2001a, 22, 200; Davidson 2001b, 98), there is no specific list of what mental states a person must have in order to count as rational<sub>1</sub>. One counter to this objection consists in the delineation of a substantial “minimal rationality” (cf. Cherniak 1981). According to this view, in order to be interpretable as agents, persons need only satisfy such minimal criteria of rationality, while the mental states or actions not satisfying these criteria could play the role of being rational<sub>2</sub>. Providing criteria for what is minimally rational ultimately makes for a substantial list of mental states sufficient for agential interpretation. Adopting this view may in turn require a modification of Davidson’s view, since a plausible interpretation of Davidson’s holism – i.e. that there are no specific mental states which a person must hold in order to be interpretable as an agent, but only a sufficient amount – could even be taken to speak against the substantial minimal rationality view by finding that some (or even *any*) mental state(s), holding which is necessary for a given person to be interpretable, could figure as being unnecessary for interpreting a different person. In other words, Davidsonian holism allows that, for any set of mental states which the substantial minimal rationality view might construe as being necessarily had by any agent in order for her to be interpretable, there is a person who is interpretable but does not have these mental states.

In any case, substantial delineations of minimal rationality are in fact unnecessary to defend our present distinction between rationality<sub>1</sub> and rationality<sub>2</sub> against the objection that it would be undermined by not being able to neatly sort mental states or actions into one or another category. That is, the fact that we might not be able to generally sort each mental state into one of the two categories opened up by our distinction does not mean there cannot be a distinction. In order to make the distinction, it is already sufficient that there is a set of mental states for each given person which is necessary for interpreting her as an agent (this is the set which is

---

the best reason one had – but it really only shows that one does not think of the best external reason as making for the best internal one (in a sense which might not be subjectively transparent).

“internal”, “cognitively accessible”, part of her “mindware”, etc. – it is the set which is psychologically efficacious), whereas the rest is unnecessary (whose content corresponds to a second set of “external” reasons, which may be rational<sub>2</sub>). So, even if a substantial “minimal rationality” could not be defended as consistent with Davidsonian holism, our distinction would remain solid.

#### 4. Conclusion

I have argued that the form of rationality methodologically required for intentional explanation comes apart from a common understanding of rationality. What the former requires is for an agent to be consistent both in terms of her mental states and in terms of her behaviour. On the other hand, the idea that human beings are notoriously irrational creatures – an idea which is both popular and also seems to be supported by recent psychological research – is grounded in the perception that people act against plausible external norms, not that there couldn’t possibly be specifiable reasons for their actions. Specifically, I have argued that even the strongest form of methodological rationalization (which I called “radical rationalization”) does not prohibit irrationality in this second sense – simply because the criteria for irrationality in the second sense, namely violating external norms of rationality, are independent of how strongly we phrase our criteria for rationality in the first sense. Whether one needs to be absolutely internally consistent or just more or less internally consistent makes no difference for whether the mental states attributed on the basis of achieving this consistency (can) violate some external norm of rationality. In other words, even on the strongest assumption – that perfect internal consistency already follows from being interpretable –, inconsistency with some external norm is nonetheless possible. Calling someone irrational in this second sense is akin to calling out their reasons as short-sighted, foolish or immoral, not as inconsistent in terms of the methodology required for attributing intentional explanation. While the first kind of rationality, by being connected to methodological rationalization, persists in internal consistency, the second kind does not.

Consequently, we can distinguish between two different uses of the term “rational”: We may say that if Anne were rational, she would vote for

the Green Party, and mean that she should do so in order for her actions to reach a maximum of consistency with her mental states (such as her beliefs about environmental policies). This falls squarely into the domain of intentional ascriptions, for we would strive to find a way to make her voting behaviour consistent with her beliefs. Or we can mean that if her actions were most consistent with what we perceive to be the state of the world, she should vote Green. We might perceive the state of the world to be one of threatening global warming and waning natural resources, and consequently think it only right to vote Green – even if none of Anne’s beliefs are actually consistent with voting Green (as she might steadfastly deny global warming and the waning of natural resources).

Reconsidering our opening quotes, it should be clear that the lamented forms of irrationality are not those methodologically excluded by intentional explanation. Neither Russell nor Stanovich meant to imply that human beings generally fail to weep over a loved one’s death, fail to value acts of kindness, fail to eat when hungry and presented with food, or fail to understand why insurgents would not freely surrender their children to the oppressor. That is, they do not mean to criticize people for failing to understand the basic concepts with which we describe their mental states and/or actions. Rather, both mean to criticize failures of critical and long-term thinking and the like: a lack of adherence to reasonable norms. Criticism of this kind does not touch the fact that we have to assume an agent to be rational in order to render her thoughts and actions intelligible, and that, barring pathological instances, we can expect to be able to generally interpret them in this way.

### Acknowledgements

I would like to thank Steffen Steinert, Ali Yousefi, Mario Günther, Stephan Sellmaier, and two anonymous referees for their helpful comments on an earlier version of this paper.

### References

ARIELY, D. (2010): *Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions*. New York: Harper Perennial.

- ARPALY, N. (2000): On Acting Rationally against One's Better Judgment. *Ethics* 110, 488-513.
- BECHTEL, W., & WRIGHT, C. (2009): What is Psychological Explanation? In: Calvo, P. & Symons, J. (eds.): *Routledge Companion to Philosophy of Psychology*. London: Routledge, 113-130.
- BRATMAN, M. (1979): Practical Reasoning and Weakness of the Will. *Noûs* 13, 153-171.
- BROADIE, A. (1968): The Practical Syllogism. *Analysis* 29, No. 1, 26-28.
- BURGE, T. (1979): Individualism and the Mental. *Midwest Studies in Philosophy* 4, No. 1, 73-122.
- BUSS, S. (1997): Weakness of Will. *Pacific Philosophical Quarterly* 78, 13-44.
- CHERNIAK, C. (1981): Minimal Rationality. *Mind* 90, No. 358, 161-183.
- CLARK, A., & CHALMERS, D. (1998): The Extended Mind. *Analysis* 58, No. 1, 7-19.
- CUMMINS, R. (2000): "How Does It Work" vs "What Are the Laws?": Two Conceptions of Psychological Explanation. In: Keil, F. & Wilson, R. (eds.): *Explanation and Cognition*. Cambridge (Mass.): MIT Press, 117-144.
- DAVIDSON, D. (1980): *Essays on Actions and Events*. Oxford: Clarendon Press.
- DAVIDSON, D. (2001a): *Inquiries into Truth and Interpretation*. 2<sup>nd</sup> ed. Oxford: Clarendon Press.
- DAVIDSON, D. (2001b): *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.
- DAVIDSON, D. (2004): *Problems of Rationality*. Oxford: Clarendon Press.
- DAVIDSON, D., MCKINSEY, J. & SUPPES, P. (1955): Outlines of a Formal Theory of Value. *Philosophy of Science* 22, 140-160.
- DENNETT, D. (1987): *The Intentional Stance*. Cambridge (Mass.): MIT Press.
- DENNETT, D. (1991): Real Patterns. *The Journal of Philosophy* 88, No. 1, 27-51.
- FODOR, J. (1989): *Psychosemantics*. Cambridge (Mass.): MIT Press.
- GIGERENZER, G. (2008): Moral Intuition = Fast and Frugal Heuristics? In: Sinnott-Armstrong, W. (ed.): *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge (Mass.): MIT Press, 1-26.
- HEMPEL, C., & OPPENHEIM, P. (1948): Studies in the Logic of Explanation. *Philosophy of Science* 15, 135-175.
- HOLTON, R. (1999): Intention and Weakness of Will. *Journal of Philosophy* 96, 241-262.
- LEWIS, D. (1983): *Philosophical Papers*. Vol. 1. Oxford: Oxford University Press.
- MCNAUGHTON, D. & RAWLING, P. (2004): Duty, Rationality, and Practical Reasons. In: Rawling, P. & Mele, A. (eds.): *The Oxford Handbook of Rationality*. Oxford: Oxford University Press, 110-131.
- MELE, A. (1987): *Irrationality*. New York: Oxford University Press.
- PERKINS, D. (1995): *Outsmarting IQ: The Emerging Science of Learnable Intelligence*. New York: Free Press.

- PICCININI, G. (2007): Computational Explanation and Mechanistic Explanation of Mind. In: de Caro, M., Ferretti, F. & Marraffa, M. (eds.): *Cartographies of the Mind: The Interface between Philosophy and Cognitive Science*. Dordrecht: Springer, 23-36.
- RAMSEY, F. (1931): *The Foundations of Mathematics and Other Logical Essays*. Braithwaite, R. (ed.), London: Routledge & Kegan Paul, Trench, Trubner & Co. – New York: Harcourt, Brace and Company.
- RUSSELL, B. (1950): *Unpopular Essays*. London – New York: Routledge.
- SAUER, H. (2012): Educated Intuitions. Automaticity and Rationality in Moral Judgment. *Philosophical Explorations* 15, No. 3, 255-275.
- SAUER, H. (2017): *Moral Judgments as Educated Intuitions*. Cambridge (Mass.): MIT Press.
- SEARLE, J. (1979): What is an Intentional State? *Mind* 88, No. 1, 74-92.
- SEARLE, J. (2000): *Mind, Language and Society: Philosophy in the Real World*. London: Phoenix.
- SELLARS, W. (1997): *Empiricism and the Philosophy of Mind*. Cambridge (Mass.) – London: Harvard University Press.
- SIE, M. & WOUTERS, A. (2010): The BCN Challenge to Compatibilist Free Will and Personal Responsibility. *Neuroethics* 3, 121-133.
- STANOVICH, K. (2003): The Fundamental Computational Biases of Human Cognition: Heuristics That (Sometimes) Impair Decision Making and Problem Solving. In: Davidson, J. & Sternberg, R. (eds.): *The Psychology of Problem Solving*. New York: Cambridge University Press, 291-342.
- STANOVICH, K. (2009): *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven: Yale University Press.
- SUNSTEIN, C. (2005): Moral Heuristics. *Behavioral and Brain Sciences* 28, No. 4, 531-573.
- TRISKIEL, J. (2016): Psychology Instead of Ethics? Why Psychological Research Is Important but Cannot Replace Ethics. In: Brand, C. (ed.): *Dual-Process Theories in Moral Psychology*. Wiesbaden: Springer, 77-98.
- TVERSKY, A. & KAHNEMAN, D. (1974): Judgment under Uncertainty: Heuristics and Biases. *Science* 185, No. 4157, 1124-1131.

# Fictional Names and Truth

RICHARD VALLÉE<sup>1</sup>

**ABSTRACT:** By addressing fictional names head on, we risk going back to familiar, ordinary names intuitions and missing what is specific about them. I propose a different strategy. My view is grounded on fictional name sentence utterances and on indexed tokens of such sentences, where an index contains the fictional narrator and the time and location of the token. Using the framework of pluri-propositionalism (Perry 2012), I argue that the semantic relation of reference – ‘x’ refers to y - where ‘x’ is a name, rather than the notion of an object, is central to the debate on fictional names. I also contend that fictional names do not enter into that relation. Tokens of fictional names are individuated with the fictional index of the sentence they originate from. This allows for dispensing with a referent. Indexed fictional name sentence tokens have semantically determined truth conditions, yet they are not truth assessed given facts. In this respect, they have cognitive significance only, and no official or referential content. Indexed fictional name token of sentences are accepted as true, but they are not true.

**KEYWORDS:** Cognitive significance – fiction – names – reference – truth.

## 1. Semantics and fictional names

There is an important distinction to be made between ordinary proper names (‘Barack Obama,’ ‘Angelina Jolie’), which designate objects located

---

<sup>1</sup> Received: 7 August 2016 / Accepted: 19 November 2017

✉ Richard Vallée

Université de Moncton, Campus de Shippagan  
218, Boul. J.-G.-Gauthier, Shippagan, NB, E8S-1X1, Canada  
e-mail: rvallee@umcs.ca

in space and time, and fictional names ('Sherlock Holmes,' 'Superman'), which, by design, do not designate space-and-time-located objects. Sherlock Holmes and Superman are creatures of fiction. They do not exist in space and time, nor do they causally interact with real objects, including speakers. If one finds data suggesting that Sherlock Holmes did exist, the name would simply be removed from the list of fictional names. According to a common, intuitive assumption, fictional names have a semantic referent. In more sophisticated versions of that view, they are said to either simply lack a referent (Braun 2005; Everett 2013) or to connect to fictional objects (Thomasson 1999; Kripke 2013). We are strongly inclined to see fictional names as ordinary names, and to think that there are objects designated by them. Such an inclination might be appropriate for fiction reading, but it gives a misleading perspective on fictional names and it should be resisted in semantics. I contend that such names have neither a referring relation to objects, nor referent. I suggest that they play a cognitive role only. If I am right, we are not in relation with fictional characters by way of reference with fictional names. Fictional names differ from ordinary names, and I will argue that that difference is echoed in the way that they are individuated.

Fictional name sentences have truth conditions and, according to common sense intuitions, some are true. Yet, such sentences raise issues because their truth is not supported by facts: there is no object referred to by 'Superman', and no facts supporting the truth of 'Superman flies,' for instance. If fictional name sentences are true, their truth does not depend on facts. That leaves the semantics of fictional names, and the truth of fictional name sentences, in need of an explanation. A view on such names should dovetail with intuitions on truth of fictional name sentences, and vice-versa.<sup>2</sup> This article aims to articulate some of the important elements of such an explanation.

The usual procedure begins with fictional name and then moves to sentences or fictional name sentence utterances. Addressing fictional names head on, however, we risk going back to familiar, ordinary names intuitions and miss what is specific about them. I propose a different strategy. My

---

<sup>2</sup> In that respect, my perspective differs from Sainsbury's, for whom 'semantics will recognize no special category of fictional sentences or fictional names. Everything will proceed just as for nonfictional regions of language.' See Sainsbury (2005, 202).

view is grounded on fictional name sentence utterances, not on fictional names. The article will be structured as such: Firstly, I will offer brief indications on the semantics of proper names. I will then focus on the truth of fictional name sentences. Secondly, I will propose a view on fictional names and utterances of fictional name sentences that is consistent with common intuition about truth. Finally I will offer some remarks on the identity of such names. It is important to know, before I move on, that my suggestion rely on the token-utterance distinction. And this distinction deserves a short explanation here. An utterance is a particular event, namely the use of a sentence by a speaker, occurring at a moment of time in a specific location. If one of these parameters is changed, a different utterance obtains. In contrast, a token of a sentence is not an event and it is not individuated by indices. There are different tokens of the sentence ‘London is a nice city’ when the latter is written on different post-its at different moment of time by the same speaker, or by different speakers on different post-its at the same moments of time. However, tokens of sentences are neither individuated by speakers, time and location, nor do they keep track of these indices.

## 2. Proper names

The Millian approach to names, according to which the only semantic value of a proper name is its bearer, is now dominant in the field. Following Kripke’s paradigm (Kripke 1980), ordinary proper names are generic names, or newly created names, assigned to specific objects located in space and time by speakers also located in space and time. The assignment procedure invoked by Kripke is vague and can remain so without any impact on the commonly accepted semantic core of names: names refer directly to the objects they are assigned to. Assigning a name to an individual establishes a convention allowing us to designate that individual with that name. Names have no reference fixing linguistic meaning, and carry no descriptive content exploiting features of the objects to which they refer. They are designators.<sup>3</sup> A definite description can fix the referent of a name, but it cannot determine its meaning.

---

<sup>3</sup> I assume, but I will not defend, the direct reference theory of names.

