

Contents ▪ Obsah

ARTICLES ▪ STATE

M. Vacek: <i>Impossibilists' Paradise on the Cheap?</i> [in English]	283
I. Hanzel: <i>Data, Phenomena and Mechanisms: The Logic of Construction of Bohr's 1913 Theory of the Hydrogen Atom</i> [in English]	302
S. E. Hatipoğlu: <i>Consciousness and Peripheral Self-Awareness</i> [in English]	323
G. Hoffmann: <i>Infallibility about the Self</i> [in English]	341
M. Kosterec: <i>Elipsa</i> [Ellipsis; in Slovak]	355
A. Blum: <i>On Changing the Past</i> [in English]	377

DISCUSSIONS ▪ DISKUSIE

M. Picha: <i>Externalismus a skeptická výzva</i> [Externalism and Skeptical Challenge; in Czech]	379
M. Taliga: <i>Pravdivosť alebo zdôvodnenie? Odpoveď M. Nublíčkov</i> [Truth or Justification? A Reply to M. Nublíček; in Slovak]	384
M. Duží: <i>Elipsa a anafora</i> [Ellipsis and Anaphora; in Czech]	399

BOOK REVIEWS ▪ RECENZIE

L. Országh: P. Kosso, <i>A Summary of Scientific Method</i> [in Slovak]	413
---	-----

REPORTS ▪ SPRÁVY

M. Vacek: <i>PhiLang 2013, Third International Conference on Philosophy of Language and Linguistics</i> [in English]	418
--	-----

NOEMA – edícia časopisu *Organon F*

Zväzok 1

Marián Zouhar: *Medzi sémantikou a epistemológiou jazyka*
Bratislava: aleph, 2010, 334 strán. ISBN 978-80-89491-01-8

Zväzok 2

Petr Kořátko – Martin Pokorný – Marcelo Sabatés:
Fictionality – Possibility – Reality
Bratislava: aleph, 2010, 290 strán. ISBN 978-80-89491-04-9

Zväzok 3

Barbora Geistová Čakovská: *Vzťahy synonymie.*
Logická sémantika verzus lingvistické teórie o synonymii
Bratislava: aleph, 2010, 156 strán. ISBN 978-80-89491-03-2

Zväzok 4

Marián Zouhar: *Problémy jazyka a počiatky analytickej filozofie*
Bratislava: aleph, 2010, 268 strán. ISBN 978-80-89491-02-5

Zväzok 5

Timothy Childers: *Co je pravděpodobnost?*
Bratislava: aleph, 2011, 228 strán. ISBN 978-80-89491-06-3

Zväzok 6

Marián Zouhar: *Význam v kontexte*
Bratislava: aleph, 2011, 354 strán. ISBN 978-80-89491-07-0

Zväzok 7

Marie Duží – Pavel Materna: *TIL jako procedurální logika.*
Průvodce zvědavého čtenáře Transparentní intensionální logikou
Bratislava: aleph, 2012, 414 strán. ISBN 978-80-89491-08-7

Impossibilists' Paradise on the Cheap?

MARTIN VACEK

Institute of Philosophy, Slovak Academy of Sciences
Klemensova 19, 813 64 Bratislava, Slovak Republic
martinvacekphilosophy@gmail.com

RECEIVED: 27-12-2012 • ACCEPTED: 04-03-2013

ABSTRACT: The aim of the paper is to discuss the ersatz theory of Lewis's impossible worlds, point out its undeniable benefits and demonstrate its costs. Firstly, I present two approaches to Lewis's impossible worlds taken as constructions out of possibilia. Secondly, I evaluate the proposals using the Lewisian criteria of success concerning the well defined conception of analysis. Although appealing, I do not find the proposals fully persuasive. Thirdly, I discuss the objection from an ad hoc distinction between possible and impossible worlds. I conclude that the objection does not present a special problem for the Lewisian theory. Finally, I motivate a theory of extended modal realism, to wit, modal realism enriched with concrete impossibilia.

KEYWORDS: David Lewis – modal realism – impossible worlds.

1. Introduction

Notoriously, David Lewis's theory of possible worlds¹ raises a lot of objections. Besides the incredulous stares on the side of 'common sense' readers, it was argued that it cannot account for phenomena concerning impossible matters. It is because of the fact that, according to Lewis, if we sup-

¹ For the full and comprehensive exposition of modal realism, see Lewis's *magnum opus* (1986).

posed that there is an impossible world, and given the fact how his theory of modality works, we would get a plain contradiction.²

It is, therefore, not a surprise that any advocate of the Lewisian conception of analysis either a) denies the impossible phenomena, b) postulates concrete impossible worlds at the cost of true contradiction in the actual worlds or c) combines the Lewisian approach to modality with the *ersatzist* project, taking possible and impossible worlds as construction out of possibilities of one sort or another. In what follows, I scrutinize the *ersatz* theory of Lewis's impossible worlds and discuss their pros and cons. Finally, I motivate a theory of extended modal realism, the theory that takes the extension of modal realism by impossible worlds seriously.

2. Impossible worlds: Why bother with them?

In his *Counterfactuals*, David Lewis formulates an argument according to which

[i]t is uncontroversially true that things might have been otherwise than they are. I believe, and you do so, that things might have been different in countless ways. But what does that mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the way they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit, 'ways things could have been'. I believe things could have been different in countless ways. I believe permissible paraphrase of what I believe, taking the paraphrase at its face value. I therefore believe in the existence of entities which might be called 'ways things could have been'. I prefer to call them 'possible worlds'. (Lewis 1973, 84)

The argument states at least two things. One of them is the fact that we believe in the ways the world, our actual world, might have been. The other thing is that the ways at issue are entities of some sort or other over which we quantify. If we agree with Quine (1948) and accept that to be is

² Since, for Lewis, worlds are 'are not like stories or story-tellers... They are like this world; and this world is no story, not even a true story', 'at so-and-so world' is a restricting modifier working in the same way as the modifier 'in Bratislava', does. See Lewis (1986, 7, fn. 3).

to be a value of a bound variable, modal notions are to be understood as quantifications over *concrete*³ entities – possible worlds.⁴

3. The second argument and dilemma

However, we can ask, while accepting Lewis's argument from ways, why should we not accept a similar argument for the existence of impossible worlds, namely 'ways the world *could not* have been'? The argument, as it goes, is a variant of the argument from ways obtained by replacement of possibility terms (could, might, possible) with corresponding impossibility terms (could not, might not, impossible). Thus, reformulating the argument, we have:

It is uncontroversially true that things could NOT have been some ways. Thus, there exist many entities of a certain description, to wit, '*ways things could NOT have been*'. I believe things could NOT have been different in countless ways. I believe permissible paraphrase of what I believe, taking the paraphrase at its face value. I therefore believe in the existence of entities which might be called '*ways things could NOT have been*'. I prefer to call them '*impossible worlds*'.⁵

Given Lewis's original argument for the existence of possible worlds, there seems to be no principal difference when applied to impossibilia.⁶ But then, a dilemma arises: either we accept Lewis's argument from ways for

³ According to Lewis, concrete entities are a) things like donkeys and puddles and protons and stars, rather than entities like numbers; b) individuals rather than sets, particulars rather than universals; c) they have spatiotemporal location, enter into causal interaction, are indiscernible one from another and d) are not abstractions from something else (see Lewis 1986, § 1.7).

⁴ Yes, Quine is quite skeptical about possible worlds. It should be noted, however, that it is the lack of criteria of identity of possible worlds that Quine is worried about. When it comes to Lewis, the worry disappears. See Divers (2007).

⁵ The reformulation can be found in Naylor (1986), Yagisawa (1988) and Vander Laan (1997).

⁶ The view that impossible worlds should inherit the metaphysical status of their possible mates is called Parity Thesis. For further arguments against drawing an ontological division between possible and impossible worlds, see Priest (1997, 580-581) or Priest (2005, 139).

neither possible nor impossible worlds, or we do so in both cases. In the former (reductio) case, we get:

1. If Lewis's argument can be applied in the case of possible worlds, then it can be applied, *mutatis mutandis*, in the case of impossible worlds as well.
2. Lewis's argument cannot be applied in the case of impossible worlds.
Therefore
3. Lewis's argument cannot be applied in the case of possible worlds either.

To support the latter alternative – the one advocated in this paper – is to accept the Lewisian argument from ways and apply it uniformly to the case of impossible worlds. The result is a *modus ponens* version which looks as follows:

1. If Lewis's argument can be applied in the case of possible worlds, then it can be applied, *mutatis mutandis*, in the case of impossible worlds as well.
- 2*. Lewis's argument can be applied in the case of possible worlds.
Therefore
- 3*. Lewis's argument can be applied in the case of impossible worlds as well.