Moreover, according to Kripke, proper names refer to the same object in worlds where that object exists. They are rigid designators. The name 'Barack Obama' has been assigned to a person who became the President of the USA. It designates the same person in counterfactual situations in which he does not become President of the USA. Were the designata of the name determined by a definite description only, the name would not rigidly refer to an object. Different people can have the name 'Barack Obama.' We then say that they have different names, written or pronounced in the same way, because they are connected to different people. Alternatively, one can say that they have the same name, but that the latter is then connected to different people.<sup>4</sup>

Fictional names are, knowingly, introduced by authors for fictional, non-existent characters. Such names are not intended to refer to real objects, and they are not assigned any. Being fictional is not a contingent feature of such names, and neither is their lack of a referent (Kripke 2013). Fictional names differ from – even if they are not always distinguished from – empty names. They contrast with 'Vulcan' and 'Zeus,' which were not intended to be fictional names. The former was used to designate an object predicted, wrongly, to exist. The latter was used to designate a god, which, as it happens, does not exist. Most importantly, fictional names have no designata or referent. As they lack real designated referents, they cannot be individuated with them.<sup>5</sup> This being so, is there a difference between the name 'Sherlock Holmes' found in Doyle's books and the name 'Sherlock Holmes' used in a 21<sup>st</sup> century television series about a 21<sup>st</sup> century character? Are these two different names? If so, why are they different? Is it the same name? If so, why is it the same? If fictional names have no referents, these questions have no easy, intuitive answers.

---

<sup>4</sup> This is a controversial issue. Cf. Kaplan (1990).

<sup>5</sup> Fictitious objects, which lack dimensions, could be invoked here. I set them aside. I focus on fictional names rather than ontology.

### 3. Fictional names and truth

Consider

- (1) Sherlock Holmes is a detective.

If proper-name sentences or utterances determine singular propositions, that is, propositions containing the object referred to by the name, and if fictional names do not designate objects, then fictional-name sentences or utterances do not determine complete truth conditions or propositions. If Sherlock Holmes is a fictional, non-existent character, and if the name has no referent, then (1) determines a proposition containing an empty slot,  $\langle \_, \text{being-detective} \rangle$ , and it is neither true nor false.<sup>6</sup> If ‘Emma Bovary’ is a fictional name, then

- (2) Emma Bovary is a detective

determines the same empty-slot proposition. However, (1) and (2) are *prima facie* respectively true and false. The negation of (1) and (2), ‘it is false that Sherlock Holmes is a detective’ and ‘it is false that Emma Bovary is a detective,’ also determine gappy propositions, and they are neither true nor false. Nonetheless, the first one is intuitively false, and the second one is intuitively true. Finally, a speaker’s motivations for using (1) differ from their motivations for using (2) and vice-versa. The speaker of (1) wants to talk about Sherlock Holmes, and not about Emma Bovary; the speaker of (2) wants to talk about Emma Bovary, and not about Sherlock Holmes. Speakers do not choose randomly between (1) and (2) before making utterances. A view on fictional names should account for differences between (1) and (2).

If fictional names lack referents, then the sentence

- (3) According to Conan Doyle’s stories, Sherlock Holmes is a detective

---

<sup>6</sup> For a classic presentation of these problems, see Braun (1993; 2005).

determines a gappy proposition and it is not true, although most speakers feel that it is true. If ‘Emma Bovary’ is a fictional name, then (3) and (4) determine the same gappy proposition

- (4) According to Conan Doyle’s stories, Emma Bovary is a detective.

Most speakers take (3) and (4) to have different truth conditions, and take (3) to be true and (4) to be false. Fictional-name sentences question the common view that affirmative sentences express, or determine, complete true or false contents.<sup>7</sup> The truth and falsity of fictional name sentences thus deserve our attention.<sup>8</sup> Fictional names belong to the category of referring terms, that is to say: expressions introducing an element into the truth conditions of a sentence or an utterance. However, by design such names lack referents and do not introduce objects into truth conditions of sentences or utterances. A view on fictional names should account for their belonging to that category *and* for their lack of referent.

Following Braun (1993, 2005), ‘Sherlock Holmes’ has no referent, and a sentence like (1) expresses a gappy proposition, ⟨ , being a detective) (see also Everett 2013). He also contends that such propositions are false. Furthermore, their negation turns a false sentence into a true sentence. However, fiction sentences, or the propositions they express, are intuitively true: Sherlock Holmes is a detective (see Taylor 2000). Finally, if Braun is right, (1) and (2) express the same gappy proposition. That does not sound right.<sup>9</sup>

‘Sherlock Holmes’ is not transparent about its referent or having a referent, and it does not carry bells and whistles indicating whether it is an ordinary or a fictional name.<sup>10</sup> Not knowing much about the UK and 19th century literature, I think that Sherlock Holmes is a historical figure. You

---

<sup>7</sup> According to strict Millianism, singular propositions contain objects, not abstract or fictitious entities. Salmon (1998) introduces abstract objects and Predelli (2002), fictitious entities.

<sup>8</sup> For an examination of these issues, see Everett (2003).

<sup>9</sup> My criticism of Braun’s view on fictional names and gappy propositions is not intended as a criticism of Braun’s entire, complex view on such names.

<sup>10</sup> See also Kripke (2013, 30).

know that ‘Sherlock Holmes’ is a fictional name. We have a discussion and both of us talk about Holmes. At one point, you realize that I am misinformed or naive, and you tell me that ‘Sherlock Holmes’ is a fictional name. I then learn something important about that name. Still, I believe that my utterance of (1) is true and that your utterance of the same sentence is also true. *Prima facie*, intuitions on the truth of (1) or utterances of (1) do not depend on the existence of Holmes. What grounds these intuitions? You and I believe that we are talking about the same thing. Fictional names having no referent, ‘Sherlock Holmes’ in my mouth and in your mouth cannot ‘corefer’. What makes us think that we were talking about the same thing when using ‘Sherlock Holmes’?<sup>11</sup> Using ‘Sherlock Holmes,’ we were both assuming to be talking about the same character and had very similar ‘Holmes’ directed beliefs. At the same time, we had very different cognitive lives. I would have asked people if Holmes once met the Queen of England. You would not. Objects here seem to be irrelevant to our shared understanding of utterances of fiction sentences like (2). A semantic theory should explain and make sense of this.

#### 4. Pluri-propositionalism

I advocate a pluri-propositionalist framework in semantics (Perry 2012; Korta & Perry 2011). Following that perspective, utterances rather than tokens or sentences as type are in the foreground. Expressions as type have linguistic meaning, which is a rule determining content constituent for utterances of linguistic expressions. Linguistic meaning fixes the semantically determined content or the truth conditions of utterances.<sup>12</sup> Consider for example Brad Pitt’s utterance of

- (5) Angelina Jolie is an actress.

---

<sup>11</sup> Everett (2000) raises these questions concerning talking and thinking about the same thing when using fictional names. Friend (2014) examines them and suggests a solution that I will not explore here.

<sup>12</sup> For simplification, I do not make a difference between spoken and written utterances of a sentence or a name.

This utterance, **u**, is individuated by the speaker, Brad Pitt, the time, say April 27 2016, and the location, Los Angeles. ‘Angelina Jolie’ is an ordinary proper name. Following the now paradigmatic view on names, it has no linguistic meaning and a referent only. The name is conventionally associated with that referent. Being linguistically competent and relying on your knowledge of language only, including your knowledge of what a proper name is, you know that

Given that (5) is an English sentence, the utterance **u** of (5) is true if and only if the individual<sup>13</sup> that the convention exploited by **u** allows us to designate by ‘Angelina Jolie’ is an actress.

Call the content giving the semantically determined truth conditions of the utterance, without considering facts, the cognitive significance of the utterance (Perry 2012). A person accepting as true an utterance<sup>14</sup> of (5) will believe the cognitive significance of the utterance (Perry 1988). If the cognitive significance classifies a thinking episode, we can take the latter to be in the speaker’s head. That content contains the utterance **u** itself as a constituent and is, hence, reflexive with regards to the utterance. Notice also that the name itself is mentioned in the cognitive significance of the utterance. What follows ‘if and only if’, and precedes ‘is an actress’, captures an important aspect of the reference or designation relation. Yet, what you then understand does not call for the referent of the utterance of the name. The name is associated with a convention tying it to **ANGELINA JOLIE** herself. So, after taking into account facts required for fixing the designation of the indexical terms, including names,

Given that (5) is an English sentence, the utterance **u** of (5) is true if and only if **ANGELINA JOLIE** is an actress.

‘**ANGELINA JOLIE** *is an actress*’ is the designational content of the utterance, giving the conditions under which the utterance is true. The designational content of the utterance of (5) does not contain the utterance of

---

<sup>13</sup> An individual is whatever is designated by a proper name.

<sup>14</sup> Accepting as true an utterance is an attitude, and it does not imply that the utterance is true.

that sentence. The designated individual has that feature, being an actress, or not, whether or not there is an utterance, and whether or not that name has been assigned to that person. All utterances of (5) with that specific name associated with the same convention have the same designational content.

### 5. A suggestion

Suppose now that Conan Doyle makes an utterance of (1). His utterance is an event individuated by the speaker, Doyle, the time of the utterance, say, October 15, 1890, and its location, London. The sentence is in English, and all the examples are in the next pages. Thanks to linguistic competence only, including knowledge of proper names, when hearing or reading Doyle's utterance of (1), you understand that

The utterance **u** of (1) is true if and only if the individual that the convention exploited by **u** allows us to designate as 'Sherlock Holmes' is a detective.

You understand (1) using the same resources you used to understand Brad Pitt's utterance, even though Doyle is actually writing fiction. Utterances are particular events, which do not last and cannot be reproduced. They cannot be found in Doyle's books. As it goes, readers are not really interested in Doyle's utterance. Doyle's utterance left a token of a sentence on the page he was writing on. What he left on the piece of paper has been reproduced thousands of times. Readers are interested in the sentences he left on the paper he was writing on. Doyle's utterance is not a relevant event when considering fiction. It is rather the token that the utterance left that should be considered.<sup>15</sup> Call it the token **t**. To dispense with utterances, it can be argued that

---

<sup>15</sup> One can argue that reading a token creates an utterance. Still, the speaker of the utterance is not then the narrator. Such a view also relies on a non-standard notion of utterance.

The token **t** of (1) is true if and only if the individual that the convention exploited by **t** allows us to designate as ‘Sherlock Holmes’ is a detective.

There are different tokens of (1) in different copies of the book. Tokens, however, are not enough to account for fiction because they do not keep track of their source. Doyle, or any speaker, could have produced the relevant token. Readers know better, and they would disagree with the idea that Doyle is the speaker. Fiction sentences are arguably used by narrators telling stories, where the narrator is a fictive creature different from the author (Currie 1990; Kania 2005). The narrator of *Planet of the Apes* is not the author, Pierre Boulle, but except for the main part of the book, a creature whose nature is unknown (an ape?); the narrator of *The Name of the Rose* is not the author, Umberto Eco, but Adso of Melk. Doyle introduces Watson, and no one else, as the fictional narrator, or narrator for short, of (1) and the producer of the token. It is quite common to understand tokens in fiction as belonging to stories told by narrators, not authors, at a location and moment in time. We also want to have figures of speech (irony, for instance) and conversational implicatures (Grice 1989) in fiction, something tokens do not allow. The narrator can make irony and conversationally implicate something in producing a token. Fiction readers follow the fictive narrator and his or her use of sentences in fiction.

Let us say that in writing fiction, authors are making utterances leaving indexed tokens of sentences. An indexed token of a sentence is individuated by the sentence itself and an index containing the narrator, the time, and the location of the token. For example, ‘Sherlock Holmes is a detective’ (narrator of **t**, time of **t** and location of **t**). I call such an index a fictional index. The indices relied on capture what is specific to that fictional token, and they are the minimum needed to individuate the indexed sentence token. The indices are also echoed in the truth conditions of the fictional token. The idea of a fictional index generalises to any sentence token in fiction, including those not containing a fictional name, and it specifies a feature that is characteristic of fiction: the presence of a narrator. It is not required to mention the name of the author. Time can cover a short or a long period. Location can be big or small. No decision on that point has to be made here. We then have, for (1),

The indexed token **t** of (1) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ is a detective.

The narrator of **t**, the time of **t** and the location of **t** index the relevant token **t**. These truth conditions are token reflexive, since **t** is mentioned in them. With that in mind, let us focus on fictional names like ‘Sherlock Holmes’.

Ordinary names, like ‘London’, are used by fictional narrators and can also be found in fiction. These names then have their regular, associated conventions and the referent that comes with it. Fictional names are different. They are used by fictional narrators and have their source in fiction only. Such names also lack both meaning and associated conventions tying it to a space-time located referent. Nonetheless, readers individuate them, and can see a difference between ‘Sherlock Holmes,’ the name of the famous detective, and ‘Sherlock Holmes,’ the name of a British civil servant. I suggest that fictional names are individuated by a sequence of letters or phonemes and the fictional index of the sentence they originate from, composed of the fictional narrator, time and location of the token: ‘Sherlock Holmes’ (narrator of **t**, time of **t**, location of **t**).<sup>16</sup> A fictional narrator introduces a fictional name, at the location and time of the writing. Of course, such individuating indices are not part of the name token or an utterance of that name. I call ‘Sherlock Holmes’ (narrator of **t**, time of **t**, location of **t**) the indexed token of the fictional name. All fictional names have a fictional index. If a name does not, it is assumed to refer directly to its designata by default. Knowing that a name is fictional means knowing that it has a fictional index, and vice-versa. The contribution of an ordinary name to the cognitive significance of an utterance differs from the contribution of a fictional name to the cognitive significance of a fictional sentence: the former is utterance bound and it is not indexed, while the latter is token bound and it is indexed.

Different fictional sentences and names, having token reflexive indexes, can have different fictional indexes, where the narrator of **t**, location of **t** and time of **t**, is the narrator of **t’**, location of **t’** and time of **t’**. The

---

<sup>16</sup> Different sequences of letters and phonemes should be considered because in different languages (Russian, Japanese, French, and so on) names are written and pronounced in different ways. I put aside this issue.

narrator of each indexed token in *Remembrance of things past* is the same, as is its location in space and in time.

### 6. Back to truth conditions

Let us go back to the truth conditions of the indexed token of (1) and take into account the presence of the fictional name.

The indexed token **t** of (1) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ (narrator of **t**, time of **t**, location of **t**) is a detective.

The index of the fictional name is part of the truth conditions of the relevant indexed sentence token. The truth conditions of this token are not metalinguistic. They contain the name and its index only, and neither the token nor its truth conditions are about that name. Every component of the fictional index can be assigned a value, a fictional narrator, a stretch of time and a location, by considering the fiction where the indexed sentence token or fictional name is found. For example, the fictional narrator of the token can be Watson, the location of the token can be London, and the time of the token can be 1890. Two different indexed tokens can be assigned the same values. Value assignation makes it that tokens are not token reflexive anymore. For an indexed token of (1), and an indexed token of ‘Sherlock Holmes,’ we then have the following options.

Narrator of <b>t</b>	Time of <b>t</b>	Location of <b>t</b>
Watson	1890	London
Watson	1890	Location of <b>t</b>
Watson	Time of <b>t</b>	London
Watson	Time of <b>t</b>	Location of <b>t</b>
Narrator of <b>t</b>	1890	London
Narrator of <b>t</b>	1890	Location of <b>t</b>
Narrator of <b>t</b>	Time of <b>t</b>	London
Narrator of <b>t</b>	Time of <b>t</b>	Location of <b>t</b>

Some of these options are token reflexive. Others are not. One can ask whether or not all these options capture the name ‘Sherlock Holmes’ in different indexed tokens. I will come back to this point. What should be done with the fictional ‘Watson’ in the index? The fictional name ‘Watson’ should be assigned a fictional index, and it can also be assigned values containing the name of the narrator of *t*: ‘Watson’ (Watson, 1890, London). The location of the fictional sentence token usually differs from the location of the fiction. The location of the narrator writing (1), Watson, remains unknown, but London, also the location of the fiction, is a plausible place. The location of the fiction could also be Calcutta. Time is rarely emphasized. The time of the indexed token, the time of the writing, usually differs from the time in the fiction. For our example, 1890 will do. The time of the indexed sentence token is usually later than the time of the events described. The narrator in Doyle’s novel, Watson, in writing that Holmes was a smart student, writes that the bearer of this indexed token of ‘Sherlock Holmes’ is a smart student before the time of his own token. Time is needed to make sense of the tense of the verbs in fiction. For example, when Doyle writes that Holmes took a cab, Holmes takes a cab earlier than the time of the utterance, and also the time of the indexed token. Yet, it uses the time of the indexed token as a reference point: Holmes takes a cab before the time of the indexed token. Indexed sentence tokens are not designed to be read with changing values for the index – for an indexed token of (1) with assigned values, the narrator is Watson, the time is 1890 and the location is London. They are assigned a fixed narrator, space and time.

Fictional indices are commonly assigned values by readers, for example, for the indexed token sentence (1)

The individual that the convention exploited by the indexed token *t* of (1) allows us to designate by ‘Sherlock Holmes’ is a detective (Watson, 1890, London).

and for the name ‘Sherlock Holmes’,

‘Sherlock Holmes’ (Watson, 1890, London).

The assigned values are in parentheses. Some values are fictive, Watson, others are not, London. At least one value, the narrator, should be

fictive. Readers assign values to the fictional indices whenever possible. Knowledge of these values depends on more than linguistic competence only. It depends on knowledge of the fiction the sentence or fictional name originates from. Information on the fiction is needed to identify the values of indices. Indexed sentence tokens with assigned values give the truth conditions of a fictional sentence:

The indexed token **t** of (1) is true if and only the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ (Watson, 1890, London) is a detective.

I will always mention the fictional index of the fictional names in truth conditions of indexed tokens.

Consider the indexed token **t** of ‘Sherlock Holmes is a detective,’ which comprises the narrator of **t**, the location of **t** and the time of **t**. The fictional index plays a pivotal role in understanding fiction. All different indexed tokens of (1), in thousands of copies of Conan Doyle’s books, have the same fictional fixed index, with the narrator of **t**, time and location of **t** all assigned the same values: Watson, London and 1890. Location and time in the novel are relative to these values. Fictional names occurring in these tokens have the same fictional index and the same assigned values. Interestingly, in keeping with our table, we can also have different assigned values for that sentence and that name, for instance:

Watson	1899	London
Miss Hudson	1999	Calcutta
Watson	2234	Boston

There appears to be no constraints on the values of the index – the narrator of a token could be named ‘Miss Hudson’ or ‘Watson,’ (s)he could be located in 1999 or 2234, in London or Boston. We can use a fictional indexed token **t** of (1), Watson as the fictional narrator of **t** and change the time of **t** to the year 2234, and the location of **t** to Boston. Introducing a distinction between indexed tokens and the indexed-token-assigned values, the name of the narrator or time for instance, opens room for altering these

values at will to obtain different values for fictional indexes, and different fictional names.<sup>17</sup> I have already mentioned that this raise issues I will come back to with respect to name identity in section 8.

We have for (1): ‘Sherlock Holmes is a detective’ (narrator of **t**, the location of **t** and the time of **t**) — and for the fictional name ‘Sherlock Holmes’ (narrator of **t**, the location of **t** and the time of **t**). I call it the cognitive significance of the indexed token sentence and the cognitive significance of the fictional name. Since there are different tokens of the sentence and the name, indexed tokens of (1) and ‘Sherlock Holmes’, in different copies of Doyle’s book, these different tokens have different cognitive significance. We also have, with assigned values, for the indexed token **t** — ‘Sherlock Holmes is a detective’ (Watson, London and 1890) — and ‘Sherlock Holmes’ (Watson, London and 1890).<sup>18</sup> Once again, different tokens of the sentence and the name occur in different copies of the book. However, the fictional index being assigned fixed values is stable. I call it the informed cognitive significance of the indexed token sentence and the informed cognitive significance of the fictional name. It is informed because it takes into account specific, story grounded factors. All indexed tokens of (1) and ‘Sherlock Holmes’, with assigned values, in different copies of Doyle’s book, have the same informed cognitive significance because there are the same sentences and name tokens. The index of a sentence-indexed token, or a fictional name, and its assigned values, as well as its informed cognitive significance, is fixed and it is not designed to individuate a real world entity — a real world token or name. Let us say that an indexed token of a fictional sentence, or fictional name, is a sentence, or name, with a token reflexive index, which is standardly assigned values.

---

<sup>17</sup> In a fictional story, there can be various fictive narrators — in Akutagawa’s *In the Bush*, or in Inoue’s *The Hunting Gun* for instance. It complicates the issues but does not fundamentally change the basic problem. In cases where there are many different fictive narrators, it is possible that the indexed tokens are not coherent in that they cannot all be simultaneously accepted as true (Akutagawa’s *In the bush*). Of course, different fictive narrators can individuate a name — Watson, or Holmes’ landlady if the latter was writing about Holmes.

<sup>18</sup> It captures Kripke’s idea (2013, Lecture 2) that fictional names are ‘pretended names’. They are introduced in a fiction sentence.

Unfortunately, fictional indexed sentence token as well as fictional names are not always specified as finely as we would like them to be. Sometimes the fictive narrator is not named, the specific location is rarely indicated, if indicated at all, and the time is commonly left in the dark. All the reader can exploit is then the narrator of *t*, time of *t*, location of *t*. Such tokens of sentences have token reflexive truth conditions only. This is the nature of fiction, and it may complicate the task of the reader. The narrator is very important. For simplification purposes, one keeps the narrator of the indexed sentence token or fictional name, and its assigned value only, say, Watson, if there is such a value. It can then be argued that all indexed sentence tokens in Conan Doyle's books for instance have the same narrator and that the same truth conditions of the indexed token apply.

Accepting as true fiction indexed tokens containing fictional names, 'Sherlock Holmes is a detective,' is a nice way to describe the reader's attitude toward fiction. The reason for acceptance as true can vary wildly – because it is in Doyle's book, because it is plausible given what is known about Holmes, and so on – as do reasons for rejection as false.<sup>19</sup> However, such reasons are not semantically relevant. Tokens accepted as true also give possible belief contents. The cognitive lives of people who accept as true indexed tokens of (1), with the same assigned values, are identical even if Holmes does not exist: they all believe that Sherlock Holmes is a detective. This belief has truth conditions containing the fictional name itself.<sup>20</sup> If such names are objects, then the content or truth conditions of fictional name utterances or indexed tokens, and the belief these utterances are used to express, do contain objects.

The fictional index of 'Sherlock Holmes' in (1) and in

(6) Sherlock Holmes went downstairs

are the same. The name does not refer to an object and hence these tokens of the fictional name cannot co-refer. But the truth conditions of the

---

<sup>19</sup> I do not need the notion of fidelity – Sainsbury (2005) – except to mention the reasons for acceptance as true.

<sup>20</sup> I take belief as an example of an attitude commonly used in the relevant literature. I will not examine other attitudes. My paper focuses on contents, not on attitudes. I let the reader examine implications of my view on other attitudes.

fictional tokens of (1) and (6) cohere if the fictional names are identical, that is, if they have the same fictional index assigned the same values. It is commonly assumed that they are. ‘Sherlock Holmes’ is very finely individuated every time.

You and I accept as true that ‘Sherlock Holmes’ in Watson’s indexed token **t** of (1) is ‘Sherlock Holmes’ in my utterance **u’** of (1), a metafictional use of the sentence, in a conversation about the book. It is the same fictional name. So,

The utterance **u’** of (1) is true if and only if the individual that the convention exploited by **u’** allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective.

The narrator’s indexed token of (1), and my utterance of (1), have something in common. They share part of their truth conditions: the token and my utterance are true if and only if *the individual that the convention exploited by the indexed token **t**, and my utterance **u’**, allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective*. It is also the ‘Sherlock Holmes’ in your utterance **u’’** of (1) in a conversation about the same books.

The utterance **u’’** of (1) is true if and only if the individual that the convention exploited by **u’’** allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective.

My utterance and your utterance share part of their truth conditions: they are true if and only if *the object designated by both utterances of (1) allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective*. As is the case with utterances of ‘Angelina Jolie is an actress,’ different utterances of (1) have different cognitive significance. The fictional index, and the values that individuate the relevant name, is not in the sentence (1) or in utterances of (1). Rather, it is made explicit in the truth conditions of the indexed tokens and utterances of (1). Such tokens and utterances are not assessed as true given facts. For such tokens and utterances, we are solely considering the cognitive significance of tokens and utterances of fiction sentences. Moreover, it is arguable that such truth conditions capture what competent speakers understand in the use of ‘Sherlock

Holmes.’ No entity is needed to account for our intuitions about the truth of tokens of fiction sentences, including intuitions on identity tokens of fiction sentences or utterances of fiction sentences like ‘Sherlock Holmes is Sherlock Holmes’: the individual that the convention exploited by **t** (or **u**) allows us to designate as ‘Sherlock Holmes’ (Watson, London, 1890) is the individual that the convention exploited by **t** (or **u**) allows us to designate as ‘Sherlock Holmes’ (Watson, London, 1890).<sup>21</sup> However, in conversation, an utterance of ‘Sherlock Holmes is Sherlock Holmes’, or any similar identity sentence utterance, needs clarification before being accepted as true, and cannot be assessed as true, or false, given facts.

Once introduced, fictional names can leave fiction and be used in utterances. I can say ‘Sherlock Holmes never went to Chile’, and my utterance is true if and only if the individual that the convention exploited by the utterance allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) never went to Chile. My utterance can be accepted as true, or rejected as false. It is not true or false. There is room for much discussion here. Korta and Perry (2011, 89) would say that my utterance is accurate or not. A 21<sup>st</sup> century teenager can wonder: would Sherlock Holmes (Watson, London, 1890) use the internet? Any answer to that question is, and remains, very speculative.

Accepting as true the relevant token of (1), with values assigned to both name and sentence in the truth conditions, I believe that Sherlock Holmes is a detective. Accepting as true the relevant token of (1) with the same assigned values to the token and the name in the truth conditions, you also believe that Sherlock Holmes is a detective. We believe the same thing made explicit by the informed cognitive significance of the indexed token. Holmes does not have to exist to share the same belief about Holmes. Can another individual in Doyle’s novels be named ‘Sherlock Holmes’ without Doyle, or the narrator, telling the reader? There is no answer to that question, and the only ‘Sherlock Holmes’ we have is the detective named

---

<sup>21</sup> For ‘Dr. Jekyll is Mr. Hyde’, we obtain, ‘Dr. Jekyll (narrator of **t**, time of **t**, location of **t**) is Mr. Hyde (narrator of **t**, time of **t**, location of **t**)’. The truth conditions of a token of this identity sentence are ‘the individual that the convention exploited by the token **t** allows us to designate by ‘Dr. Jekyll’ (narrator of **t**, time of **t**, location of **t**) is the individual that the convention exploited by the token **t** allows us to designate by ‘Mr. Hyde’ (narrator of **t**, time of **t**, location of **t**)’.

‘Sherlock Holmes’ mentioned by Doyle and Watson. The fictional index and its assigned values are sufficient to individuate it.

## 7. Indexed tokens and fictional names

Consider a token of (1) in a copy of a Doyle’s novel. As we have seen, competent and informed speakers understand that

The indexed token **t** of (1) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective.

Suppose that we have a discussion about the novel and that I say (1). Which ‘Sherlock Holmes’ am I using? A fictional name, individuated by a fictional index, or a real name individuated by its bearer? What I said needs precision before being assigned truth conditions. I was talking about the novel, used a specific fictional ‘Sherlock Holmes,’ and said that the bearer of ‘Sherlock Holmes’ (Watson, London, 1890) is a detective – where the index individuates the used name. The truth conditions of my metafictional utterance **u** are given by

The utterance **u** of (1) is true if and only if the individual that the convention exploited by **u** allows us to designate by ‘Sherlock Holmes’ (Watson, London, 1890) is a detective.

My utterance is true if and only if the bearer of the fictional name is a detective. The truth conditions of the indexed token of (1) are easy to obtain. If the relevant name is properly individuated, there is no major difference between a fictive and a metafictional use of fictional name sentences. Except for reflexivity to the token and to the utterance, they have the same truth conditions and are, hence, in a sense equivalent. Such reflexivity echoes the fact that Watson is the fictive narrator of the token, and I am the actual speaker of the utterance. Now, consider my transfictional utterance **u** of

(7) Sherlock Holmes is a better detective than Hercule Poirot.

My utterance has truth conditions

The utterance **u** of (7) is true if and only if the individual that the convention exploited by **u** allows us to designate by ‘Sherlock Holmes’ (Watson, 1890, London) is a better detective than the individual that the convention exploited by **u** allows us to designate by ‘Hercule Poirot’ (narrator of **t**, time of **t**, location of **t**).

Suppose that you accept as true my utterance. It is understood that the elements indexing the names are different. Comparing the qualities of Holmes and Poirot is not grounded on facts and it remains a very speculative activity. The important philosophical difference between the fictive, metafictive and transfictive use of fictional names impacts the cognitive significance of indexed tokens and utterances only.

Let us go back to (2), ‘Emma Bovary is a detective,’ which is an English sentence and where ‘Emma Bovary’ is a fictional name. I will not consider the assignment of values to the index. A fictive token of this sentence has truth conditions

The indexed token **t** of (2) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Emma Bovary’ (narrator of **t**, time of **t**, location of **t**) is a detective.

The reader will not accept as true the relevant token of that sentence. He can reject as false my utterance of (2) for the same reasons

The utterance **u** of (2) is true if and only if the individual that the convention exploited by **u** allows us to designate by ‘Emma Bovary’ (narrator of **t**, time of **t**, location of **t**) is a detective.

The same goes for (4), ‘According to Conan Doyle’s stories, Emma Bovary is a detective.’

The utterance **u** of (4) is true if and only if according to Conan Doyle’s stories, the individual that the convention exploited by **u** allows us to designate by ‘Emma Bovary’ (narrator of **t**, time of **t**, location of **t**) is a detective.

‘Conan Doyle’ is a real name and, given facts about the utterance, it can be assigned a designata: according to **CONAN DOYLE**’s stories, the person that the convention exploited by **u** allows us to designate by ‘Emma Bovary’ (narrator of **t**, time of **t**, location of **t**) is a detective. It is also an utterance that will be rejected as false.

The semantically determined content and cognitive significance<sup>22</sup> of fiction tokens containing fictional names have no echo in official truth conditions or designational content, that is, the truth conditions that are obtained after determining the relevant, referred-to objects once facts about the utterance are taken into account (Perry 2012). Such tokens cannot be truth assessed. For fictional name sentence utterances or tokens, there is no designational content. In contrast to ordinary proper name sentence utterances, fictional name utterances and indexed tokens have utterance, or indexed token, dependent truth conditions only.<sup>23</sup> In this sense, Sherlock Holmes, the detective we all like, could not have existed without Doyle’s books. He is a creature of fiction. There is no content like ‘**SHERLOCK HOLMES** is a detective’ with Holmes himself as a constituent. We could introduce truth conditions with an empty slot. However, if the name is a fictional name, which by design has no referent, then there is no empty slot official content by design.<sup>24</sup> Only cognitive significance or informed cognitive significance is relevant to the author and reader. Some call utterances that fail by design to have designational content ‘pretence’ or ‘make-believe’ because they contain fictional names. I simply call it ‘fiction writing.’ Fiction readers consider the narrator’s story, not facts, because there are no facts. My view on fictional names also captures an aspect of fictional indexed tokens, like indexed tokens of

(8) Sherlock Holmes lives in London.

Following our model, if ‘London’ is an ordinary name,

---

<sup>22</sup> Or informed cognitive significance.

<sup>23</sup> In that respect, my view on the semantics of fictional names and fictional name sentences strongly differs from Braun’s.

<sup>24</sup> In that respect, my view contrasts with Braun’s and Taylor’s.

The indexed token **t** of (8) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ (Watson, 1890, London) lives in the place the convention exploited by the narrator of **t** allows us to designate by ‘London.’

Given the facts, we obtain

The indexed token **t** of (8) is true if and only if the individual that the convention exploited by **t** allows us to designate by ‘Sherlock Holmes’ (Watson, 1890, London) lives in **LONDON**

**LONDON** is the city itself. The truth conditions of fiction tokens that are partly about reality – or that have mixed contents – are not captured by the idea that fiction is pretence or make-believe because these truth conditions are too fine grained.