A quite different reason for the acceptance of impossible worlds is the so-called granularity problem. Famously, Lewis, following Quine, accepts the identity conditions for sets according to which a set A is identical to a set B *iff* for all members x of A it holds that they are also members of B and vice versa. Also, Lewis accepts a so-called Natural Class Theory (Armstrong 1989) according to which a property is ontologically identified with a set of its instances. But then, some problems arise. Namely, if there are cases where two mutually *distinct* properties are *necessarily* coextensive (including properties had by no Lewis's individual whatsoever), Lewis's theory lacks the resources to distinguish them. For example, is the property of, say, 'being a married bachelor' the same as the property of 'being a round square', namely the empty set? In the same manner, is the property of 'being trilateral' the same as the property of 'being triangular'? If we suppose that a) the properties in fact are different, and b) do not include married bachelors, round squares or trilateral but not triangular individuals into

modal realists' ontology, our pre-theoretical opinions about their distinctness are violated.⁷ Unless we can do better.

4. Concreteness abandoned

Famously, Lewis's denial of genuine *impossibilia* – worldbound individuals instantiating *genuine* impossibilities (Divers 2002, 67) – rests on a simple argument, according to which a contradiction in some impossible world would infect the world we are part of. And since there is no (actual) subject matter about which you can tell the truth by contradicting yourself (Lewis 1986, 7, fn. 3), impossible worlds – at least those concrete – do not exist.

However, one could argue that we can keep the Lewisian *possibilia* and thus save his analyses about possible matters and, at the same time, construct various impossible worlds out of them. Lewis himself notes that he could construct excellent *ersatz* worlds 'drawing on the genuine worlds for raw material':

...there is room for different versions of 'ways things might be' or 'possibilities' or 'structures'. If you mean to include ways that something less than an entire world might be, then again you need sets of possible individuals. If part of the role you associate with these terms is a principle of identity of indiscernibles, you had better take not just any sets but only those that include both or neither of any pair of indiscernible individuals. If, on the other hand, you want to confine your attention to maximal 'ways', and you do not require guaranteed identity of indiscernibles, then you might as well replace unit sets by their sole members. If you associate with 'state of affairs' a role involving predication, I would recommend individual-property pairs, where the property in turn is taken as a set of possible individuals. And so on. (Lewis 1986, 185-186)

If so, *ersatzists* argue, why do not go one step further and construct *ersatz impossibilia* out of the stock of genuine *possibilia*? For example, given the (unrestricted) existence of concrete actual and possible individuals, (*ersatz*)

⁷ Since, for Lewis, propositions are special kinds of properties, the same line of argument applies to the indistinguishability of necessary coextensive propositions considered as sets of whole possible worlds.

impossible worlds could be set-theoretic constructions out of *propositions* or set-theoretic constructions out of *states of affairs*. In any case, the idea that impossible worlds are a free lunch can be seen as serviceable and, given the Lewisian criteria of success, methodologically more appealing than the extension of our ontological commitments.

5. Hybrid modal realism

Following Divers's suggestion (Divers 2002, Chap. 5), Franz Berto (2010) presents one possible solution to the problem of *impossibilia*. In particular, Berto points out that in order to avoid the granularity problem, the world semantics should be extended by impossible worlds (Berto 2010, 471–472). Moreover, the extension should not infect the Lewisian ontology, should sustain the non-modal analysis of modality and stay as conservative as possible.

By accepting the Lewisian ontology – concrete possible worlds and set-theoretic constructions out of them – Berto develops such a modal theory according to which a) *de dicto* possibilities are represented by (genuine) possible worlds, while b) those phenomena of (at least logical) impossibility are represented by *ersatz* entities playing the roles of impossible worlds. Since this account avoids the consequence that concrete individuals instantiate genuine impossibilities, Lewis's objection does not apply here. Thus, one can be a genuine realist about possibilia, while a so-called *ersatzist* about their impossible mates.

The basis of Berto's proposal rests on the Lewisian account of propositions which, briefly, says that propositions A and B are the same propositions *iff* A and B are true at the same possible worlds and, that is the case *iff* A and B are the same set of worlds (cf. Berto 2010, 471). Since, according to Lewis, worlds represent *de dicto* possibilities by *instantiating* propositions directly, basic propositions like 'it is raining', 'it is not raining' or 'there are married bachelors' are simply those sets of concrete possible worlds that the propositions 'it is raining', 'it is not raining' and 'there are married bachelors' hold at them, respectively. The question, however, is whether genuine modal realism has resources for constructing *ersatz* impossible worlds without violating its thorough extensional character. To put the problem otherwise: does genuine modal realism contain the resources for constructing *ersatz* impossible worlds qua sets of propositions? Berto thinks that it does.

In particular, if we consider basic atomic propositions to be sets of concrete worlds, and, using the set-theoretic machinery, consider impossible worlds to be sets of proposition, we get a fully extensional account of possible worlds as well as impossible worlds. Since impossible worlds, or world-books,⁸ are reducible to sets of propositions, and propositions are identified with sets of worlds, impossible worlds can be simply identified with certain sets of sets of concrete possible worlds.

For instance, let us have two distinct contradictory propositions, 'the sun is shining and the sun is not shining' and 'it is raining and it is not raining'. Let us also suppose that the Lewisian logical space consists of exactly six worlds $\{w_1, w_2, w_3, w_4, w_5, w_6\}$. Provided that the proposition 'it is raining', A , is identified with the set $\{w_1, w_2, w_3\}$ and the proposition 'it is not raining', $\sim A$, with the set $\{w_4, w_5, w_6\}$, the contradictory proposition 'it is raining and it is not raining' – $(A \text{ and } \sim A)$ – is, by the same reasoning, identified with the set of the above sets, namely $\{\{w_1, w_2, w_3\}, \{w_4, w_5, w_6\}\}$.⁹ The resultant set is an impossible world, i_1 , because it represents an inconsistent situation. Now, let us also suppose that the proposition 'the sun is shining', B , is identified with the set $\{w_1, w_3, w_5\}$ and its negation, $\sim B$, with $\{w_2, w_4, w_6\}$. Similarly, the contradictory proposition 'the sun is shining and the sun is not shining' – $(B \text{ and } \sim B)$ – is then the set $\{\{w_1, w_3, w_5\}, \{w_2, w_4, w_6\}\}$. Let call the impossible world i_2 . No doubt, impossible worlds i_1 and i_2 are different. While i_1 is identified with the set of the form $\{\{w_1, w_2, w_3\}, \{w_4, w_5, w_6\}\}$, the form of i_2 is quite different – $\{\{w_1, w_3, w_5\}, \{w_2, w_4, w_6\}\}$.

Next, beside the distinctness of (any two) contradictory propositions, Berto's approach can represent various impossible situations where there are three or more true, but mutually incompatible distinct atomic propositions. To use Berto's own example, the propositions 'Nassau Street runs east-west', 'the Railroad nearby runs north-south' and 'Nassau Street is parallel to the railroad nearby' are mutually incompatible. We can, however, construct such an *ersatz* impossible world which represents the impossibility.

⁸ The term 'world-book' is from Plantinga (1974, especially Chap. 4).

⁹ Here, Kiourti (2010, 80) quite correctly notes that if the aim was to differentiate between different contradictions true at some *ersatz*-world, we will have to define conjunction and, consequently, all logical connectives recursively. That means, however, that such an approach commits its proponents to a particular logical behavior for the logical connectives at such impossible worlds.

Namely, let us suppose that the proposition ‘Nassau Street runs east-west’ holds in w_1, w_3, w_6 , the proposition ‘the Railroad nearby runs north-south’ holds in w_2, w_3, w_5 and the proposition ‘Nassau Street is parallel to the railroad nearby’ hold just in the world w_6 . Any world at which all of them hold must be impossible, because there is no Lewis’s world in which all of them hold. But that is no problem anymore. We can represent the impossibility, nonetheless: $\{\{w_1, w_3, w_6\}, \{w_2, w_3, w_5\}, \{w_6\}\}$. End of theory (Berto 2010).

A slightly different approach to *ersatz* theory proposes Edwin Mares (Mares 1997). Trying to stay ontologically as neutral as possible, Mares assumes there to be a set of all possible worlds (whatever they are), individuals existing in those worlds, relations, a possibility to collect those individual into a unique set and, finally, sets themselves (Mares 1997, 519).

Then, Mares introduces states of affairs, meaning structures having the following form:

$$\langle R, a_1, \dots, a_n, \pi \rangle$$

Here, R is an n -place relation, a_1, \dots, a_n are individuals, and π , called polarity, is either 1 or 0. The mention of 1 and 0 indicates, that while some of the states of affairs obtain, others do not. However, the story is even more complicated, because unless we want to face the idea of negative states of affairs, we might think of the structure as a piece of information rather than a state of affairs (see Kiourti 2010, 84). In this case, informational sequences can tell us whether some relation R does or does not hold between the individuals in the sequence. This, subsequently, allows us to delineate those negative states of affairs that fail to obtain from those whose negations obtain instead, and thus fulfill Mares’s desideratum that negative information does not reduce to the absence of positive information (Mares 1997, 520). Thus, the information that it is not raining in Auckland does not reduce to not having the information that it is raining in Auckland.