We need cognitive significance only or, better, informed cognitive significance, to understand fictional indexed tokens, and fictional names, and to follow stories. Of course, we can believe such contents. And it is fun. Our belief about fiction has truth conditions, something like *the individual that the convention exploited by ‘Sherlock Holmes’ (Watson, 1890, London) is a detective*. But we do not then believe assessed-as-true content containing fictional objects. Fiction does not require designational content and truth assessment, quite the opposite. In that respect, writers do not pretend that their utterances are true given features of utterances. They can be described as just supposing that the tokens they left are indexed and can be accepted as true. The notion of truth involved in the intuitions mentioned at the beginning of this paper plausibly fits acceptance as true. Let us go back to a token of (1) in Doyle’s novel. It is not true, but it is accepted as true. You can believe, or not, in the existence of Holmes and still accept the token of (1) as true. In our (earlier) discussion, you took ‘Sherlock Holmes’ to be a fictional name and knew that there was no designational content that was assigned a truth-value. I took ‘Sherlock Holmes’ to be a real name and thought that there was a designational content that was assigned a truth-value. This is the cognitive difference between you and me. Knowing what a name is an important aspect of our knowledge of language; knowing that some names are fictional names is also a major aspect of our linguistic competence.

Let us go back to the truth conditions of the indexed token of (1). If the name is a properly individuated fictional name, one cannot substitute a different fictional name, ‘Emma Bovary’ for instance, in the token and accept it as true. In any case, any token of ‘Emma Bovary is a detective’ has a different cognitive significance from a token of (1). Consider (1) in a television series, taking place in the 21<sup>st</sup> century. The name differs from the name in Doyle’s books – because the time of the fictional name is different – and cannot be substituted for the first version while preserving acceptability. The intuition that it can do so needs arguments. I suspect that accepting free substitution of ‘Sherlock Holmes’ depends on focusing on the sequence of letters or phonemes only and disregarding a more fine-grained individuation condition for the name.

### 8. Fictional names as complex objects

Kripke (2013, 78) evokes issues raised by characters appearing in different fictional stories. The same problems show up for names themselves. Is the name ‘Sherlock Holmes’ found in Conan Doyle’s novels the same as the name ‘Sherlock Holmes’ heard in a film or a television series? Various fictional names can be introduced, by different fictive narrators in different fictions. Ordinary names are simple objects individuated by a sequence of letter or sounds and the referent they are assigned to. Fictional names are complex objects with no assigned referent. Their fictional index, as well as the word itself, can be modified. The same sequence of letters, ‘Holmes’, can be individuated by fictional indexes assigned different values: ‘Sherlock Holmes’ (Watson, London, 1890), ‘Sherlock Holmes’ (Narrator of *t*, Calcutta, 1918), ‘Sherlock Holmes’ (narrator of *t*, Moscow, 1954), and so on. Suppose that you use ‘Sherlock Holmes.’ You plausibly have a specific, well-identified name in mind, be it from Doyle or from 1930s movies. And maybe you do not. If you do not, it *prima facie* does not always really matter in communication. The source can be Conan Doyle’s book or a TV series, and the names can be different. Unless details are needed, there will be no question concerning the specific name used. Due to lack of a specific index for a used fictional name, fictional name sentence utterances very often have no clear truth conditions, lack determined cognitive significance and they cannot always clearly be accepted as true, or rejected as false.

Conversations involving such names frequently show a high degree of indeterminacy. There is room for speculation on name identity and identification in communication that I will not engage in here. Nonetheless, a view on fictional names that does not address the identity and identification of fictional names, and that fails to take into account the many different ‘Sherlock Holmes,’ does not tell the whole story about such names. In so far as the complex object can be altered, we have a manifold of potential fictional names originating from the initial fictional token of ‘Sherlock Holmes’. We have a problem with the identity of names in stories. In this respect, such names contrast with names like ‘Zeus’ and ‘Vulcan.’ I will not explore that metaphysical issue here.

## 9. Conclusion

Speakers who are semantically competent with names know the difference between ordinary names and fictional names even if they are unable to say to which category a specific name belongs. The cognitive role of a specific name will differ accordingly. Reasoning with an ordinary name sentence or utterance and reasoning with a fictional name sentence or utterance does not have the same scope. My view has no impact on the conceptions of fictional objects; for example, Thomasson’s sophisticated picture of these objects. Reading fiction, we rely on accepted as true indexed tokens and use imagination about fictional objects and characters. However, this is of no semantic relevance. In this respect, my suggestion separates semantics from the ontology connected to fictional names. It binds the category of fictional objects to fiction, not to the semantics of fictional names. Creatures of fiction are motivated neither by semantics, nor by referential issues. Attributing a referential role to fictional names and seeing them as designating objects only help readers to engage with fiction. The point of my paper is that fictional names have a cognitive aspect only, and no referential role. These features, and the role of the reader in identifying names and characters, deserve further exploration.

## References

- ADAMS, F., FULLER, G. & STECKER, R. (1997): The Semantics of Fictional Names. *Pacific Philosophical Quarterly* 78, 128-148.
- AKUTAGAWA, R. (2007): *Rashomon and Seventeen Other Stories*. Trans. Jay Rubin. London: Penguin Classics.
- BOULLE, P. (1964): *Planet of the Apes*. New York: Signet books.
- BRAUN, D. (1993): Empty Names. *Noûs* 27, 4, 449-469.
- BRAUN, D. (2005): Empty Names, Fictional Names, Mythical Names. *Noûs* 39, No. 4, 596-631.
- BROCK, S. (2004): The Ubiquitous Problem of Empty Names. *Journal of Philosophy* 101, No. 6, 277-298.
- CURRIE, G. (1990): *The Nature of Fiction*. Cambridge: Cambridge University Press.
- ECO, U. (1983): *The Name of the Rose*. San Diego: Harcourt.
- EVERETT, A. (2003): Empty Names and 'Gappy' Propositions. *Philosophical Studies* 116, 1-36.
- EVERETT, A. (2000): Referentialism and Empty Names. In Everett, A. & Hofweber, T. (eds.): *Empty Names, Fiction and the Puzzles of Non-Existence*. Stanford: CSLI Publications, 37-60.
- FRIEND, S. (2014): Notions of Nothing. In García-Carpintero, M. & Martí, G. (eds.): *Empty Representations*. Oxford: Oxford University Press, 307-332.
- GRICE, H.P. (1989): Logic and Conversation. In: Grice, H.P.: *Studies in the Ways of Word*. Harvard University Press, 22-40.
- INOUE, Y. (1977): *The Hunting Gun*. North Clarendon: Tuttle Publishing.
- KANIA, A. (2005): Against the Ubiquity of Fictional Narrators. *The Journal of Aesthetics and Art Criticism* 63, No. 1, 47-54.
- KAPLAN, D. (1990): Words. *Proceeding of the Aristotelian Society*, 93-119.
- KORTA, K. & PERRY, J. (2011): *Critical Pragmatics*. Cambridge: Cambridge University Press.
- KRIPKE, S. (1980): *Naming and Necessity*. Cambridge (Mass.): Harvard University Press.
- KRIPKE, S. (2013): *Reference and Existence*. Oxford: Oxford University Press.
- MURDAY, B. (2011): Two-Dimensionalism and Fictional Names. In: Lihoreau, F. L. (ed.): *Truth in Fiction*. Berlin: Ontos Verlag.
- PERRY, J. (1988): Cognitive Significance and the New Theory of Reference. *Noûs* 12, 1-18.
- PERRY, J. (2012): *Reference and Reflexivity*. Second edition. Stanford: CSLI Publications.
- PREDELLI, S. (2002): 'Holmes' and Holmes – A Millian Analysis of Names from Fiction. *Dialectica* 56, No. 3, 261-279.

- REIMER, M. (2001): The Problem of Empty Names. *Australasian Journal of Philosophy* 79, No. 4, 491-506.
- SAINSBURY, R.M. (2010): *Fiction and Fictionalism*. London: Routledge.
- SAINSBURY, R.M. (2014): *Reference without Referents*. Oxford: Oxford University Press.
- SALMON, N. (1998): Nonexistence. *Noûs* 32, No. 3, 277-319.
- TAYLOR, K. (2000): Emptiness without Compromise. In: Everett, A. & Hofweber, T. (eds.): *Empty Names, Fiction and the Puzzle of Non-existence*. Stanford: CSLI Publications, 17-36.
- THOMASSON, A. (1999): *Fiction and Metaphysics*. Cambridge: Cambridge University Press.
- VOLTOLINI, A. (2006): *How Ficta Follow Fiction. A Syncretistic Account of Fictional Entities*. Dordrecht: Springer.

## *Grundprobleme*, or Popper Meets Kant

ONDŘEJ BÍBA – JITKA PAITLOVÁ<sup>1</sup>

ABSTRACT: First part of the text presents a historical excursion searching for the genesis of Popper's philosophical views in the interwar Vienna. It analyzes the actual writing process and circumstances that surrounded Popper's work on *Die beiden Grundprobleme der Erkenntnistheorie*. The aim of this section is to evaluate Popper's reception and intellectual self-development through the denial of logical positivism. The second "internalist" segment of this article further examines the *Grundprobleme* itself through the analysis of Popper's specific interpretation of Kant's transcendental idealism. We will confront Seubert's claim that through *Die beiden Grundprobleme der Erkenntnistheorie* Popper definitely and knowingly accepts Kant's stance. We show that even though Popper adopted Kant's transcendental method of questioning, he had later criticized certain aspects of Kant's transcendental method. As a result, Popper establishes the so called genetic apriorism, which dwells on his own version of the deductive psychology of knowledge.

KEYWORDS: Epistemology – interwar Vienna – *Immanuel Kant* – Karl Popper – logical positivism – neopositivism

---

<sup>1</sup> Received: 28 September 2017 / Accepted: 28 december 2017

✉ Ondřej Bíba

Department of Philosophy  
Faculty of Philosophy and Arts, The University of West Bohemia  
Sedláčkova 19, 306 14 Plzeň  
e-mail: biba@kfi.zcu.cz

✉ Jitka Paitlová

Department of Philosophy  
Faculty of Philosophy and Arts, The University of West Bohemia  
Sedláčkova 19, 306 14 Plzeň  
e-mail: paitlova@kfi.zcu.cz

## 1. Externalist perspective: genesis of the views of Karl Popper

Despite the increasingly widespread perception of Popper as an intellectual solitaire, we present a different picture of the thinker. This insight could be seen while we look into an early stage of Popper's thought-forming process. Moyal (1997, 50) notes that "it does not hurt to take seriously that Popper was self-taught in philosophy – he was perhaps the last great autodidact at the history of philosophy." That is also why it is not easy to find one decisive element in Popper's intellectual development that would mark the turn from his philosophically mature thinking towards a "recognizably Popperian Popper" (Naraniecki 2014, 45) whom we know today. Similarly, it is difficult to identify a particular philosopher that would help us to better understand the genesis of Popper's ideas through the prism of its work.

All scholars who study Popper's early work, however, agree that an essential key to decrypt his intellectual development lies in a work *Die beiden Grundprobleme der Erkenntnistheorie*.<sup>2</sup>

### 1.1. Two problems surrounding the publication of *Grundprobleme*

Until the publication of *The Logic of Scientific Discovery* in 1935 the only available bits of information about Popper's ideas were available either through his personal contact with Viennese philosophers, or through comments published in the magazine of logical positivism *Erkenntnis*. Popper's methodological breakthrough reached on the pages of *Grundprobleme* was thus largely unknown to broader philosophical circles.

In December 1932 Carnap publishes a brief report on Popper's manuscript in *Erkenntnis* in the article *Über Protokollsätze* (Carnap 1932). He supports in it Popper's deductive model of theory testing, while expressing the hope that *Grundprobleme* is soon going to be published. Nothing like that does happen though, and a discussion surrounding Popper's work takes place only in a narrow circle of "insiders". Hacohen describes these events in detail while he also mentions a conflict about intellectual priority that

---

<sup>2</sup> Let us mention that the German title of his work (*Die beiden Grundprobleme der Erkenntnistheorie*) is an indirect allusion to Schopenhauer's book *Die beiden Grundprobleme der Ethik*.

arose between Popper and Neurath in the case of non-foundationalism: “His originality and independence were at stake. [...] He asked Carnap to emphasize his independence, and Carnap agreed” (Hacohen 2002, 218). One of the tangible results of Carnap’s assistance is that, in 1933, *Erkenntnis* publishes a most succinct excerpt of the whole book, a summarizing two page report (Popper 1933, 426–427). Carnap’s gesture, however, failed to satisfy Popper.

*Grundprobleme* did not make it to print (in the originally intended seven hundred pages long release) any sooner than in 1978 through its German edition. Many misconceptions associated with Popper’s work may thus have its basis in the lack of contextual understanding of the specific problem situation of the Viennese interwar philosophy. The English edition of *Grundprobleme* occurs in 2008, long after Popper’s death. On a more positive note, its belated publication helped to bring a new wave of interest in Popper’s work. The English version of the book now represents a strong impulse to reopen discussion on topic of an early Popperian philosophy.

Popper was well aware that, in order to build an international reputation, he must hold a close relationship with the most discussed philosophical movement of his epoch, the logical positivism. Popper himself never was a member of this philosophical movement – at least in a classic conception of this school of thought, broadly defined by its identification with the Vienna Circle and with attending Schlick’s seminars.<sup>3</sup> Despite the fact that both Popper and logical positivists discussed the same philosophical questions, each of them ended up with a completely different outcome.

On a philosophical level, Popper is an anti-inductivist, anti-verificationist and he is also against the cumulative approach to science. Beside that he does not curse metaphysics for its worthlessness. Despite these differences regarding scientific method, some members of the Vienna Circle (as well as other philosophical contemporaries) regarded Popper as an heir to Viennese analytic tradition.

---

<sup>3</sup> Popper did not receive a personal invitation to visit Schlick’s seminars (see Hacohen 2002, 188–190).

### 1.2. *The myth of Popper the neo-positivist*

We claim that an origin of the *myth of Popper the neo-positivist* had been affected by three major events. First being Popper's forced immigration to New Zealand and a resulting separation from "philosophical venues" of Europe; a fifty years delay in *Grundprobleme's* publication constitutes another important factor while the last impact was caused by another postponement, this time a twenty-five years long waiting for the English translation of the *Logic of Scientific Discovery*.<sup>4</sup> It is these actualities that collectively brought up a rift in the conceptualization of Popper by the upcoming generation of philosophers and gave an impulse to his inclusion in the neo-positivist school of thought.

On the other side, we should not completely ignore that some form of cooperation undoubtedly prevailed between Popper and members of the Vienna Circle. Despite their criticism, at least some of the members of the inner circle proved to be intellectually open towards Popper's ideas. Especially Carnap, a prominent philosopher in Viennese analytic tradition, often assisted Popper on his path towards broader recognition. Naraniecki (2000, 514) points out a following fact:

Through Popper's private correspondence we get a remarkably different picture of his relationship to the members of the Circle. As the Vienna Circle constituted the pre-eminent philosophical school in Austria, Popper was invariably drawn towards them. He formed life-long intellectual relationships as well as close friendships with members of this group.

However not all members of the circle were supportive to a famously-eccentric Popper.

According to Hacohen, when Popper first met Moritz Schlick, it was a complete disaster:

Popper gave a lecture to the Gomperz circle at which Schlick, Carnap, Kraft, and other circle members were present. Popper was nervous, and

---

<sup>4</sup> The specifics of Popper's stay at New Zealand can of course be also applied to the above mentioned problem that presents a mistaken identification of Popper as a logical positivist.

this brought out the worst in him. He targeted Wittgenstein for criticism and, during the discussion, confronted Schlick. Wittgenstein, he told Schlick, was a dogmatist. Like the Catholic Church, he prohibited discussion of philosophical problems that he could not solve, declaring them nonexistent. Schlick left angrily in the middle of the discussion. He told Carnap later that Popper had misinterpreted Wittgenstein, and there was nothing new in his paper. (Hacohen 2002, 219)

Popper thus finds himself standing on the edge of an abyss, as his road to academic success leads both through the denial of fundamental theses of neo-positivism as well as through the close contact with members of the Vienna Circle – the proponents of the doctrine whose principles Popper so strongly criticized.