The next step in Mares’s strategy is the introduction of an index. According to him, an index is a (maximal) set of states of affairs such that for any n -place relation R and any sequence of individuals a_1, \dots, a_n , either $\langle R, a_1, \dots, a_n, 1 \rangle$ is in the index, $\langle R, a_1, \dots, a_n, 0 \rangle$ is in the index or both states of affairs are in the index. The indices represent worlds. Although not any index represents a possible world, any index represents some world. For, for all states of affairs

- (M) $\langle R, a_1, \dots, a_n, 1 \rangle$ and $\langle R, a_1, \dots, a_n, 0 \rangle$
 $\langle R, a_1, \dots, a_n, 1 \rangle \in i$ iff $\langle a_1, \dots, a_n \rangle$ is in the extension of R at w
 and
 $\langle R, a_1, \dots, a_n, 0 \rangle \in i$ iff $\langle a_1, \dots, a_n \rangle$ is not in the extension of R at w .

Now, how can the theory help the Lewisian to accommodate impossible worlds? Naturally, the easiest way is to consider Mares's individuals as genuine full-blooded individuals and relations (including properties) as sets of ordered individual pairs (or just sets of individuals in the case of properties). Polarity could be understood as an inclusion of the individual(s) at issue in the sets. For instance, an outright contradiction 'it is raining and it is not raining' can be easily represented by an index $\langle \langle R, a, 1 \rangle, \langle R, a, 0 \rangle \rangle$, where a stands for a world and R stands for the set of worlds such that it is raining at them. While the former sequence of the index is positive, thus representing the case of raining in a , the latter is apparently its negation. Consequently, we get an index representing logical contradiction and, a fortiori, an impossible world, without going beyond the Lewisian ontology.

Similarly, the contradictory proposition 'the sun is shining and the sun is not shining' can be represented by an index $\langle \langle S, a, 1 \rangle, \langle S, a, 0 \rangle \rangle$, where, as in the previous case, a stands for a particular world, S stands for the property of shining and 1 and 0 for the inclusion and exclusion of the respective world in the respective set-theoretic construction. Since the indices $\langle \langle R, a, 1 \rangle, \langle R, a, 0 \rangle \rangle$ and $\langle \langle S, a, 1 \rangle, \langle S, a, 0 \rangle \rangle$ are obviously distinct,¹⁰ the apparatus is strong enough to differentiate between various purely logical impossibilities. That is Mares's theory in a nutshell (cf. Mares 1997).

6. Evaluating hybrid modal realism

In sum, it is undeniable that both of the above proposals present a strong improvement of genuine modal realism. As it seems, they are more conservative than the original theory, because they can a) account for more of our pre-theoretical opinions than the original theory and, at the same time, b) do not extend the Lewisian ontology. All the work is done by the movements in the definitions, namely the definition of impossible world.

¹⁰ Apparently, the distinctness rests on the fact that S and R are distinct sets of worlds. While S denotes the set of worlds being such that the sun is shining at them, R denotes the set of 'rainy' worlds.

Also, the proposals have certain theoretical as well as methodological virtues. Firstly, the reduction of impossible worlds to the Lewisian individuals and sets is fully extensional. At the base level, we have individuals, sets, and set-theoretic constructions out of them. Then, impossible worlds – being certain set-theoretic constructions – are a free lunch, ‘lying around the house of almost any possible world theorist’ (Mares 1997, 525). Secondly, given the fundamental difference between possible and impossible worlds and, a fortiori, a non-modal distinction between them, we can sustain the very definition of possibility without any resort to primitive modality. Namely, the Lewisian dictum

(P) It is possible that P *iff* there is a world, w, such that at w, P

still holds, while the quantifier in the right-hand side ranges only over the maximal mereological sums of spatiotemporally interrelated individuals. On the other side, the impossibility claims get different interpretation. Instead of the Lewisian account of impossibility

(I) It is impossible that P *iff* there is no possible world, i, such that at i, P,

we have the *ersatzist* version, namely

(I*) It is impossible that P *iff* there is a world, i, such that at i, P

while the domain of the right hand side quantifier comprises sets of sets of concrete possible worlds only.

Finally, and most importantly, allowing there to be *ersatz* impossible worlds constructed out of concrete individuals avoids Lewis’s lethal objection that contradiction obtaining at those worlds amounts to a plain contradiction in the actual world. Recall that Berto’s and Mares’s impossible worlds *are* books – either some sets of propositions or some sets of informational sequences – and so although it still holds that the phrase ‘at (concrete world) w’ in (P) works as a modifier restricting the (explicit or implicit) quantification to particular domains, the phrase ‘at i’ in (I*) works quite differently. For, Berto’s as well as Mares’s *ersatz impossibilia* are abstract set-theoretic entities in the case of which the phrase ‘according to some *ersatz* impossible world’ is not a restricting modifier. For that reason, there is an impossible world *according to which* a contradiction is true without the contradiction infecting the actual world. And those are very good results.

7. Ad hoc ontological distinction?

As indicated above, some virtues of the proposals above are remarkable and that is the reason why we should take them seriously. Berto, as well as Mares, are successful in their analyses of modality in non-modal terms. They also score in distinguishing (at least some) apparently (not only logically impossible) propositions, without any extension of the Lewisian ontology. Moreover, they do not (radically) contract the Opinion and thus preserve conservativeness without any violation of economy.

It was argued by several philosophers, however, that the fundamental difference between possible and impossible worlds is not justified. According to Lewis, the actual world is not an abstract entity. So why other worlds should be? By the parity of reasoning, possible worlds are not abstract entities so why impossible worlds should be? The possible objection to the above proposals, therefore, is whether any approach deeming possible worlds as concrete universes, while deeming impossible worlds to be mere *ersatz* constructions, does not pose an ad hoc and, a fortiori, unjustified distinction.

In a similar vein, Graham Priest writes:

... there is, as far as I can see, absolutely no cogent (in particular, non-question begging) reason to suppose that there is an ontological difference between merely possible worlds and impossible worlds – any more than there is for supposing there to be such a difference between merely possible worlds which are physically possible and those which are physically impossible. To differentiate between some non-actual worlds and others would seem entirely arbitrary. (Priest 1997, 582)

As the passage concludes, to make a fundamental distinction between possible and impossible worlds does not seem to be justified without begging the question. Also, David Vander Laan (1997) argues that such a slight difference in our language as the difference between 'the way the world could be' and 'the way the world could *not* be' should not justify such a radical difference between possible and impossible worlds. He asks:

Is there a principle of ontology which would justify our construing these similar parts of our modal language in such dissimilar ways? (Vander Laan 1997, 600)

However, any attempt to weaken the hybrid theories just by raising the objection from an ad hoc difference must fail. Granted, it may seem arbi-

trary or unjustified to start with the presumption that possible worlds and impossible worlds are quite different kinds of things. But if the best overall theory plays out that way – that is, if conservativeness-cum-economy is maximised by identifying possible worlds with one kind of construct and impossible worlds with another – the putative intuition for or against the parity has nothing to do with it.

Thus, it seems that Vander Laan is plainly wrong when saying that there is not such a principle of ontology that would justify our construing these similar parts of our modal language in such dissimilar ways. Of course, there is such a principle: the parity or imparity thesis is justified if that is how the best analysis plays out. Put otherwise, it is not an input of any analysis whether the parity thesis is justified or not. On the contrary. There is nothing wrong in saying that impossible worlds are ontologically distinct from possibilities, given that the theory posing the distinction does not lead to trouble while the theory that poses the parity does. What we as philosophers are looking for is the ideal ‘definitions + ontology and ideology’ package combining a) the best balance of conservativeness and economy b) the preservation of all (or almost all) of our pre-theoretical opinions and c) the sufficient advantages to its rivals. An opinion whether the parity thesis holds or not plays no role at all.

I, therefore, conclude that the objection to hybrid versions of modal realism that rests on a putative illegitimacy of the ontological distinction between possible and impossible worlds is not a good way to dismiss the proposals. Since it is the objection itself, not what is objected, that is unjustified by the Lewisian well-defined conception of analysis, it does not present serious troubles for the above approaches. Since it is an input, rather than an output of the analysis how the nature is carved along the joints, there is simply no non-question-begging reason to think that possible and impossible worlds are of the same metaphysical kind. It is fully in the competence of the total theory to decide.¹¹

¹¹ A quite different objection would be to say that some contradictory worlds are abstract, even if we think that they are possible. Then, the ontological distinction between various possible worlds would be, of course, illegitimate.