In the first part of *Grundprobleme* Popper expresses some *comments on the relationship of the book to the current theory of knowledge*:

On account of its formulation of the problem and its method, which is oriented towards natural sciences, the book is close to modern ('logically' oriented) '*positivism*' (Bertrand Russell, Moritz Schlick, Philipp Frank, Rudolf Carnap, Hans Reichenbach and Ludwig Wittgenstein). Yet for this very reason, it devotes its most detailed *criticism* to this movement, and it attempts to expose the 'fundamental contradiction of positivism' through which this philosophy fails. (Popper 2008, xxxiii-xxxiv)

Popper is thus a victim of a following dilemma: he does not know whether the philosophical destruction of neo-positivism will not endanger his current ties to many of its supporters. Hence, it is their positive evaluation that he will soon need for thanks to political development in Europe. Let us now focus in more detail on the critique of logical positivism present on the pages of *Grundprobleme*.

### 1.3. *The centerpiece problems of Grundprobleme and their current reception*

Through the philosophically challenging text of *Grundprobleme* Popper presents his ideas concerning neo-positivism and its representatives.

Two problems which Popper deemed crucial with respect to a possibility of any further development in epistemology were the problem of induction and the problem of demarcation. The first issue is largely associated with questions that have been already sketched by Hume, namely to what extent can we legitimately use our experience to predict future events. The second one can be traced back to Kant and reflects a problem of demarcation of science from metaphysics and pseudo-science.

Hacohen (2002, 196) offers an explanation that these two fundamental problems originally started as one, the problem of induction. “Popper intended to follow Gomperz’s model of a ‘dialectical critique’: a critique exposing internal contradictions in the positivist views of induction, and leading to his own solution.” It is the final part of Hacohen’s comment which highlights the fundamental problem of Popper’s manuscript. The competitive “disadvantage” of *Grundprobleme* is hidden in the very length (and exceptionally difficult readability) of the text itself, which could not be, as we mentioned above, published for its bulkiness.

Even though Popper was never a neo-positivist, he still shared many neo-Kantian tendencies with members of the Vienna Circle. In the early thirties of the 20<sup>th</sup> century, Kant’s work enjoyed a great popularity in both philosophical and scientific circles. At the same time, many of Kant’s conclusions did not seem nearly as obsolete as in 1979. We mention these circumstances only to put further emphasis on the fact that thanks to a failed attempt to publish *Grundprobleme* in 1933, the work lost a great part of its potential popularity.

For its overly sophisticated argumentative structure *Grundprobleme* cannot speak clearly to many of its readers by the time of their publication. Moreover, Popper does not enjoy much popularity today in analytic philosophy, but it is paradoxically exactly in these circles where the complicated terminology of *Grundprobleme* could be best understood. Furthermore, the main issues that Popper addressed in *Grundprobleme* – Kant-Fries’ problem, Dingler’s conventionalism, as well as critiques of the philosophies of Selz and Buhler – were already losing on popularity back in 1930’s, but even worse, fifty years later those discussion lost almost all importance whatsoever to all but the small group of historiographers of the philosophy of science.

## 2. Internalist perspective: Popper's reinterpretation of Kant

This part of the study introduces Popper's early theory of knowledge as was constituted by Popper in *Grundprobleme*. We focus on the internalist analysis of classical philosophical approaches (classical rationalism and empiricism as well as Kant's transcendental idealism), that could be traced as sources of influence for young Popper. We aim to show that Popper approached the philosophies in question in a very critical manner.

It is almost unbelievable how high levelled are the philosophical arguments of barely thirty years old Popper. As his explicit disagreement with neo-positivism is usually well known thanks to *The Logic of Scientific Discovery*, we will hereby focus on another subject: his critical approach towards Kant. Seubert (2016, 8) argues that "Popper definitely and deliberately reveals the acceptance of Kant's position." We will show, however, that this acceptance is indeed critical and brings Kant's apriorism to its new consequences.

### 2.1. *The synthesis of rationalist and empiricist elements*

Popper refers to his theory of knowledge as the *deductivist-empiricism*. He bases his theory on two fundamental assumptions:

1. Deductivism: an assumption that all scientific methods of justification are based on strictly logical deduction. Deduction is applied here to *all* scientific justifications (Popper uses no exceptions) while it completely ignores any traces of the inductive method. He explicitly writes:

The view advanced here may be called radical 'deductivism'. It holds that all scientific methods of justification are, without exception, based on strictly logical deduction, and that there is no induction of any sort qua scientific method. (Popper 2008, 8)

For Popper, the deductive inference as it is used in science is based on *modus tollens*. His position on this issue is explained in the *Grundprobleme* (Popper 2008, 8) in detail:

The only admissible inferences in an inductive direction – that is, proceeding from a theory's minor premises to its major premises – are the deductive inferences of the *modus tollens*, the falsification

of major premises by way of falsifying the conclusions deduced from them.

Popper's methodology could thus be traced back to his deductivism.<sup>5</sup>

2. Empiricism: a proposition that the truth or falsity of specific matters of facts could only be decided on the basis of (our) experience. This statement literally represents "the fundamental thesis of empiricism" for Popper (*Grundthese des Empirismus*): "Only experience can decide the truth or falsity of an empirical statement"<sup>6</sup> (Popper 2008, 8). In Popper's philosophical system empiricism is associated with the so-called *one-sided falsifiability*. It means that while scientific theories (as general statements about reality) cannot ever be definitively verified, they can still be falsified.<sup>7</sup>

When Popper writes about his *deductivist-empiricism*, he literally describes it as "a synthesis of two classical theories of knowledge" – rationalism and empiricism. According to Popper, classical rationalism is characterized by its deductivist consequences. It enables us to deductively derive single statements from rationally (*a priori*) knowable universal laws of natural sciences. Classical empiricism represents the opposite position together with its inductivist consequences, as the truth of each statement is derived from experience (*a posteriori*).

Popper claims that the dispute between rationalism and empiricism concerns a question of the validity of statements about reality. It is the classic question that asks whether *there are any synthetic a priori judgments*. Popper formulates this in his own words: "Is there any ground of validity for

---

<sup>5</sup> When we evaluate this deductivist assumption of Popper, we shall keep in mind that we comment here on the work of "early Popper", who is more dogmatic than "classical" Popper, which we know well from his later texts. Later Popper is certain to say that there is no method of justification, but of criticism, and this method is, admittedly, a deductive one. However, if we look in the *Grundprobleme*, Popper's former position (as we have tried to show) differed significantly.

<sup>6</sup> Popper considers the basic thesis of empiricism and the fundamental transcendental thesis as analogical (see Popper 2008, 62).

<sup>7</sup> In *The Logic of Scientific Discovery* Popper calls this aspect the Asymmetry between *Falsification* and *Verification*.

non-logical statements other than experience” (Popper 2008, 15)? Rationalism gives a positive answer to this question, but because it does not refer to logic as the basis of validity it must either give up any such foundation or provide a different a priori basis of validity. Popper believes that, for rationalism, this foundation could be *evidence*. Empiricism answers the aforementioned question in the negative, because in addition to logic as the basis of validity for analytical statements, it does not provide any other sources of validity than the empirical verification. Popper (2008, 16) proposes a following solution:

The fundamental *rationalist* idea – there are *a priori* synthetic judgments – can be separated from the idea of *deductivism* with which it is connected, and that these two ideas are by no means logically tied to each other; in the same way, inductivism may be separated from the fundamental thesis of empiricism.

Popper understands his *deductivist-empiricist* view as a special connection between rationalist focus on *axiomatic-deductivist systems* of geometry and a primordial empirical hypothesis that these systems (if they are applicable onto reality) could be decided only by means of our *experience*. The existence of synthetic a priori statements and of inductive inferences is therefore excluded from Popper’s epistemology.

## 2.2. Popper’s transcendentalism

In *Grundprobleme*, Popper refers to his theory of knowledge as to *transcendentalism* with an explicit reference to Kant. Since transcendentalism is based on the so-called methodological or transcendental method, he says:

The term ‘transcendentalism’ will denote the view that epistemological assertions and concepts can and must be critically examined – exclusively – in terms of the actual justification procedure of the empirical sciences. (Popper 2008, 7)<sup>8</sup>

---

<sup>8</sup> This statement clearly shows Popper’s (demarcation) accent on “empirical science”, which is further raised in an even more uncompromising form in his *Logic*.

This means that the theory of knowledge is not an “individual” empirical science but it is a strictly *theoretical science* that relates to empirical sciences in the same way as are empirical sciences related to our experiential reality. Popper explicitly means that just as empirical science gives us the rules to understand reality, then analogically, the theory of knowledge should provide certain principles for understanding empirical science.<sup>9</sup> Put simply, theory of knowledge provides theoretical principles, by which empirical science realizes its processes. Popper specifically writes (Popper 2008, 7): “The theory of knowledge is a science of science. [...] The transcendental method is an analogue of the empirical method.”<sup>10</sup>

Popper characterizes his unique method of research as a general critique of all problem-solving epistemological attempts. Any given criticism is thus focused on finding *contradictions*. Popper further distinguishes between different methods of criticism: there is an example of a *logical method* that seeks “internal” contradictions in the assertions themselves, or we can use an *empirical method* which is supposed to demonstrate “external” contradictions with the facts and experience (see Popper 2008, 57). But the crucial question for Popper is: “*Is there a specifically epistemological method?*” Popper’s answer is affirmative with a remark that we can count as sufficient Kant’s transcendental methods:

Kant was the first who saw this problem. What is alluded to here by the phrase ‘specifically epistemological’, in Kant’s terminology would have to be rendered by the term ‘transcendental’. (Popper 2008, 60)

But even then Popper (2008, 60) has his reservations towards Kant:

It has often been doubted that there is another procedure of immanent criticism in addition to the logical and the empirical testing procedures;

---

<sup>9</sup> Later on, Popper mentions in *Grundprobleme*: “That the theoretical natural sciences exist is a fact. It is the task of the theory of knowledge not to doubt this fact, but rather to explain it” (Popper 2008, 64).

<sup>10</sup> However, we should also acknowledge, that according to Popper, theory of knowledge is not falsifiable; therefore, it is not a theoretical empirical science as it cannot relate to our experiential reality. In other words, the questions of theory of knowledge cannot be answered by experience or experiment.

for while Kant's definition of the *task* of his 'transcendental method' is quite unequivocal, his *solution* of this task, and the more concrete description of the transcendental *procedure itself*, are often rather abstruse and contradictory.

We will mention Popper's critique of Kant's transcendental method (i.e., the transcendental deduction) later, but let us already state that its sole role lies in the fact that the theory of knowledge has to "present a bill" (*Rechnung zu tragen*) to the actual methods of natural sciences. On that basis, Popper completes his initial claim and formulates the so called "fundamental transcendental thesis" (*transzendente Grundthese*):

Epistemological assertions and definitions must be critically examined in the light of the actual procedure of justification employed by the empirical sciences; and only this – transcendental – examination can determine the fate of such assertions. (Popper 2008, 62)

### 2.3. Popper's involvement in Kantian philosophy

Kant's theory of knowledge (i.e., transcendental idealism) is recognized by Popper as the first attempt at a synthesis between classical contradictions of rationalism and empiricism. Kant grasps *the formal aspect* of knowledge by overtaking some elements of rationalism while *the material aspect* is reached through empiricist elements. The central part of Kant's *Critique of Pure Reason* consists of a so-called *Transcendental Analytics* that is regarded by Popper as an elaboration of the problem of induction (in the form of Hume's infamous problem).<sup>11</sup> Kant's transcendental deduction represents for Popper a real solution to the logical aspect of Hume's famous problem, that general statements about reality cannot be drawn from experience. Kant's work "proves" an existence of synthetic a priori judgements on a basis of general formal presuppositions of all material experience. Popper adds to this:

The success of the 'transcendental deduction' depends on the proof that all experience, even singular empirical statements, and thus all

---

<sup>11</sup> Kant's transcendental dialectic is deemed by Popper as an elaboration on the problem of demarcation.

knowledge of reality, are made possible only by specific presuppositions, and that these presuppositions are of the same type as principles of induction; this means, however, that these presuppositions are statements about law-like regularities.<sup>12</sup> (Popper 2008, 68)

It was Kant's discovery that all knowledge of reality is made possible only by specific presuppositions which are statements about law-like regularities. And it is exactly this point that has become essential to Popper's theory of knowledge. However, Popper rejects the "synthetic turnover" (i.e., Kant's formal apriorism that reminds Popper of an old dogmatic reminiscence of traditional rationalism) and it is by these conclusions how Popper later discovers the whole new angle for Kant's critique. The justification of transcendental idealism through the results of transcendental deduction is extremely problematic for Popper. He thinks that Kant made an error because while he was deriving transcendental idealism from transcendental deduction, he confused psychological problems with epistemological ones. Popper claims: "In carrying out the 'transcendental deduction', Kant employs both psychological and – in our sense – transcendental arguments ... in order to establish the formal components of all knowledge" (Popper 2008, 68). For Popper, transcendentalism means the above-mentioned methodological method, or science of science, which critically examines – in Kant's terminology – the conditions of scientific method and scientific knowledge as such.

Popper criticizes Kant's transcendental idealism, which relates these conditions to 1) unconceivable (transcendentally ideal) "thing-in-itself" and also to 2) synthetic a priori cognitions, whose apodictic certainty Kant proves through the so-called transcendental deduction of categories. Popper, however, considers this type of deduction as circular: the assurance of synthetic cognition a priori is derived from the existence of pure a priori forms of knowledge, on the basis of which Kant derives pure a priori concepts (i.e., categories) through transcendental deduction. And it is these categories what guarantees the certainty of those synthetic a priori judgments.

Popper further argues that Kant's transcendental deduction is also uncritical (see Popper 2008, 72). While examining the theory of knowledge,

---

<sup>12</sup> Popper refers to Kant (Kant 1998, B257-262).

Popper really takes into account the results of empirical sciences. In contrast, Kant primarily focuses on such forms of cognition, which are yet to determine the empirical. To sum up: all experience is conditional for Kant. It is thus completely formed by a priori forms of cognition that are 1) pure (i.e. not empirical), 2) the grounds of all empirical experience (e.g. the conditions of the possibility of appearances), 3) apodictic certain, so therefore they are based on the “logic of truth”, respectively on the laws of nature.<sup>13</sup>

#### 2.4. *The initial Kantian question*

How is it possible for certain subjective conditions to be objectively valid? That is the central question for Kant. We can also rephrase it as follows: “How can these subjective conditions at the same time be the most general laws of nature” (Popper 2008, 89)? Popper’s answer here is largely dependent on his explanation of the agreement of any knowledge with its object. This explanation has the following three options for Popper:

1. Our knowledge is determined by its object.
2. The object is determined by our knowledge.
3. *Mittelweg*: We have knowledge as an inborn disposition that is *performed* such that it agrees with its object (see Popper 2008, 90).<sup>14</sup>

---

<sup>13</sup> See Kant (1998, B38-B39).

<sup>14</sup> For additional explanation of Kant's position, see Kant (1998, B166-169): “This cognition, which is limited merely to objects of experience, is not on that account all borrowed from experience; rather, with regard to the pure intuitions as well as the pure concepts of the understanding, there are elements of cognition that are to be encountered in us a priori. Now there are only two ways in which a necessary agreement of experience with the concepts of its objects can be thought: either the experience makes these concepts possible or these concepts make the experience possible. The first is not the case with the categories (nor with pure sensible intuition); for they are a priori concepts, hence independent of experience (the assertion of an empirical origin would be a sort of *generatio aequivoca*). Consequently only the second way remains (as it were a system of the epigenesis of pure reason): namely that the categories contain the grounds of the possibility of all experience in general from the side of the understanding. ... If someone still wanted to propose a middle way between the only two, already named ways, namely, that the categories were neither self thought a priori first principles a of our cognition

Kant (2001, §36) chooses the second option, namely that the “understanding does not derive its laws (a priori) from, but prescribes them to nature”. Popper considers this to be the core statement of transcendental idealism. However, according to him, Kantian doctrine on the subjectivity of the laws of nature can be justified neither by the doctrine of the subjectivity of form of appearances nor by the antinomy doctrine as we can find it in Kant. Such reasoning is considered by Popper a typical example of the confusion between epistemological and genetic (or psychological) aspects of knowledge. As we shall see below, Popper draws a strict distinction between these two positions.<sup>15</sup>

Later in *Grundprobleme*, Popper asks: “How can agreements of the (subjective) conditions of possible experience with (objective) laws of nature be explained”? (Popper 2008, 92) Popper aims to answer this question solely from genetic and psychological perspectives. He further adds:

Any attempt to explain that we can really have knowledge – lies beyond the scope of science (it is ‘metaphysical’). It does not matter if, like Kant, one looks for the basis of explanation in us – in the properties of our understanding, which prescribes laws to nature – or perhaps in the general properties of the world. (Popper 2008, 93)

We are thus acquainted with the properties of our world (including our own reason) exclusively through the natural laws that we seek by means of the methods of natural sciences.