8. Why still bother with concrete impossibilia?

Given the proposals above, one would think that some hybrid theory of genuine modal realism is the most reasonable way to go. Since both proposals have sources to represent outright contradictions, they can distinguish between various contradictory propositions, provide non-trivial interpretations of counterpossible conditional, differentiate between impossible propositional states and so forth. Also, at first sight, the proposals at issue have certain similar features. As Berto distinguishes between contradictory propositions of the form $(A \text{ and } \sim A)$ and $(B \text{ and } \sim B)$ by constructing impossible worlds out of some sets of possible worlds, Mares constructs his impossible worlds out of sets of possible worlds, sets of individuals and the polarity.

But why does Mares need those structured states of affairs instead of mere sets of worlds? The answer is clear: there are not only outright contradictions among the impossibilities that we intuitively distinguish. There is much more. What about such impossibilities as 'there are married bachelors' or 'some triangular individuals are not trilateral', etc.? Those impossibilities are not purely logical impossibilities, since they are not of the form $(A \text{ and } \sim A)$.

Granted, Berto's theory gives us various representations of different impossibilities by using genuine worlds. However, there are several limits the hybrid theory has. Recall, that one of the motivations for *impossibilia* was the need to differentiate between various necessary coextensive and impossible propositions, and thus provide a finer-grained distinction absent in the original theory. For, according to Lewis, the property of 'being trilateral' is identified with the property of 'being triangular', because any (actual and possible) individual which is trilateral is also triangular, and vice versa. But the problem seems to persist even in these theories.

To see the problem, let us have two (distinct) propositions, 'Toby is triangular' and 'Toby is trilateral'. Apparently, the aim is to show that the propositions are different, and thus that the proposition 'Toby is triangular but not trilateral' can be represented by an (*ersatz*) impossible world distinct from a world according to which 'Toby is triangular but not triangular'. Provided Berto's strategy, and given a simple logical space consisting of three worlds – w_1, w_2, w_3 – we can identify the proposition 'Toby is trilateral' with, say, the set of worlds $\{w_1, w_2\}$ and its negation with $\{w_3\}$. Given the Lewisian pluriverse, the same holds for the proposition 'Toby is trian-

gular', namely that the proposition is identified with the set $\{w_1, w_2\}$,¹² while its negation with $\{w_3\}$. But then the impossible proposition 'Toby is triangular but not trilateral' – $\{\{w_1, w_2\}, \{w_3\}\}$ – is not different from the impossibility 'Toby is triangular but not triangular' $\{\{w_1, w_2\}, \{w_3\}\}$. But we wanted to say that!

Also, our initial motivation to differentiate between various propositions holding at no (possible) world is not fulfilled here. For instance, the propositions 'there is a married bachelor' or 'there is a non-identical individual' are still conflated in Berto's account. Since there are neither worlds such that there are married bachelors in them, nor worlds such that non-identical individuals exist in them, we simply lack the 'raw material' to delineate the distinction.

Berto, being aware of the problems, admits that everything goes fine when we want to represent impossibilities which are adequately phrased as explicit contradictions (Berto 2010, 484). For, such impossibilities are of the conjunctive form (A and $\sim A$), resp. (B and $\sim B$). The problems arise when it comes to representing distinct (non-conjunctive) impossibilities. That means that although Berto's proposal does differentiate between various distinct impossible propositions, namely those which are the results of conjoining two (basic) propositions that are not true together in any Lewis's world, it does not allow us to differentiate between impossibilities being not of the conjunctive nature.

Mares's approach, on the other side, is not limited in this way. Since the main point of the proposal is that it allows us to attribute properties and relations to individuals that do not have them in any Lewis's world (Kiourti 2010, 85), it can accommodate several subtler impossibilities. Not only every outright contradiction of the type (A and $\sim A$), but also impossibilities like 'something is blue and yellow all over' are analyzable in the theory. For example, an impossibility of there being something, say car, that is blue and green all over is represented by a world of the following form: $\langle\langle B, a, 1 \rangle, \langle Y, a, 1 \rangle, \langle C, a, 1 \rangle\rangle$. Here, a stands for an individual, B stands for the property of 'being blue', Y for the property of 'being yellow' and C for the property of 'being a car'. It is because of the fact that the set of all (unrestrictedly speaking) blue things is different from the set of all (unrestrictedly speaking) yellow things as well as from the set of all (unrestrictedly

¹² It is because of the fact that there is no Lewis's world, w , such that something is triangular but not trilateral in w .

speaking) cars. Similarly, to use Berto's own example, the impossibility of there being negatively and positively charged particle and a quite different impossibility of there being positively and not positively charged particle can be distinguished, since the former world – $\langle\langle E, a, 1 \rangle, \langle P, a, 1 \rangle, \langle N, a, 1 \rangle\rangle$ – differs from the latter: $\langle\langle E, a, 1 \rangle, \langle P, a, 1 \rangle, \langle P, a, 0 \rangle\rangle$. Here, E stands for the property of 'being an electron', P and N correspond to the properties of 'being positively charged' and 'being negatively charged', respectively.

So far so good. However, even this approach has certain limitations. For example, it cannot deal with some necessarily coextensive properties and propositions. For example, let have the property of 'being trilateral' and the property of 'being triangular'. If we identify the properties with their instances we, unfortunately, do not get two *distinct* sets, namely the set of all trilateral individuals and the set of all triangular genuine individuals. Since there is no world in the Lewisian pluriverse that contains a trilateral but non-triangular individual, the sequences $\langle A, a, 1 \rangle$ and $\langle S, a, 1 \rangle$, where S stands for the property of 'being trilateral' and A stands for the property of being 'triangular' will be identified with the same sequence. Simply, if A and S denote the same set, the indices $\langle\langle A, a, 1 \rangle, \langle S, a, 0 \rangle\rangle$ and $\langle\langle A, a, 1 \rangle, \langle A, a, 0 \rangle\rangle$ represent the same (impossible) world. But the impossibilities 'Toby is triangular but not trilateral' and 'Toby is triangular and not triangular' are not the same.

Moreover, even this proposal does not express the intuitive difference between various (distinct) impossible properties. Recall, that what any theorist preferring this proposal has at disposal is the Lewisian ontological base. As we already know, it includes (only logically consistent) individuals and set theoretic constructions out of them, none of which is non-self-identical or contradictory. Yet, all of them are self-identical and non-contradictory. But if we want to express the difference between the properties, we have to take the set of non-self-identical individual and the set of contradictory individuals as the constituents of the sequences. Consequently, we get the sequences $\langle N, a, 0 \rangle$ and $\langle C, a, 0 \rangle$, respectively. Given that the letters N and C stand for the same set – the set of all non-self-identical individuals and the set of all contradictory individuals, respectively – the sequences are identical.

On the other hand, every individual in the Lewisian ontology is consistent as well as every individual is such that it is identical to itself. Both of the properties are thus identified with the same set of individuals – the whole plurality of possible individuals. Thus, if I stands for the property of

being ‘identical to itself’ and C for the property of ‘being a consistent individual’, the sequences $\langle I, a, 1 \rangle$ and $\langle C, a, 1 \rangle$ are identical, since the letters I and C denote the same set. Then, the propositions ‘something is self-identical but not consistent – $\langle \langle I, a, 1 \rangle \langle C, a, 0 \rangle \rangle$ – and ‘something is self-identical and not-self-identical – $\langle \langle I, a, 1 \rangle \langle I, a, 0 \rangle \rangle$ – are, due to the identity between I and C, the same. The intuitive difference is lost.

The reason, as far as I can see, is that neither Berto’s, nor Mares’s proposal admits there to be genuine *impossibilia* at the ontological base. Then, it is just to be expected that certain kinds of properties – especially those instantiated by every or no individual in the ontological base – will have to be identified. Thus, although the raw material used in the constructions is extensive enough for a variety of distinctions, the fact that it does not contain full-blooded *impossibilia* does bring certain limitations (see Kiourti 2010, 82).

Therefore, whether the project of construction of *ersatz* impossible worlds out of concrete *possibilia* is successful is one thing. How we get into it is quite another. To look at the problem from the Lewisian well-defined conception of analysis, it is very easy to trace the (rather similar) strategies. Berto and Mares stress the fact that we pre-theoretically presuppose the difference between impossible phenomena and thus our theory should reflect it. They also accept and (definitely do not extend) the ontology and so propose to modify the definition of impossibility. And that is a huge virtue of the proposals. On the other side, we are still left with a variety of unresolved problems, namely the inability to differentiate among certain impossible properties, the inability to differentiate among certain necessary properties or even necessarily coextensive properties. Thus, as usual, the question is whether the advantages of the above proposals outweigh their costs on one side, preserve the Lewisian crucial features, on the other.

Next, it would seem that the existence of genuine (concrete) *impossibilia* – instead of mere *ersatz* constructions out of *possibilia* – seems to be such as to provide a more unified, simpler and more systematic treatment of modal matters (see Yagisawa 1988). Surely, the hypothesis of the extended modal realism is just the beginning. It can turn out that the extension of the Lewisian ontology by genuine *impossibilia* would harm the theory to such an extent that no other moves in the pre-theoretical opinions, the definitions or the ontology would ‘cure’ the consequences. Yet, it is still worth of considering how the proposal to extend Lewis’s ontology by concrete *impossibilia* squares with the overall theory.