---

nor drawn from experience, but were rather subjective predispositions for thinking, implanted in us along with our existence by our author in such a way that their use would agree exactly with the laws of nature along which experience runs (a kind of preformation-system of pure reason), then ... this would be decisive against the supposed middle way: that in such a case the categories would lack the necessity that is essential to their concept.”

<sup>15</sup> Not only Naraniecki (2014, 65) refers to “the distinction between epistemology and psychological experience of knowledge” as to the central aspect of Popper’s revision of Kant. Wettersten also speaks of “intimate” relationship between methodology and psychology at Popper (see Wettersten 1990, 303).

## 2.5. Popper's genetic apriorism

Popper argues that the original Kantian question (“how can subjective conditions immediately create the general laws of nature”) cannot be answered on the basis of the theory of knowledge, but it should be interpreted from *genetically-biological point of view*. Namely as a question of “how can the agreement of the (subjective) conditions of our cognitive apparatus – of the laws governing the function of our mind – with the (objective) conditions of our environment be explained” (Popper 2008, 94). As a result, Popper formulates a general biological question of the adaptation of living organisms to objective conditions of their surroundings.

Popper affirms the fact that we – as humans – are searching for regularities in every aspect of the outer world that surrounds us. For Popper, this statement presents a basic condition for our intellectual adaptation (a human *a priori preformation*): “Only the existence of this basic intellectual function, namely the searching for regularities, makes possible the process of intellectual adaptation (cognition)” (Popper 2008, 95). Afterwards he explains a biologically proven (this always means a clear a priori sign for Popper) fact that we are able to reason through (or by means of) hypothesis and that we gathered these “reasoning functions” through our sheer adaptation to reality.

Kant's inquiry into the agreement between our intellect and relations in the world is interpreted by Popper as a purely biological question of genetic adaptation. Popper's approach can be described as *genetic apriorism* since the basic intellectual functions are innate to us. Out of the three previously mentioned options, Popper favours the third one (*Mittelweg*). In connection with this Popper distinguishes between external and internal conditions: The World as our environment (*Umwelt*) can be regarded as the embodiment of biologically relevant *external conditions*. However, what is biologically relevant and how it is relevant depends largely on *internal conditions*. These *internal factors* should help to illustrate the fact that our knowledge is anthropomorphic (see Popper 2008, 97).<sup>16</sup>

---

<sup>16</sup> The topic is further explained by Kant himself: “A middle course may be proposed between the two above mentioned, namely, that the categories are neither self-thought first principles a priori of our knowledge nor derived from experience, but subjective dispositions of thought, implanted in us from the first moment of our existence, and so ordered by our Creator that their employment is in complete harmony with the laws of

## 2.6. *The problem of anthropomorphism and apriorism*

Popper always claimed that *anthropomorphism* has an essential place in Kant's theory of knowledge, because the doctrine of transcendental idealism and the notion of *ding an sich* can be understood (from a biological standpoint) as a fact that, as humans, we cannot overcome anthropomorphic limits of our understanding and knowledge at all. From the epistemological perspective, however, Popper calls this a *problem of anthropomorphism* (or a *problem of subjectivity of our knowledge*), which he describes as "banal rather than subtle" (Popper 2008, 97).

Kant's optimistic epistemological viewpoint<sup>17</sup> is seen as untenable by Popper. Furthermore he considers Kant's apriorism entangled in many inner contradictions. As to the circular argumentation, Popper refers especially to Kant's concern with the apriority of the principle of induction which justifies the "necessary existence" of general laws of nature through the thesis that these laws are prescribed to nature through our own understanding. Popper hence argues that transcendental idealism can only explain the psychological apriorism of natural laws, not their epistemological priority.

## 2.7. *Deductive psychology of knowledge*

The deductive psychology of knowledge (which stands right beside the deductive epistemology) represents for Popper an alternative examination of the emergence of knowledge in the biological and psychological sense.<sup>18</sup> Popper talks in this context about the so-called *genetic deductivism*. According to him, our thoughts must be characterized as a series of

---

nature in accordance with which experience proceeds – a kind of preformation-system of pure reason. Apart, however, from the objection that on such a hypothesis we can set no limit to the assumption of predetermined dispositions to future judgments, there is this decisive objection against the suggested middle course, that the necessity of the categories, which belongs to their very conception, would then have to be sacrificed" (Kant 1998, B167-168).

<sup>17</sup> This viewpoint is described as one that compensates necessary anthropomorphic limitations of our knowledge by a priori valid synthetic views.

<sup>18</sup> Popper draws his inspiration here partly from Fries' psychological interpretation of Kant. Naraniecki (2014, 52) summarizes it as follows: "Fries' acceptance of psychologism remained unsatisfactory for Popper as this psychological revision of Kant merely deferred

intellectual responses that are subjectively preformed. But how to explain that reactions that are subjectively preformed and therefore do not come from experience – those ones that have been found in objective situations of the external world – show themselves as biologically valuable? Popper refers to these subjectively preformed reactions as our *anticipation* and argues:

According to the deductivist view, we do not attain our empirical knowledge by abstraction or generalization from sense-perceptions, but by trying out anticipations tentatively assigned to the ‘material’ of the receptions. Whether this tentative assignment will be abandoned or not is decided by its biological value. The method of deciding is a selective one. [...] Success in the environment determines the fate of preformed anticipations. (Popper 2008, 28)

Popper admits that *genetic deductivism* cannot explain the creation of new anticipations. For Popper, there is no inevitable or rational path that would lead from new receptions to new reactions. It is a system based on selection, more precisely on a “method of Trial and Error” (Popper 2008, 29).

According to Popper, it is possible to understand the term “a priori” in a psychological sense as *something that does not arise from experience*. Psychological use of this term is synonymous with (the already mentioned term) *anticipations*. These can therefore be understood as synthetic a priori judgments: “But these ‘a priori synthetic judgments’ would be only tentative anticipations, they would only exist *a priori*, that is, prior to being empirically corroborated; *a posteriori* they could still be rejected, refuted by experience” (Popper 2008, 33).

---

lawfulness from Kant’s consciousness (*Verstandesgesetzlichkeit*) to psychology, thus accomplishing nothing. Like the rest of science for Popper, epistemology required some means of its acceptance over competing theories even if we cannot ultimately justify such theories.” Naraniecki considers this as a proof of Popper’s epistemological and methodological non-foundationalism (see Naraniecki 2014, 52).

### 3. Summary

Based on an externalist approach, the first part of our study explains the circumstances of the origin of Popper's first major work. We have showed that the interactions with the members of the Vienna Circle as well as discussions with other notable neo-positivists played an all-important part in the formation of Popper's thinking. Our research has revealed its two concrete consequences:

1. Popper had to clearly distinguish himself from the neo-positivism. This has contributed to formulating of his own authentic position, which is so well known today through its declaration in *The Logic of Scientific Discovery*.
2. His first major treatise, *Die Beiden Grundprobleme der Erkenntnistheorie* – whose core consisted in a confrontation with 18<sup>th</sup> and 19<sup>th</sup> centuries epistemologies – did not see the light of day. In addition to its difficult readability and bulkiness, there was the fact that traditional theories of knowledge were considered an absurd metaphysics by neo-positivists.

The publication of Popper's work is thus postponed for another fifty years. The majority of the text of *Grundprobleme* is almost completely abandoned by Popper when he publishes *The Logic of Scientific Discovery* and "Hume's" or "Kant's" problems are mentioned only briefly, with no deeper explanation.

In the second part, we showed that Seubert's assertion<sup>19</sup> is somewhat simplified. First, we can track a vast source of Popper's inspiration in the work of Kant. Secondly, we can observe that Popper's approach towards Kant's formal epistemological apriorism is mostly critical. We can see a positive aspect of this inspiration in a significantly Kantian "transcendental questioning" by which Popper builds his very own theory of knowledge and places it on the pedestal of transcendental method. This method examines an intrinsic possibility of empirical science as such, respectively the

---

<sup>19</sup> Seubert claims that through *Die beiden Grundprobleme der Erkenntnistheorie* Popper "definitely and knowingly accepts Kant's stance" (Seubert 2016, 8).

validity of the laws of nature. Another source of Popperian inspiration is Kant's attempt at a synthesis of classical rationalism and empiricism. Although Popper himself performs a similar synthesis, it does not result in formal apriorism, but in *deductivist-empiricist epistemology*. That consists of two main aspects for Popper: 1) a rationalist method of scientific reasoning (deduction); 2) an empiricist assumption that the veracity or falsity of singular empirical statements can only be decided on the basis of experience.

Moreover this critique is concerned with the fundamental problem that Popper finds in Kant – his advancement from the basis of transcendental deduction towards transcendental idealism. Popper points out that Kant is confusing psychological with epistemological problems and argues that the Kantian question, “how to explain a compliance of (subjective) conditions of experience with (objective) laws of nature?”, cannot be answered from the epistemological point of view, but only in genetic and psychological terms. He therefore reinterprets this question with the help of “his” *genetic apriorism* as a purely biological one that concerns genetic adaptation. According to this theory, our basic intellectual functions are inborn and they are *performed* in a way of human natural adaptation. This innate adaptation represents the so-called internal conditions of our knowledge and leads to the problem of anthropomorphism (the problem of the subjectivity of our knowledge).

On this account, Popper simply says that such anthropomorphic boundaries of our knowledge cannot be overcome. However, from an epistemological perspective, he considers this issue not only a circular one but also unquestionably trivial. A theory of knowledge does not wonder that there is knowledge itself, but seeks only to state its possibilities and limits.

#### Acknowledgments

This research was supported by Faculty of Philosophy and Arts, The University of West Bohemia (Specific Research Grant: *The Roots of Critical Rationalism, Its Legacy in the Postmodern Discourse and Actual Reflection*, No. SGS-2016-009).

## References

- CARNAP, R. (1932): Über protokollsätze. *Erkenntnis* 3, No. 1, 215-228.
- HACOHEN, M. (1998): Karl Popper, the Vienna Circle and Red Vienna. *Journal of the History of Ideas* 59, No. 4, 711-734.
- HACOHEN, M. (2002): *Karl Popper – The Formative Years, 1902-1945*. Cambridge: Cambridge University Press.
- KANT, I. (1998): *Critique of Pure Reason*. Cambridge: Cambridge University Press.
- KANT, I. (2001): *Prolegomena to Any Future Metaphysics*. Indianapolis: Hackett Publishing Company.
- MOURAL, J. (1997): Karl Popper česky. *Kritický sborník* 17, 47-60.
- NARANIECKI, A. (2010): Neo-Positivist or Neo-Kantian? Karl Popper and the Vienna Circle. *Philosophy* 85, No. 4, 511-530.
- NARANIECKI, A. (2014): *Returning to Karl Popper*. Amsterdam: Rodopi.
- POPPER, K. (1933): Ein Kriterium des empirischen Charakters theoretischer Systeme. *Erkenntnis* 3, 426-427.
- POPPER, K. (2002a): *The Logic of Scientific Discovery*. London: Routledge.
- POPPER, K. (2002b): *Unended Quest: An Intellectual Autobiography*. London & New York: Routledge.
- POPPER, K. (2008): *The Two Fundamental Problems of the Theory of Knowledge*. London: Routledge.
- SEUBERT, H. (2016): Karl Popper wiedergelesen – was bleibt. Häretische Annäherung an einen Klassiker der Rationalität. *Aufklärung und Kritik* 23, No. 1, 4-19.
- WETTERSTEN, J. (1990): Integrating Psychology and Methodology: How can Psychology and Methodology Be Integrated? *Journal for General Philosophy of Science* 21, No. 2, 293-308.

## Some Problems of Glavaničová's Approach to Fictional Names

MILOŠ KOSTEREC<sup>1</sup>

In a recent issue of this journal, Daniela Glavaničová presented her theory of fictional names (see Glavaničová 2017, 396-402). As a part of the project within the field of philosophical analysis of fictional discourse (see at least Haraldsen 2017, Vacek 2017, Zouhar 2017 and Zvolenszky 2015), Glavaničová provided an analysis of fictional names based on the idea of fictional names modelled by necessarily empty individual roles. I limit my discussion of the proposal to the system of Transparent Intensional Logic (henceforth TIL),<sup>2</sup> since Glavaničová states: "The paper examines two possible analyses of fictional names within Pavel Tichý's Transparent Intensional Logic" (Glavaničová 2017, 384). Before that, though, I will sketch her proposal first.

First, the proposal is based on the notion of an *individual role*.<sup>3</sup> The notion is framed within the system of TIL according to which an individual role is a model for a position (e.g. office like the president of USA) an in-

---

<sup>1</sup> ✉ Miloš Kosterec

Institute of Philosophy, Slovak Academy of Sciences  
Klemensova 19, 811 09 Bratislava, Slovak Republic  
e-mail: milos.kosterec@gmail.com

<sup>2</sup> For a detailed presentation of TIL see at least Duží, Jespersen & Materna (2010) and Tichý (1988).

<sup>3</sup> I advise the reader unfamiliar with the notion to read Glavaničová's introduction of the notion on pp. 394-396 of her article.

dividual might occupy. From the technical point of view, a role is a *functional object*. It is a function from possible worlds and times to individual (if any). Note that in TIL it is not hyperintension (i.e. a construction). Similarly as with offices in the real world, it is possible for a role to be unoccupied.

Such a property is grasped by the fact that, within TIL, functions are partial, i.e. they need not have a value on an argument. For example, consider an actual king of France. The office is currently empty, meaning the king of France actually does not exist (at least of one of its readings). And this it seems to be the beneficial core of Glavaničová's proposal.

Glavaničová states:

Fictional names should be analysed in terms of individual roles... However, these individual roles are necessarily non-occupied (empty). As it is with the other expressions, we can pronounce a fictional name to speak about its sense (a construction), about its reference (a role), or about its extension (which, as it should be clear, does not exist). (Glavaničová 2017, 396-397)

There are several supposed positives Glavaničová suggests, namely: a) we can analyse an ascription of a property to fictional character as ascribing requisities to an individual role; b) we can explain the creation of fictional character by author picking an expression for the role; c) roles are abstract entities, therefore they are intersubjective; d) no new types of entities are presupposed, roles are no new queer entities; e) the roles might not be described completely; f) roles can be empty, according to Glavaničová's proposal, there is no *real* individual Sherlock Holmes, therefore it is rather easy to analyse negative existential claims about fictional characters (e.g. Sherlock Holmes does not exist.) as true statements; g) we can model quantification over fictional characters; h) we can use constructions to differentiate among various empty roles; h) we can model personal attitudes towards fictional characters.

Glavaničová also goes throughout several objections against her proposal (see Glavaničová 2017, 399-402) yet, in my eyes, unconvincingly. My main objection is, that *in TIL, there is only one necessarily empty individual role*. Glavaničová does not seem to fully admit the fact. She states:

Since the reference of a fictional name is a necessarily empty role, one needs something to differentiate *between various empty roles*. This falls within the competence of constructions. They can embody the ‘Holmes-ish’ and ‘Watson-ish’ ways of believing, in an exact way. (Glavaničová 2017, 399; emphasis M.K.)

The reason is that an individual role is a functional object, whose identity is not given by its construction. Rather, it is given by the input/output mapping from definition range to the range of values. In other words, individual roles have the extensional criterion of identity. That means that two individual roles are different if and only if there is at least one argument of the proper type (tuple of possible world and time) upon which the values of these two roles differ. But, according to the proposal at issue, the fictional roles stand for, and only for, *necessarily empty* roles. A necessarily empty role is such a function which does not have any value for any proper argument. Since there is only one such role every fictional name stands for the same object.