For modal realist, a certain sentence is true at a world *iff* it is true when we quantify over all the things in that world. By the same token, when Lewis advocates the existence of mere possibilia, he strictly differentiates between actual truths and truths *simpliciter*. For example, the claim

- (1) There is no beer

can be true, when quantifying over the empty fridge, although false when quantifying over, say, the whole country. Similarly, the claim

- (2) There are flying pigs

is false provided that we take the actual world into the account. Hence:

- (3) Flying pigs do not *actually* exist

is true. That does not, however, mean that the flying pigs do not exist *simpliciter*. Of course, sometimes it is reasonable to ignore them and quantify over us and our surroundings only. But to ignore something does not imply that the entity at issue does not exist.¹³ In other words, Lewis denies (A):

- (A) *Actually* P *iff* (unrestrictedly) P,

where P stands for a proposition such as, say, 'there are flying pigs'. Since there are no flying pigs in the actual world, the left-hand side of the biconditional is false. But, as concretist's ontology dictates (Lewis 1986, § 1.1), individuals of this kind *do* exist, none the less.

Interestingly enough, Lewis's strategy changes when it comes to impossibilities. In comments on Lycan, he states: "He [Lycan] is not suggesting that I claim to quantify over incomplete or inconsistent Meinongian objects – of course I do not – but only that I claim to quantify beyond actuality" (Lewis 1986, 98). Therefore:

- (N) *Necessarily* P *iff* (unrestrictedly) P¹⁴

is, for Lewis, true. For instance, it is a truly remarkable fact that the Law of Non-Contradiction is true *simpliciter* because it holds with respect to any *possible* world. Analogously, the Law of Non-Contradiction is true necessar-

¹³ That is, I think, the moral Lewis's modal realism implies.

¹⁴ For a similar strategy, see Kiourti (2009, Chap. 4).

ily, because there is no *possible* world which would violate the principle. As a result:

- (4) *Necessarily* the Law of Non-Contradiction holds *iff* the Law of Non-Contradiction holds *simpliciter*

is true in traditional Lewisianism.

But, as we have already learned,¹⁵ impossibilia may be objects of beliefs, counterfactuals with impossible antecedents are not all trivially true, there are different impossible properties and propositions etc. Thus, it seems as (4) is false. Since there are (excellent) reasons for proceeding to wide quantification so as to include *possibilia as well as impossibilia* within the scope of logic (Routley 1980, 83), to widen the scope of quantification beyond *possibilia* seems nothing but a regimentation of our pre-theoretical modal talk.¹⁶ Consequently

- (5) *Necessarily* the Law of Non-Contradiction holds *iff* the Law of Non-Contradiction holds when quantifying over *possibilia*

is true, while

- (6) *Necessarily* the Law of Non-Contradiction holds *iff* the Law of Non-Contradiction holds when quantifying over *possibilia and impossibilia*

is false.

To conclude, we have seen that Lewis's possibilist's strategy rests on the difference between two kinds of truths, namely *actual* truths and truths *simpliciter*. However, Lewis ignores the analogous difference, i.e. the difference between *necessary* truths and truths *simpliciter*. But, by noting the very similarity between the principles (A) and (N), Lewis's commitment to the latter principle seems question-begging.¹⁷ More accurately: Why do not stretch an extra mile and provide a similar argument for the existence of *impossibilia* as we did in the case of mere *possibilia*? Surely, the violation of the Law of Non-Contradiction is a tricky move in any theory. But in order

¹⁵ To mention just few *prima facie* contributions to the debate, see Yagisawa (1988), Priest (1997), Berto (2008; 2010), Kalhat (2008) or Kiourti (2009).

¹⁶ How such logic could look like, see Routley (1980), Kiourti (2009) or Berto (2010).

¹⁷ That Lewis begs the question when it comes to *impossibilia*, see also Kalhat (2008).

to deal with inconsistency (and indeterminacy) logically, what else could we expect. Once we change (or rather liberalize) semantics to the extent that it accommodates possible and impossible individuals without falling into an utter nonsense, the impossible phenomena can be treated on a logical basis.

References

- ARMSTRONG, D. M. (1989): *Universal: An Opinionated Introduction*. Westview Press.
- BERTO, F. (2009): Impossible Worlds. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*. Available at: <<http://plato.stanford.edu/archives/fall2009/entries/impossible-worlds/>>.
- BERTO, F. (2010): Impossible Worlds and Propositions: Against the Parity Thesis. *Philosophical Quarterly* 60, No. 240, 471-487.
- DIVERS, J. (2002): *Possible Worlds*. London: Routledge.
- DIVERS, J. (2007): Quinean Scepticism about De Re Modality after David Lewis. *European Journal of Philosophy* 15, No. 1, 40-62.
- DIVERS, J. (forthcoming): The Analysis of Possibility and the Extend of Possibility.
- KALHAT, J. (2008): Has Lewis Reduced Modality? *European Journal of Philosophy* 17, No. 4, 504-526.
- KIOURTI, I. (2010): *Real Impossible Worlds: The Bounds of Possibility*. PhD thesis, University of St. Andrews.
- LEWIS, D. K. (1973): *Counterfactuals*. Oxford: Blackwell.
- LEWIS, D. K. (1986): *On the Plurality of Worlds*. Oxford: Blackwell.
- LEWIS, D. K. (1991): *Parts of Classes*. Oxford: Blackwell Publishers.
- MARES, E. (1997): Who's Afraid of Impossible Worlds? *Notre Dame Journal of Formal Logic* 38, No. 4, 516-526.
- NAYLOR, M. (1986): A Note on David Lewis's Realism about Possible Worlds. *Analysis* 46, No. 1, 28-29.
- QUINE, W. V. O. (1948): On What There Is. *Review of Metaphysics* 2, 21-38.
- PLANTINGA, A. (1974): *The Nature of Necessity*. Oxford: Clarendon.
- PRIEST, G. (1997): Sylvan's Box. *Notre Dame Journal of Formal Logic* 38, No. 4, 481-487.
- PRIEST, G. (2005): *Towards Non-Being: the Semantics and Metaphysics of Intentionality*. Oxford: Oxford University Press.
- ROUTLEY, R. (1980): *Exploring Meinong's Jungle and Beyond. An Investigation of Noneism and the Theory of Items*. Canberra: Research School of Social Sciences, Australian National University.
- VANDER LAAN, D. (1997): The Ontology of Impossible Worlds. *Notre Dame Journal of Formal Logic* 38, No. 4, 597-620.
- YAGISAWA, T. (1988): Beyond Possible Worlds. *Philosophical Studies* 53, No. 2, 175-204.
- YAGISAWA, T. (2010): *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.

Data, Phenomena and Mechanisms: The Logic of Construction of Bohr's 1913 Theory of the Hydrogen Atom¹

IGOR HANZEL

Department of Logic and Methodology of Sciences. Faculty of Philosophy
Comenius University in Bratislava. Šafárikovo nám. 6. 814 99 Bratislava. Slovak Republic
hanzel@fphil.uniba.sk

RECEIVED: 10-09-2012 • ACCEPTED: 14-03-2013

ABSTRACT: The paper aims at a unification of the two directions in contemporary philosophy of science: the direction which deals with the relation of data to phenomena with the direction which deals with the knowledge about mechanism and its employment in scientific thinking. It aims also at a reconstruction of the development of scientific knowledge which is characterized in contemporary philosophy of science as a movement from data, via phenomena, to mechanisms. An attempt will be made to show that this in fact amounts to an assignment of philosophical categories like data, phenomena, mechanism, etc. This unification and reconstruction draws also on the reconstruction of the main stages of the development of knowledge leading from A.-J. Ångström's measurement of the wave-lengths of spectral lines of hydrogen in 1868 to N. Bohr's theory of the hydrogen atom proposed in 1913.

KEYWORDS: Bohr's hydrogen atom – data – laws of phenomena – mechanisms – phenomena.

1. Introduction

This paper has two aims. The first is to unify two directions in contemporary philosophy of science: a) the direction which draws on the

¹ This paper was written with the support of the grant VEGA, number V/11/092/00.

seminal works of J. Bogen and J. Woodward to deal with the relation of data to phenomena and b) the direction which utilizes here the article Machamer – Darden – Craver (2000) to deal with the knowledge about mechanism and its employment in scientific thinking. This unification aims to delineate the main stages of the development and growth of scientific knowledge via thinking and shall draw on the results achieved in the framework of philosophy of science in the last 20 years or so, as well as on the reconstruction of the main stages of development of knowledge leading from A.-J. Ångström's measurement of the wavelengths of the spectral lines of hydrogen in 1868 to N. Bohr's theory of the hydrogen atom in proposed 1913.

The second aim, located at a "meta" level, is to show that the reconstruction of the development of scientific knowledge (which is characterized in contemporary philosophy of science as a movement from data, via phenomena, to mechanisms) in fact amounts to an assignment of philosophical categories (e.g., *data*, *phenomena*, *mechanism*, etc.) to the respective stages of this development. From this "meta" point of view it will become obvious that our choice of the above-given episode in the history of physics can contribute to the explication of these categories, and at the same time, by bringing in additional categories, serve to broaden the network of categories by means of which philosophy of science approaches and reflects the natural sciences. The categories that we will introduce in addition to those used in the more recent philosophy of science literature (e.g., data, phenomena, and mechanism) are as follows: *phenomena as derived from the law of phenomena*, the latter being derived from the knowledge of and by thinking about a mechanism, and where these phenomena and laws of phenomena have a different epistemological status as compared to those *phenomena derived from data*, and from *laws of phenomena derived from these latter phenomena*. With respect to the category mechanism we will also introduce the additional category pairs *ground-grounded* and *reason-reasoned*.

2. J. Bogen, J. Woodward and P. Machamer on data, phenomena, and mechanisms

J. Bogen and J. Woodward in their seminal works criticize "those who hold that scientific theories explain what we observe and who then go on to tie the relevant notion of observation rather closely to sensory perception" (Bogen – Woodward 1988, 306). In opposition to those who ap-

proach scientific theories and explanations in this way, they propose a differentiation between the category *data* and the category *phenomena* as follows (1988, 305-306):

Data ... play the role of evidence for the existence of phenomena, for the most part [they] can be straightforwardly observed. However, data typically cannot be predicted or systematically explained by theory. By contrast, well developed scientific theories do predict and explain facts about phenomena. Phenomena are detected through the use of data, but in most cases are not observable in any interesting sense of the term² ... Facts about phenomena ... are evidence for the high-level general theories by which they are explained.

So, in their view, while claims about data serve as evidence for claims (facts) about phenomena, claims about phenomena serve as evidence only for claims which are already part of general theories explaining/predicting claims about phenomena but not about data. Stated in a more general way: “we need to distinguish what theories explain (phenomena or facts about phenomena) from what is uncontroversially observable (data)” (Bogen – Woodward 1988, 314). Data can also be subjected to explanations, but because they are produced by irregular coincidences of myriad particular causes, such explanations:

when they can be given at all, will be highly complex and closely tied to the details of particular experimental arrangements ... Thus, explanations of data will often lack generality ... Moreover, the factors involved in the production of any given bit of data may be so disparate and so numerous, and their co-occurrence so rare, that the details of their interaction may be both epistemically inaccessible and difficult to model theoretically. Exhibitions of dependency-relations of the sort that would be achieved by explicit derivations or the tracing of specific causal mechanisms may prove impossible because of computational intractabilities. (Bogen – Woodward 1988, 326)

² This claim was recently restated as follows: “phenomena need not be observable ... asking whether phenomena are observable is often not the right question to understand how ... reasoning works. This is because the reliability of ... reasoning often has little to do with how human perception works” (Woodward 2011, 171).

For the relation of data to phenomena it holds that the latter are derived from the former:

The problem of detecting a phenomenon is the problem of identifying a signal in a sea of noise, of identifying a relatively stable and invariant pattern of some simplicity and generality with the recurrent features – a pattern which is not just an artifact of the particular detection technique we employ or the local environment in which we operate. (Woodward 1989, 396-397)

Bogen and Woodward also explain, as already stated above, the difference between the category *phenomena* and the category *data* in terms of the category *observation*. While the former need not be observable, the latter can usually be directly observed in the following sense:

Data are *records* or *reports* – accessible to the human perceptual system and available for public inspection. Some data (e.g., reports of measurement results written in laboratory notebooks, or the drawings of the field geologists) are produced by human perceivers. But ... many data are produced by nonhuman measurement and recording devices ... Data constitute observational evidence to investigate phenomena. (Bogen – Woodward 1992, 593)

So, for example, they claim that when measuring the temperature at which a sample of lead melts, the value – the datum – of each thermometer reading is observable (cf. Bogen – Woodward 1988, 319). The following diagrammatic summary of the views of Bogen and Woodward can be given (cf. Schindler 2007, 167):

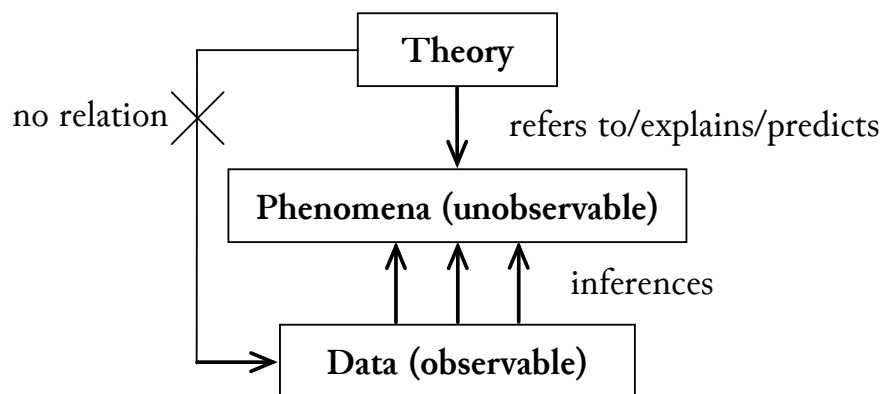


Diagram 1: S. Schindler on Bogen's and Woodward's views on data, phenomena and theory

What “Theory” in this diagram stands for can be understood if one takes into account the fact that Bogen and Woodward characterize the process of systematic explanation to which claims about phenomena are subjected as “explanation of an outcome due to the assertions of the details of a mechanism” (Bogen – Woodward 1988, 323). The category of mechanism was delineated by means of additional categories explicated in Machamer – Darden – Craver (2000) and Darden (2008). In Machamer – Darden – Craver (2000) one can identify the following triple of categories falling under the category *working of the mechanism*: *start-up/set-up conditions*, *intermediate activities*, and *terminating activities*. The first category involves the following subordinated categories: *entities*, *relevant properties of entities given at the beginning of the activities*, *the mutual interactions of these entities*, *entities and activities as mutually correlative*, *the spatial distribution, orientation, and relations of the entities and activities*, *the temporal characteristics of the activities*. To the category *intermediate activities* are subordinated categories *intervening entities and activities*, understood as *producers of new states in entities and/or of new entities and activities*. The third category involves categories *endpoint of activities*, *parameters and state of the mechanism*.

This triple of categories with their subordinated categories, once unified with the set of categories given by Darden (2008, 965), can be summarized in Table 1.³

Main categories					
		<i>Components</i>	<i>Spatial arrangement of components</i>	<i>Temporal aspects of components</i>	<i>Contextual locations</i>
Subordinated Categories		set-up/start-up & terminating conditions, entities, initial & intermediate & terminating activities, modules	localization structure orientation connectivity compartmentalization	order rate duration frequency	location within a hierarchy location within a series

Table 1: Categories involved in the category *mechanism*

³ The category *module* stands for reasoning about groups of components of the mechanism.

3. From Ångström to Bohr and “back”

3.1. From data to phenomena: A.-J. Ångström

We now start to deal with a level of scientific knowledge to which A. Sommerfeld assigned, in a backward glance, the category – when expressed in German as *empirisches* (Sommerfeld 1921, 222), understood in English as *empirical data* (Sommerfeld 1923, 68).

Ångström’s aim was to provide a detailed atlas of the wavelengths of spectral lines given in solar light.⁴ In his experiments he used two ready-made gratings, numbered as (I) and (II), the former with 4501 lines, the latter with 2701 lines engraved on a length, declared by the producer of the gratings as being equal to 9 Paris lines. The basic equation used by Ångström for the computation of the wavelength λ of the particular lines was $\lambda = e \cdot \sin\varphi$ for the spectrum of the first order, for its N -th order holds $\lambda = N \cdot e \cdot \sin\varphi$, where e stands for the width of the grating space and φ for the angle of diffraction of the light waves, the latter being measured by a theodolite. In addition, he modified these equations in order to take into account the variation of temperature and orientation of the gratings, and then computed the data for the spectra – in Fraunhofer notation – C, F of hydrogen. The resulting values were as follows (Ångström 1868, 18):

Line	1. spectrum in 10^{-7} mm	Number of observations	2. spectrum in 10^{-7} mm	Number of observations	3. spectrum in 10^{-7} mm	Number of observations
C	6562.27	2	6561.97	3	6562.16	1
F	4860.67	2	4860.62	2	4860.80	3

Table 2: Data for the spectral lines C and F of hydrogen
computed in Ångström (1868)

Based on all the data for grating (I), as well as all the data for grating (II), he then computed the *weighted arithmetic mean* values of the wavelengths of the spectral lines of solar light. For the four lines of hydrogen C, F, near G, and h – in the more modern notation H_α , H_β , H_γ , and H_δ to

⁴ For a detailed analysis of his experiments and the conceptual basis of his computations see Landauer (1898, 6-33) and Baly (1912, 12-29).

which we will hold from now on – he obtained the following mean values (Ångström 1868, 31 – 32):

Line	Wavelength in 10^{-7} mm
H_{α}	6562.10
H_{β}	4860.74
H_{γ}	4340.10
H_{δ}	4101.20

Table 3: Average wavelengths of the spectral lines of hydrogen computed in Ångström (1868)

These values were used by J. J. Balmer in his 1884 formulation of the first truly unifying formula for the wavelengths of the spectral lines of hydrogen.