Consequently, speaking about Sherlock Holmes does not differ from speaking about Watson, Moriarty, etc., going thus against the desiderata to differentiate between fictional characters. Being aware of the problem, Glavaničová adds:

One of the reviewers claimed that there could be only one trivialization of necessarily empty role in TIL, so every fictional name would have the same meaning. It depends on the identity criteria of constructions in general and constructions of *necessarily empty roles* in particular. I do not see any obstacle in differentiating between different (constructions of) empty roles on the basis of their requisites. E.g., the meaning of Holmes is different from the meaning of Watson, because the requisites for being Holmes differ from the requisites for being Watson. (Glavaničová 2017, 398, footnote 22; emphasis M.K.)

The above mentioned problem stems from the fact that there is only one necessarily empty role within TIL. For, the analysis of proper name in TIL goes via trivialisation, according to which there is only one trivialisation for each object. The problem, however, is that there is only one necessarily empty individual role.

In what follows, Glavaničová seems to suggest the possibility of new criteria of identity of constructions. It can be done, but at the cost of leaving TIL (at least in its present forms). One way or the other, her proposal will still have problems in both intensional and extensional contexts, since the contexts respect the substitution of equivalent constructions *salva veritate*. To repeat, Glavaničová's proposal assigns, undesirably, all fictional names to the same object.

The second objection is that the properties of individual office are type theoretically different from the properties of individual. Although Glavaničová benefits from this fact when analysing negative existential claims about Sherlock, it is hard to see how the proposal would cope with sentence such like:

“Sherlock Holmes weighs 160 pounds.”

We do not seem to be talking about the property of any office. The sentence (if taken genuinely) must include ascription of weight to some individual on type theoretical analysis. If author stated, instead of intending that the ascription was not type-theoretically assigned to an individual, but rather to some other type of object, then she would seem to propose that ordinary relations (as weighs) have some non-ordinary meaning within fiction. But if the sentence includes ascription to an individual, that is Sherlock Holmes used *de re*, then the sentence will not have a truth value (since there is not any such an individual in the proposal). And this goes against any informativeness of any sentence containing fictional name used *de re*.

The third objection concerns Glavaničová's proposal regarding modalities within the fiction. As far as “Sherlock Holmes need not be a detective” is true *within the fiction*, the proposal has bad results. According to it, properties are ascribed to individual offices as their requisites. That means, however, that a fictional character (denoted by the use of a fictional name) has all its properties necessarily. If Sherlock Holmes is a detective it is necessary that Sherlock Holmes is a detective. In short, as far as there is a nontrivial modality within fiction (i.e. characters have at least some properties merely possibly), Glavaničová's proposal fails.

The fourth objection concerns the purported use of requisites when discerning necessarily empty roles. Following the reasons above, it evokes an attempt to discern two different numbers 1 within the same arithmetics.

There is only one such thing. Granted, Glavaničová understands that the notion of requisite (see Duží, Jespersen & Materna 2010, chapter 4.1), needs a different definition. She thus suggests the following:

Another option is to use definitions of requisites from the above book, but change the material implication for some other sort of implication. A further option is to treat the notion of requisites as primitive. Finally, my preferred option is to define the requisites in terms of the content of the respective work of fiction. Note, however, that there are principal reasons why the essential properties cannot be defined once and for all: identity of fictional characters is interest relative, and so is the extent of their essential properties... (Glavaničová 2017, 400)

The question now stands: which kind of implication should we use? A strict implication would not help, at least in the case of necessarily empty roles. Another option is a scientifically empty way of solving the problem. In short, if we state that two objects (empty roles) differ because there are other objects (sets of requisites) that differ, we seem to have a criterion. But if we claim that the requisite is a primitive object, we do not explain the difference, but only state a presupposition. The third, preferred, way is too vague.

In conclusion, I find Glavaničová's proposal original, yet unintuitive. Although I see some positives of her proposal, I think that the negative consequences are too much to bear. I know that it is only an outline of a theory and the proper formulation will come soon. I therefore hope that this discussion note will contribute to improvement, as well as better understanding, of the view.

#### Acknowledgments

This paper was supported by VEGA grant No. 2/0049/16 *Fictionalism in Philosophy and Science*.

## References

- DUŽÍ, M., JESPERSEN, B. & MATERNA, P. (2010): *Procedural Semantics for Hyperintensional Logic. Foundations and Applications of Transparent Intensional Logic*. Berlin: Springer.
- GLAVANIČOVÁ, D. (2017): Tichý and Fictional Names. *Organon F* 24, No. 3, 384-404.
- HARALDSEN, F. (2017): The Truth about Sherlock Holmes. *Organon F* 24, No. 3, 339-365.
- TICHÝ, P. (1988): *The Foundations of Frege's Logic*. Berlin & New York: de Gruyter.
- VACEK, M. (2017): Fiction: Impossible! *Axiomathes*, DOI 10.1007/s10516-017-9353-3.
- ZOUHAR, M. (2017): The Fiction of Fictionalism. *Filozofia* 72, No. 8, 603-615.
- ZVOLENSZKY, Z. (2015): An Argument for Authorial Creation. *Organon F* 22, No. 4, 461-487.

František Gahér and Vladimír Marko: *Method, Problem, and Task*  
[*Metóda, problém a úloha*]  
Vydavateľstvo UK, Bratislava, 2017, 192 pp.<sup>1</sup>

*Method, Problem, and Task* is a recent book on method authored by František Gahér and Vladimír Marko. The book is relatively slim, and this can be surprising in view of the fact that its topic remains rather underdeveloped in the current philosophy of science. Nevertheless, the authors succeeded in presenting an interesting conception of method which is accessible to wide variety of readers from different fields of interests, not necessarily scientific ones. Yet, the book has its drawbacks as well.

The first two chapters (*Introduction* and *The Question of Method*) state the central motives and aims of the book, and describe the plan of the subsequent inquiry. According to Gahér and Marko, the definitions of method that are available in the philosophy of science literature are generally rather unsatisfactory. They often take the form of an ostensive definition, or just describe the etymology of the term in question (p. 13). Moreover, and this is a crucial point, these definitions suffer from being inapplicable outside the disciplines in which they were developed (p. 10). Since important methodological concepts such as *method*, *problem*, *task* and the related concepts, as explicated in the philosophy of science, were unsatisfactory for the purposes of other scientific disciplines, the latter were forced to come up with their own fundamental methodological concepts in order to be able to keep operating in their competitive environment (p. 11). Due to the growing and deepening interdisciplinarity, there is natural tendency to find a common ground for these particular results; it seems, however, that the philosophy of science is largely excluded from undertaking this project.

The aim of the authors is to provide a definition of method that would both comply with the aims and results reached in scientific disciplines such as AI, program languages, cognitive sciences and contribute to the philosophy of science.

---

<sup>1</sup> ✉ Tomáš Kollárik

Department of Logic and Methodology of Sciences, Faculty of Arts  
Comenius University in Bratislava  
Gondova 2, 814 99 Bratislava, Slovakia  
e-mail: tomas.kollarik@uniba.sk

In chapter *The Question of Method*, the authors map changing attitudes towards the importance of the concept of *context of discovery* in the philosophy of science, which was clearly underrated by leading persons like Reichenbach and Popper. The authors illustrate this claim by a quotation in which Popper maintains that this concept (i.e. the concept of the context of discovery) belongs to the scope of empirical psychology and cannot be an object of logical analysis, in contrast to the closely related concept of context of justification (p. 15). Scales were slightly tilted in favor of the context of discovery in second half of the 20<sup>th</sup> century, when prominent authors (Gahér and Marko mention Kuhn, Laudan, Hintikka, Nickles) studied scientific progress and mechanisms of scientific discovery, which began to be understood as a special case of the mechanism involved in problem solving.

Probably the most emphasized claim in the book says that a problem is always a problem of a subject who aims at achieving a certain goal. An immediate consequence of this presupposition is that there are no problems without someone who acts with the aim to achieve, reach, change, create, or destroy something. Differences in goals, purposes, knowledge, interests of subjects aimed at obtaining a certain goal lead to another important consequence: the conditions that constitute a problem for one subject do not necessarily constitute a problem for another subject. The key variable on which the main stress of the book is placed is knowledge. Gahér and Marko claim that in order to solve a problem one must alter her own *knowledge*.

How problems arise according to the authors? Simply speaking, there must be (i) a goal-oriented subject. (Let's denote the "subject" Mary). (ii) Mary tries to do something (e.g., she tries to replace a bulb), but (iii) she does *not know* how to do so. Mary is thus unable, at least temporarily, to fulfill her goal – she is facing a problem. When Mary finally acquires all information necessary for replacing the bulb (she *knows* what to do) she faces a *task*. In other words, what Mary actually did, according to the authors, was what all subjects who strive for obtaining a certain goal attempt to do when facing a problem, namely transforming a problem into a task (p. 170). The differences between a problem and a task are stressed repeatedly and in different ways throughout the whole book, but the demarcating line is drawn as early as in the chapters *Problematic and Unproblematic Situation* and *Problem and Task*.

The most extensive chapter entitled *Problem* introduces a number of key notions like *problem space* or *problem representation*, *knowledge space*, *relevant knowledge*, *problem solution space*. The *process* of transforming a problem into a task is articulated in this chapter in the following way. We never find ourselves in a problematic situation without having *any* knowledge. Knowledge space consists

of all knowledge that is at our disposal at the moment of problem arising. Only part of it (namely relevant knowledge) is used in order to shape the problem space or the problem representation. The problem space is a way of how we understand or represent a problem at a given time. It consists of these types of element: a) the set of *states* (an initial state, transitional states and a target state), b) the set of *operators* (legitimate and effective procedures of transforming one state into another), c) the set of *local information* (information about the current state and preceding states), and d) the *set of constraints and requirements* (simply speaking, the set of constraints and requirements excludes some possible ways that lead us from an initial state through some transitional states to a target state).

Moreover, Gahér and Marko introduce a distinction between *abstraction* space and *execution* space (p. 39). The need for this distinction becomes clear once we realize that we often deal not only with simple problems or tasks, but also with *complex* or *hard* problems and tasks (pp. 32-33). These kinds of problem often involve several problematic segments including a multitude of stages opened to different types of operators or are such that their execution is in some sense very difficult, time-demanding etc. *Planning* at the level of *abstract* space often includes *simplification* in the sense of eliminating marginal, or minor, tasks or subproblems and focusing on the main ones. If the abstraction level is not divided into several stages of abstraction (it depends on the strategy accepted by a solver and the nature of a problem), then the solution of the problem proceeds at two different levels. The problem is initially solved at the abstraction level and this solution is subsequently tested at the execution level.

The second half of the book is focused on the notions of task and method. The reader can now see the demarcation line between task and problem more clearly. According to Gahér and Marko, tasks and problems are structurally similar – both can be presented as consisting of the sets of stages, operators, local information, constraints and requirements. Both can be simple or complex and both can be initially planned at the abstraction level, then projected onto a lower level and finally executed. The key difference concerns the subject whose knowledge space, in the case of task, includes information about how to transform an initial state into a target state.

It is natural to expect that methods can be applied repeatedly at different times and places and that they produce the same type of result in the same type of situations. Therefore, procedures, as realized in particular circumstances, which happens especially in the case of complex tasks, are not good candidates for being methods. This is because procedures occur in rather unique conditions that are very unlikely to repeat. In relation to these considerations, I find very interesting the

part of the book in which the two related concepts of *generic task* and *generalized task* are introduced (p. 109f). Generic task is an *elementary abstract scheme* and represents a *pattern* of how to achieve goals of a specific type (e.g., classification, comparison, evaluation of objects). Generalized task is an abstract scheme that consists of several generic tasks (e.g., classification *and* evaluation). If someone tries to fulfill a complex task, she should a) decompose it in order to b) get subtasks and then c) identify relevant generic tasks. The subtasks are then d) organized in order to accomplish the complex task in accordance with certain constraints and requirements.

The need to accomplish specific types of generic tasks is related to certain conditions that are specified by *production rules* which are conditional in their form: IF (some conditions obtain), THEN (accomplish this procedure). The “IF” is followed by a sufficient condition, or a set of conditions that together constitute a sufficient condition for applying a particular procedure specified after “THEN”.

Now it should be clear that the authors refuse an understanding of method as a way of solving problems or a *manual* (the readers may find a refined view of method as a way of solving problems in Zouhar, Bielik & Kosterec 2017). Using a method is always related to a solver who has certain motivations, preferences etc., and to a particular situation. These two elements are often so unique that the view according to which method is simply some kind of manual applicable in specific circumstances would lead to a *proliferation* of methods, and consequently to diminishing their value. Moreover, the authors deny the widespread opinion that we produce and learn methods in order to solve problems. According to them, problems occur if and only if there is a goal-oriented subject, and a proper method that could be used to achieve the goal is absent (p. 123). The existing methods, generic tasks, or abstract patterns are results of transforming problematic situations into unproblematic situations.

The above considerations lead the authors to the conclusion that scientific work is – at least at the stage of developing “normal science” – largely a routine enterprise. Scientists use well-tested methods in order to achieve desired goals. Regarding the criteria of *being scientific* that are applicable to methods, the authors believe that they are historically dependent and therefore not absolute. Nonetheless, they add that method must meet some minimal requirements in order to be truly scientific: every transition from one stage to another stage, when moving in a problem or task space, should be governed by warranted rules that represent a causal or some other kind of universal relation. In comparison to a simple manual in which many connections between states of executing a task are left unpronounced for practical reasons, a scientific method should be open and transparent and based on

explicit and accessible reasons. This last feature is related to the *predictive* and *explanatory* power of scientific method – each state of a task (except for the first state and the last state) should be identifiable as a consequent *and* an antecedent of another state of task (pp. 140f).

The last chapter of the book deals with method from more formal point of view, but formalism is reduced to a necessary minimum. The conceptual apparatus of the book is enriched here with new notions, mostly those originated by Tichý, and we can even see traces of his influence at the level of literary style. The authors describe sentence meanings as procedures. Procedures have results. One can understand a sentence (i.e., know which procedure it expresses) without knowing the result of the procedure. If one wants to know what the result of a procedure is, she can express her *cognitive attitude* toward the result by a question. And, finally, orders are impulses to achieve, realize or find results of procedures. In other words, orders are impulses to perform methods in order to obtain results. In contrast to other abstract entities (e.g., sets, numbers etc.), method is a kind of procedure that can be an object of *conative* attitudes of subject. In other words, their execution leads to results that can be not just abstract entities but also space-time objects (statues, buildings, etc.) Orders can be more or less specified. If orders are spelled out in great detail, an agent cannot but execute *the* procedure expressed by the order without taking the liberty of carrying out some free steps of her own. In the case of less specific or more general orders, an agent must choose from a class of *functionally equivalent* procedures that are similar in leading to the same result.

The book is written in a readable style. Gahér and Marko succeeded in developing an original conception of method that integrated the results obtained in many different scientific disciplines. Apart from introducing their own notions, the authors use terms that already are in circulation in other fields of inquiry. The reader can find in the book a lot of illuminating illustrations and examples. That is why the book is accessible to wider audience.

There are some drawbacks in the book as well. Apart from the syntactical and grammatical ones, the list of contents (p. 3) does not capture the actual structure of the book. The names of each chapter and subchapter are provided but their order is sometimes switched. Furthermore, on pp. 40-43, the authors present a distinction between *objective* and *subjective* problem space, and remind the reader that the expression “objective” characterizes a *type* of problem space (i.e., a type of *representation* or *model* of problem) and is not used to suggest that problems exist *independently of subjects* (p. 41). But few pages below (p. 46), the objective problem space is described (in contrast to the subjective problem space) as a *real* one. This

comparison does not seem to be correct because both the objective and the subjective problem space are representations and as such they do not differ ontologically, but rather epistemologically. While the objective space always involves solution and all possible trajectories from initial state to the target state, the subjective space does not.

The above errors are minor. They can cause a little confusion but careful readers can cope with them easily. The book offers a lot of insightful considerations, and that is why it can be recommended to every reader who is interested in the methodology of sciences.

*Tomáš Kollárik*

### References

- ZOUHAR, M., BIELIK, L. & KOSTEREC, M. (2017): *Metóda: metodologické a formálne aspekty* [Method: Methodological and Formal Aspects]. Bratislava: Comenius University in Bratislava.