From the *epistemological* point of view it holds that in his *thought* operations Ångström employed the statistical analysis based on the weighted statistical mean, which enabled him to pass in mind from what we label by the philosophical category *data* to what can be labeled by the philosophical category *phenomena*. Thus, *data* as a philosophical category stand in our approach for an *entity in mind* and simultaneously for *a certain level of attained knowledge about the world*. Both are information about occurrences given to us in our practical operations (manipulations) with things in the world; *occurrences in the world of things are thought in mind as data*.

In the case of the above given Table 2, the data for spectra stand for the meaning of a sentence like “The wavelength of the line C (H_{α}) in the 1st spectrum on grating (I) is $6562 \cdot 10^{-7}$ mm”. In our view, contrary to that of Bogen and Woodward, the meaning of sentences of this type does not refer to something observable in the world, or, stated in more general categories, to something accessible to human sensory perception. *Occurrences as states of affairs in the world in the range of 10^{-7} mm are not perceptible* and, of course, the *very meaning, in the sense of information, stated in a sentence like the one given above, is not perceptible either*. So, in our reconstruction here *data do not refer to perceptible state of affairs, and so neither do phenomena derived from them*. In fact, the meaning of a statement is never perceptible, since meanings as information can only be thought of because they are entities in the mind on which one can perform thought operations which yield other thought entities. In Ångström’s approach such operations were,

for example, the computations of the weighted arithmetic mean for the wavelength and his abstraction from the impact of changes of the atmospheric pressure.

3.2. From phenomena to laws of phenomena:

J. J. Balmer, J. R. Rydberg, and W. Ritz

The results given in Ångström (1868) were the point of departure for the development of spectral analysis leading to what W. Ritz labeled by the epistemological category *empirische Gesetze* (Ritz 1903, 266) and N. Bohr by the category-concept hybrid *empirical spectral laws* (Bohr 1922, 1).

J. J. Balmer was able to derive one formula, which he viewed as an “expression of a law” (Balmer 1885a, 552), and which unified the knowledge of the wavelengths of the four spectral lines H_α , H_β , H_γ , and H_δ .⁵ He drew on the wavelengths of the spectral lines of hydrogen as computed by Ångström: $\lambda_\alpha = 6562.10 \cdot 10^{-7}$ mm, $\lambda_\beta = 4860.74 \cdot 10^{-7}$ mm, $\lambda_\gamma = 4340.10 \cdot 10^{-7}$ mm, and $\lambda_\delta = 4101.20 \cdot 10^{-7}$ mm. He started by unifying the above given four wavelengths by means of a basic wavelength $\lambda_0 = 3645.6 \cdot 10^{-7}$ mm, so

that they are given as multiples of the following coefficients $\frac{9}{5}$, $\frac{4}{3}$, $\frac{25}{21}$, and $\frac{9}{8}$, respectively. By multiplying the second and third coefficients by $\frac{4}{4}$ he

obtained four coefficients which can be written as $\frac{9}{9-4}$, $\frac{16}{16-4}$, $\frac{25}{25-4}$,

and $\frac{36}{36-4}$, and thus unified into the formula $\frac{m^2}{m^2-2^2}$, where m stands

for the positive whole numbers 3, 4, 5, and 6. The four wavelengths λ_α , λ_β , λ_γ , and λ_δ can be then unified, by means of a basic wavelength $\lambda_0 = 3645.6 \cdot 10^{-7}$ mm, into the formula

$$(1) \quad \lambda = \lambda_0 \frac{m^2}{m^2 - 2^2} \quad (m = 3, 4, 5, 6)$$

Formula (1) can be viewed as embedded into a phenomenal spectral law, so that holds:

⁵ For an analysis of this derivation see Banet (1966) and (1970), respectively; the latter article draws on Balmer's manuscripts.

$$(2) \quad \forall x [H(x) \ \& \ I(x) \rightarrow \lambda(x) = \lambda_0 \left(\frac{m^2}{m^2 - 2^2} \right)]$$

Here “ $H(x)$ ” stands “ x is hydrogen,” “ $I(x)$ ” for “ x is incandesced,” and “ $\lambda(x)$ ” for “the wavelength of radiation emitted by x .” This law has the following two features. It can be used already for the purpose of explanation of the already known four different wavelengths of the spectral lines of hydrogen. It thus fulfills the requirement of the so-called *functional interdependence* imposed on scientific explanation by J. Woodward:

The law occurring in the explanans of a scientific explanation of some explanandum E must be stated in terms of variables or parameters variations in the value of which will permit the derivation of other explananda which are appropriately different from E . (Woodward 1979, 46)

We can provide here the first epistemological delineation of the category *scientific law of phenomena: it unifies in mind a number of different, already known phenomena pertaining to entities of the same kind and provides a unified account of these phenomena.*

The importance of such a phenomenal law is that it stood in Balmer’s reasoning not only for a unification of antecedently known phenomena, but was also used by him for the derivation of phenomena whose referents were not known to him when he stated formula (1). In fact, on the basis of formula (1) and the hypothesized value $m = 7$, he predicted the existence of another spectral line of hydrogen, with the predicted wavelength $\lambda = 3969.65 \cdot 10^{-7}$ mm. Only later did he learn that such a spectral line had already been discovered together with other spectral lines symbolized as H_ϵ , H_ζ , H_η , H_θ , and H_i . Thus, the scope of applicability of formula (1) was extended to the values 7 through 11 of the parameter m . In addition, in the second communication in Balmer (1885b), he again broadened, based on the additional knowledge he obtained about spectral lines, the scope of application of the formula (1), so that the values of the parameter m ranged now for the positive whole numbers 3 through 16.

In the ensuing development of spectral analysis, phenomenal spectral laws were generalized in order to cover also the spectral lines of substances other than hydrogen. As we shall see, the works of J. R. Rydberg and W. Ritz appear as crucial to this investigation.

Rydberg in (1890) employed the term “wave number,” abbreviated as n in the sense of “*the number of wavelengths in 1 centimeter*” (1890a, 13), that

is, $n = 10^8 \cdot \lambda^{-1}$, where λ is expressed in units used by Ångström. He also drew on the already acquired knowledge that the spectra of particular substances are ordered in series. Based on this he found that the wave numbers assigned to the spectral lines of different substances can be described by the following unified formula (1890, 40):

$$(3) \quad n = n_0 - \frac{N_0}{(m + \mu)^2}$$

Here “ m ” stands for a positive whole number labeled “number of the term;” N_0 is a universal constant, while constants n_0 and μ stand for specific constants characterizing particular spectral series of a substance, and where n_0 is the limit which n approaches when $m = \infty$. By transforming Blamer’s formula (2) so that it now holds for wave numbers ($\lambda = \frac{10^8}{n}$ and $\lambda_0 = \frac{10^8}{n_0}$)

he obtained the formula $n = n_0 \cdot \frac{m^2 - 2^2}{m^2}$ and thus $n = n_0 - \frac{4n_0}{m^2}$, thus $n = 4n_0 \left(\frac{1}{2^2} - \frac{1}{m^2} \right)$. By comparing the former formula with formula (3) he

found out that for its universal constant holds $N_0 = 4n_0$ and so as $\lambda_0 = \frac{10^8}{n_0}$,

he computed in (1890) that $N_0 = 109721.60$.