## Modal Metaphysics: Issues on the (Im)Possible V August 17-18, 2017 Bratislava<sup>1</sup>

The Department of Analytic Philosophy of Slovak Academy of Sciences has been organising several international conference last five years. The topics of the conferences range from metaphysics (Issues on the (Im)Possible series), through philosophy of language (Philosophy of Language (I): Semantics of Fictional Discourse) to logic (Current Trends in Deontic Logic). In August 17-18, 2017 the first conference listed above instantiated its fifth instance. The conference, as usual, took place at the Institute of Philosophy of the Slovak Academy of Sciences. Divided into two days and two parallel sessions the conference was able, and the organisers pleased, to accommodate two keynote talks and twenty-five accepted contributions. The former were given by Philip Bricker (University of Massachusetts) and Seahwa Kim (Ewha Womans University), while the latter included scholars from all around the world.

The first day of the conference commenced with a session split into Áron Dombrovski (Eötvös Loránd University)'s 'The New Theory of Modal Fictionalism' and Borut Cerkovnik (University of Ljubljana)'s 'Modality and Ontology of the Tractatus' papers, commented by Maciej Sendłak and Yujian Zheng respectively. Next session was shared between Yujian Zheng (Lingnan University) and his 'Backtracking Counterfactuals in Causal Reasoning' followed by Shyane Siriwardena's comments, and 'Ascription, the Regress(es) of Instantiation and the Problem of Relatedness' given by Francesco Spada (University of Modena and Reggio Emilia) and commented by Bing-Cheng Huang. William Bondi Knowles (University of Edinburgh) reviewed 'Conceivability Arguments and Their Use in Philosophy' (commented by Peter Marton) while, at the same time, Mark Garron (McMaster University) discussed the triad 'Quine and Priest and the Excluded Middle'.

---

<sup>1</sup> ✉ Martin Vacek

Institute of Philosophy  
Slovak Academy of Sciences  
Klemensova 19, 811 09 Bratislava, Slovakia  
e-mail: martinvacekphilosophy@gmail.com

After lunch, the sessions continued with a pair of talks, namely Tien-Chun Lo's (National Taiwan University) 'On the Relation between Temporal Necessity and Metaphysical Necessity' and Zsófia Zvolenszky's (Slovak Academy of Sciences) 'Inadvertently Created Fictional Characters are Nothing New' followed by 'Facts, Truths, and Liars' delivered by Peter Marton (Clark University) and commented by Mark Garron, and 'A Modal Semantics for Essence and Ground' given by Jonas Werner (University of Hamburg) commented by Mattia Sorgon. The last three slots of contributed papers were filled in by 'Expressivity of Language without Commitment to Possible Worlds', 'The Representational Deficiencies of Hybrid Modal Realism' and 'Towards a Modally Harmonious Theory of Counterfactuals: Challenging an Edgingtonian Approach' given by Bing-Cheng Huang (National Taiwan University), Matthew Collier (University of Oxford) and Shyane Siriwardena (University of Leeds) (and in the same order commented by Matteo Pascucci, Robert Michels and William Bondi Knowles). The end of the first day was reserved for the first keynote address, given by Philip Bricker. In his 'Is there a Humean Account of Qualities', Bricker addressed the Humean problem with quantities and presented what he takes to be the best account of quantities, combining his own commitment to quidditism as well as his comparativist inclinations.

The second day of the conference started with a pair of talks, namely Ilaria Acquaviva (Fondazione Collegio San Carlo, Scuola Alti Studi)'s 'Francisco Suarez's Model Theory: An Actualists and Abstractionist Paradigm on Real Possible Beings' and Aaron Ben-Ze'ev (University of Haifa)'s 'I Want to Know Where Love Is: The Extended Epistemological and Ontological Status of Romantic Experiences'. After them, Mattia Sorgon (University of Alberta) proposed 'The Modal Account of Essence: An Analysis of the Notion of Sparseness' (with Francesco Spada as a commentator) and, at the same time, Péter Susánszky (Central European University) discussed 'Fictionalism and Mere Possibilia' (followed by Áron Dombróvski's comments). Matteo Pascucci (University of Salzburg) discussed 'Modalities between a Logical and an Empirical Account of Indeterminism' (featuring Tien-Chun Lo as a commentator) and Robert Michels (Université de Genève) provided an account of 'Cross-World Comparatives for Lewisians' (commented by Michael Nelson).

After lunch, four more contributed papers were presented. Penultimate session was filled in by Lukas Skiba (University of Hamburg) and Maciej Sendłak (University of Warsaw) with their 'Fictionalism and the Modal Status of Fictions' and 'Worlds as Non-existent Objects' respectively, commented (in the same order) by Zsófia Zvolenszky and Sergio Genovesi. Finally, Michael Nelson (University of California-Riverside) presented 'Contingent Existents' paper with Lukas Skiba in

the role of a commentator, and Sergio Genovesi (Rheinische Friedrich-Wilhelms-Universität Bonn), on a similar note, spoke about ‘Contingent Events’ (commented by Matthew Collier).

The last talk of the conference, the second keynote address, was delivered by Seahwa Kim. In her ‘Fictionalism, Modal Fictionalism and Truth in Fiction’, Kim discussed Jason Stanley’s objection that hermeneutic fictionalism is revolutionary fictionalism as well as Stephen Yablo’s if-thenism. She then argued that Stanley’s objection is based on the conflation of two usages of ‘hermeneutic vs. revolutionary’ and, subsequently, pointed out that fictionalism is incompatible with if-thenism. Finally, Kim commented on an ordinary notion of truth in fiction.

Issues on the (Im)Possible V picked on the previous conference’s temporal parts. Again, the presence of brilliant philosophers and personalities have made the organisers thinking about its future continuants. Readers thus should stay tuned and keep checking conference sites ([www.metaphysics.sk](http://www.metaphysics.sk)) which, beside the history of the Issues on the (Im)Possible, tracks and will track its further evolution.

*Martin Vacek*

### Current Trends in Deontic Logic November 22-24, 2017, Bratislava, Slovakia<sup>1</sup>

The first volume of the conference *Current Trends in Deontic Logic* was organized by the *Department of Logic and Methodology of Sciences* (Comenius University in Bratislava; D. Glavaničová, T. Kollárik, M. Zouhar), co-organized by the *Department of Analytic Philosophy* (Slovak Academy of Sciences; M. Vacek) and supported by the *Jan Hus Educational Foundation*. To our knowledge, it was the first conference on deontic logic in Slovakia.

A prelude to the conference was a seminar with Sven Ove Hansson. The seminar was devoted to the paper Hansson (2006) on the ideal-world construction in deontic logic and its criticism. In a nutshell, the crucial question is the following: Is it a good approach to analyse what we ought to do in terms of what we ought to

---

<sup>1</sup> ✉ Daniela Glavaničová

Department of Logic and Methodology of Sciences  
Comenius University in Bratislava  
Faculty of Arts, Gondova 2, 814 99 Bratislava, Slovak Republic  
e-mail: [dada.baudelaire@gmail.com](mailto:dada.baudelaire@gmail.com)

do in an ideal world (as the standard deontic logic and its siblings suggest)? Hansson argues that the answer is *no*. He abandons the standard approach and opts for a preference-based account instead. The above paper was introduced by Sven Ove Hansson himself, followed by commentaries (Zsófia Zvolenszky, Matteo Pascucci) and further discussed by the seminar participants.

The first official day of the conference was opened by Sven Ove Hansson's invited talk *The Intuitive Base of Deontic Logic*, which naturally continued the topic opened by the pre-conference seminar. How to isolate deontic concepts? How to analyse them formally? Standard analysis and its dyadic variant were introduced, their problems were exposed, and some alternative semantic constructions were considered.

The first contributed talk was given by Robert Trypuz and Piotr Kulicki, presenting about (multi-valued deontic) logics for normative conflicts, whilst self-driving cars were serving as a motivation. The next talk by Piotr Kulicki (a joint work with Xin Sun) was focused on the quantum imperative logic. After that, a sequence of three presentations devoted to hyperintensional deontic logic ensued. The first one was given by Federico L. G. Faroldi and Tudor Protopopescu. They employed Artemov's justification logic to analyse practical reasons. The second one was given by Daniela Glavaničová, who was talking about the hyperintensional logic of responsibility. The third talk was given by Albert Angelberger, who was talking about the analysis of free choice permission within the truth-maker semantics for deontic logic. The topic of the free choice permission reappeared with the talk by Igor Sedlár (a joint work with Frederik Van De Putte). After that, Zsófia Zvolenszky presented on common problems for the analysis of fictional and deontic discourse, focusing mainly on the analysis of conditionals. The last talk of the day was given by Stef Frijters (once more, a joint work with Frederik Van De Putte), who focused on the factual detachment.

The second official day of the conference started with the invited talk given by Olivier Roy, *Dynamic Logic of Power and Immunity*. As the title suggests, the talk was devoted mainly to Hohfeldian notions of power and immunity. Dynamic logic was employed, and the functioning of the model was illustrated by some legal examples.

The Hohfeldian theme continued with the presentation by Réka Markovich, who offered a formal representation of Hohfeld's categories. The next talk was given by Pere Pardo, who analysed obligations as maps on intentions. After that, Alessandra Marra presented about the Miners' Paradox, considering ways of determining when the Reasoning by Cases is a valid principle. The Miners' Paradox reappeared, and the analysis of oughts within the STIT logic without action types

was suggested by Aldo Iván Ramírez Abarca (a joint work with Jan Broersen). And that was the last talk of the day, and the last talk of the conference.

On the behalf of the organising committee it can be revealed that there is an incentive to organize a similar conference again in two years. The idea behind the first volume of this event was to bring deontic logic to Slovakia. However, the success of the event motivated the committee members to continue, and to set up the (more ambitious) goals for the future: to bring together deontic logicians with academics from different areas of logic and philosophy, such as modal metaphysics, theory of fiction, epistemic logic, hyperintensionality, and so on. Therefore, we hope to provide for an intriguing meeting of academics as well as topics in 2019.

*Daniela Glavaničová*

## References

HANSSON, S. O. (2006): Ideal Worlds – Wishful Thinking in Deontic Logic. *Studia Logica* 82, No. 3, 329-336.

## Deflationism in Metaphysics December 15-16, 2017, University of Vienna<sup>1</sup>

Deflationism in philosophy is usually understood as a position according to which to utter a philosophical statement is to invoke that the statement is true, since truth plays no substantial role in philosophy. Deflationism in metaphysics is, by the same manner, to be understood as a position that to state a metaphysical statement is to state that the statement is true, period. Such (a family of) view(s) has provoked a large debate among metaphysicians and, unsurprisingly, the research program gathers philosophers from all corners of the world. The Emmy Noether

---

<sup>1</sup> ✉ Martin Vacek

Institute of Philosophy  
Slovak Academy of Sciences  
Klemensova 19, 811 09 Bratislava, Slovakia  
e-mail: martinvacekphilosophy@gmail.com

Research Group Ontology After Quine: Fictionalism and Fundamentality (Ontologie nach Quine: Fiktionalismus und Fundamentalität), together with the Institute of Philosophy at the University of Vienna, organised an event focused on several aspects of deflationism in metaphysics. Bearing the same name, the workshop put together well-known philosophers working in metaphysics in general, its connection to truth in particular.

A line-up of the workshop included eight speakers. The first talk of the first day was given by Matti Eklund (Uppsala) entitled 'Personites and Existence'. Eklund stressed the significance of metaontological considerations in ethics and metaethics, and generalised it beyond the so-called personite problem. Jade Fletcher's (University of Leeds) 'Truth for Metaontological Deflationists' expressed an inclination to a specific easy-ontology-style conception of truth, given two requirements she identified. The last two talks were given by Tim Button (University of Cambridge) and Esa Diaz-Leon (University of Barcelona), discussing 'A Dogma of Contemporary Metaphysics' (in recorded version) and 'Substantive Metaphysical Debates about Gender and Race: Verbal Disputes and Metaphysical Deflationism' respectively.

The second day commenced with one of the organisers's talk, Delia Belleri (University of Vienna). In 'A Methodological Challenge to Deflationary Ontology', Belleri argued that either easy ontology 'changes the subject', or has the same difficulties as non-deflationary ontology. David Liggins's (University of Manchester) provided the audience with an excellent outline of Nicholas's Jones's argument for second-order realism, as well as well as challenges the argument faces. 'The Abductive Method and the Haecceities Argument for Necessitism' by Lukas Skiba (University of Hamburg) did both, looked at relevant theoretical virtues in the contingentism/necessitism debate, and scrutinised how they square with the Haecceities Argument. It is probably not a coincidence that the last slot belonged to the author of *Ontology Made Easy*. In her 'Conceptual Ethics and the Work of Metaphysics', Amie Thomasson (Dartmouth College) argued that metaphysics-first approaches to conceptual choice face several (epistemological) difficulties and, surprisingly or not, sketched a different one: the pragmatics-first approach.

Deflationism in Metaphysics workshop was a nice and successful event. The topic(s), speakers it hosted, after talks discussions and informal chats over the coffee breaks foreshadowed the direction towards which the debate may move. It is only to be seen whether the debate is on the right track.

*Martin Vacek*

Institute of Philosophy, Slovak Academy of Sciences  
and  
Slovak Philosophical Association

*Modal Metaphysics:  
Issues on the (Im)Possible VI*

**August 2-3, 2018**  
(Bratislava, Slovakia)

**Keynote speakers**

SÒNIA ROCA ROYES (University of Stirling)  
GONZALO RODRIGUEZ-PEREYRA (University of Oxford)

We invite submissions for a 30 minute presentation followed by 10 minute comments and a 15 minute discussion. Areas of interest might include any aspect of analytic metaphysics, epistemology and logic of modality.

A paper of approximately 3000 words should be prepared for blind review and include a cover page with the full name, title, institution and contact information. Papers can be submitted in pdf or doc(x) and should be sent to **modalmetaphysics@gmail.com**. Talks will be followed by commentaries.

Deadline for submission: **March 15, 2018**

Notification of acceptance: **April 30, 2018**

If you wish to submit a paper, comment on an accepted paper or would need any further details, please, email us to the above address, or visit the conference website **www.metaphysics.sk**

---

## CALL FOR PAPERS

for a special issue of  
*Organon F – international journal of analytic philosophy*  
on

### Fictionalism as an (In)Appropriate Philosophical Methodology

Submission deadline: **June 15, 2018**

It is sometimes said that you can't have your cake and eat it too. Fictionalism in various fields of philosophy can be regarded as an attempt to turn this proverb into a falsehood. On the one hand, it is admitted that realistic ontologies of possible worlds, fictional characters, theoretical objects, propositions, numbers and other abstract entities is able to provide elegant explanations in various areas; on the other hand, the realistic ontologies are sometimes challenged as populating the universe with queer entities that bring about insurmountable problems. Fictionalism tries to steer a middle course by retaining the explanatory power of the realistic ontologies but eliminating the problems they might induce. The way in which fictionalism tries to achieve this result is bold—it approaches the realistic ontologies as useful fictions, meaning that they both deliver their explanations of phenomena and are stripped of their burdensome realistic load.

We believe that fictionalism, despite being pursued for some dozens of years, is still very attractive to many philosophers as well as rather worrisome to many others. Considering pros and cons of fictionalism is still capable to provide an opportunity for bringing new arguments and analyses. That is why we decided to prepare a special issue of *Organon F – international journal of analytic philosophy* in which fictionalism will be discussed from various perspectives. Particularly welcome are articles in which fictionalist methodology is applied to various fields like philosophy of modalities, philosophy of mathematics and other sciences, philosophy of fiction, semantics and philosophy of language, to name just a few. Critical articles in

which problems of fictionalism are primarily discussed are welcome as well. Last, but not least, authors of historically oriented articles that are intended to study the origins of fictionalism and its developments are also encouraged to submit their contributions.

All papers submitted to the special issue will be peer-reviewed by at least two anonymous referees. Authors are asked to submit anonymized versions of their papers with a separate file containing author information, affiliation and e-mail address. There are no restrictions regarding the extent of papers. If you would like to ask whether your contribution is relevant to the scope of this special issue, feel free to contact us. All queries and submissions can be sent via the e-mail address **[organonf@gmail.com](mailto:organonf@gmail.com)**.

---