Finally, in early 20th century, W. Ritz formulated “a new law of the series spectra” (Ritz 1908, 521), making it possible to combine the already known formulas of spectral series of a substance, so that one could obtain new formulas of other series of this substance that can be used for the prediction of the existence of these other series. This law he stated in an auto-seminar paper added to his article (1908) as follows:

While the hitherto known laws of the series spectra connect with one another the lines of *one* series ... it is shown here, that there exists a simple relation also between different series of one element: by means of additive or subtractive combinations, either of the series formula or constants given in them, new formulas are being formed, which allow to compute completely the new lines discovered in recent years from those known earlier. (Ritz 1911, 162)

So, for example, as shown above, for the lines H_α , H_β , and H_γ , Balmer's formula (1) in Rydberg's reformulation yields:

$$n_\alpha = N_0 \left(\frac{1}{2^2} - \frac{1}{3^2} \right), n_\beta = N_0 \left(\frac{1}{2^2} - \frac{1}{4^2} \right), n_\gamma = N_0 \left(\frac{1}{2^2} - \frac{1}{5^2} \right)$$

By applying Ritz's combination principle, one obtains from these three formulas the following two additional formulas:

$$n_\beta - n_\alpha = N_0 \left(\frac{1}{3^2} - \frac{1}{4^2} \right), n_\gamma - n_\alpha = N_0 \left(\frac{1}{3^2} - \frac{1}{5^2} \right)$$

Thus, one is able to predict the existence of a spectral line for which should hold the formula

$$n = N_0 \left(\frac{1}{3^2} - \frac{1}{m^2} \right) \quad (m = 4, 5)$$

This prediction was confirmed by Paschen (1908), who – based on a series of experiments – proved the existence of a spectral series corresponding to the last formula.

Based on the last formula and Balmer's formula (1), Ritz stated the universal formula:

$$(4) \quad n = N_0 \left(\frac{1}{n^2} - \frac{1}{m^2} \right)$$

The phenomenal spectral law (2) then turns into the phenomenal spectral law

$$(5) \quad \forall x [H(x) \ \& \ I(x) \rightarrow n(x) = N_0 \left(\frac{1}{n^2} - \frac{1}{m^2} \right)]$$

Here " $n(x)$ " stands for "the wave number of the spectral line emitted by x ."

From the *epistemological* point of view it holds that Ritz's law (5) can be used not only (as can like Balmer's formula (1)) to explain and predict particular wavelengths, but also to explain the laws for Balmer's and Paschen's series. We can thus provide a reinterpretation of the category *law of phenomena* given above. *A scientific law of phenomena unifies in mind a number of different phenomena and laws of phenomena which pertain to entities of the same*

kind and provides a unified account of them. One can give also a generalization of the above quoted requirement of *functional interdependence* imposed by Woodward on scientific explanations:

The law occurring in the explanans of a scientific explanation of some explanandum phenomenon E and explanandum law of phenomena L must be stated in terms of variables or parameters variations in the value of which will permit the derivation of other explananda phenomena and explananda laws of phenomena which are appropriately different from E and L .

This generalization is based on Woodward's view that "the scientific explanations typically have as their explananda generalisations rather than singular sentences ... the scientific explanation of particular facts is an activity which is derivative or parasitic on explanation of generalisations" (Woodward 1979, 63).

4. Bohr's mechanism of the hydrogen atom

In Part I of his article (1913), Bohr sets as his aim to deal with "the mechanism of the binding of electrons by a positive nucleus" (Bohr 1913, 2). He conceptually approaches this mechanism by means of Planck's view of radiation which he employs drawing on the "general acknowledgment of the inadequacy of classical electrodynamics in describing the behavior of system of atomic size" (Bohr 1913, 2). At the same time, Bohr clearly differentiates in his approach two distinct but still interrelated phases, *one*, in which he will provide "a basis for a theory of the constitution of atoms", and another one, based on the first one, in which he will give a "an account ... for the law of the line spectra of hydrogen" (cf. Bohr 1913, 2).

As the object of his thought operations Bohr chooses "a simple system consisting of a positively charged nucleus of very small dimensions and an electron describing closed orbits around it" (Bohr 1913, 3), and where this object is subjected to two additional suppositions: the mass of the electron is negligible in size as compared to that of the nucleus, and the velocity v of the electron is much smaller than the velocity of light c . That choice enables him to bypass the issue of mechanical instability, and, when unified with the first supposition, provides the conceptual framework for the treatment of the hydrogen atom. The latter supposition, characterized by

means of the epistemological category *idealization*, states that $\frac{v}{c} = 0$ holds.

By the introduction of this idealization Bohr can abstract in mind from the impact of relativistic effects.

Bohr initially presupposes that no energy is radiated; thus he can apply classical mechanics to the case of the movement of the electron, on an elliptical orbit, around the nucleus located in the focus of this orbit. If e and m stand for the charge and mass of the orbiting electron, E for the charge of the nucleus, and $2a$ for the major axis of the orbit, then it holds (Bohr 1913, 3):

$$(6) \quad \omega = \frac{\sqrt{2}}{\pi} \cdot \frac{W^{\frac{3}{2}}}{eE\sqrt{m}}, \quad 2a = \frac{eE}{W}$$

Here ω stands for the orbital frequency of the electron, while W stands for the energy to remove it from the vicinity of the nucleus to infinity.

Then, Bohr gives up the supposition that energy is not radiated; this requires stabilizing the atom radioactively. To achieve this stabilization, he brings in Planck's idea of a discrete energy being radiated by a vibrator with the frequency ν , and where the energy radiated in one emission is $\tau h\nu$, where τ is positive whole number and h Planck's universal constant. He applies this idea of Planck in such a way that he now supposes the electron, being initially at a great distance from the nucleus, as bound by the nucleus and settled in an orbit around it. To this binding Bohr applies Planck's idea in such a way that the energy being emitted in this process is given by the relation $E = \tau h \frac{\omega}{2}$, where ω stands for the orbital frequency of the electron on its final orbit in which it settles.⁶ This relation, when combined with relations given in (6), yields

$$(7) \quad W = \frac{2\pi^2 m e^2 E^2}{\tau^2 h^2}, \quad \omega = \frac{4\pi^2 m e^2 E^2}{\tau^3 h^3}, \quad 2a = \frac{\tau^2 h^2}{2\pi^2 m e E}$$

Bohr interprets these relations in such a way that for the respective values of τ one obtains the values for W , ω , and $2a$ which characterize the

⁶ Bohr uses in Part I two additional applications of Planck's view on quanta of energy; we do not deal with them in this paper. On this see Heilbron – Kuhn (1969).

configurations of the atom when no energy is radiated by the atom. These configurations he views as characterizing the states of the atom; so he assigns to them the term “stationary states.” By bringing in the values for e , m , and h , and by supposing that $e = E$, he computes for $\tau = 1$ the linear dimension $2a$ of the atom of hydrogen, its optical frequency ω , and its ionization potential $\frac{W}{e}$; all these computed values correspond to those which are computed on the basis of experiments. At the end of his treatment of the atom itself, Bohr gives the following brief summary of his main assumptions:

- (1) The dynamical equilibrium of the systems in the stationary states can be discussed by help of the ordinary mechanics, while the passing of the systems between different stationary states cannot be treated on that basis.
- (2) That the latter is followed by the emission of a homogeneous radiation, for which the relation between the frequency and the amount of energy emitted is the one given by Planck’s theory. (Bohr 1913, 7)

In our view to Bohr’s treatment of the constitution of the hydrogen atom one can assign the category *ground* understood as *working of the ground’s mechanism*. This category involves as subordinated those categories which are given in Machamer – Darden – Craver (2000) and Darden (2008) and which were stated already above.

5. The spectra explained and predicted

After Bohr dealt with the working of the mechanism of the hydrogen atom, he moved on to the derivation of the formula for the spectral lines of hydrogen. As shown above, Rydberg’s formula (5) holds for the spectral lines of hydrogen. It can be transformed as follows. The relation $\lambda = c \cdot T$, where c is the speed of light and T the period of its movement, is transformed into $n = \frac{1}{c \cdot T}$, and by introducing the magnitude frequency ν as $\nu = \frac{1}{T}$ one obtains $\nu = c \cdot n$. So, by multiplying the formula (5) by the magnitude c , one obtains

$$(8) \quad v = c \cdot N_0 \left(\frac{1}{n^2} - \frac{1}{m^2} \right)$$

For $c = 3 \cdot 10^{10} \text{ cms}^{-1}$ and $N_0 = 109675.0 \text{ cm}^{-1}$ one obtains $c \cdot N_0 = 3.29 \cdot 10^{15} \text{ s}^{-1}$. Bohr compared this value, obtained by computation from the historically prior form of the formula for hydrogen, with the value he computed on the basis of the formula he derived from his theory about the hydrogen atom. He proceeded in the following way. He interprets the experimental process of discharge in vacuum tube when the spectrum is experimentally produced as a process of moving the orbital electron to a great distance from the nucleus. This latter process he views as corresponding to the creation of a stationary state when the electron is bound by the nucleus. For the energy W radiated in this binding holds the relation (7), and so as the object of his thought derivations is still hydrogen, he puts the absolute values of the charge of the orbiting electron and that of the nucleus as identical. He thus obtains from (7) the relation

$$(9) \quad W_{\tau} = \frac{2\pi^2 m e^4}{\tau^2 h^2}$$

Then, the energy produced when the atom passes from the state characterized by τ_1 to that characterized by τ_2 is as follows:

$$(10) \quad W_{\tau_2} - W_{\tau_1} = \frac{2\pi^2 m e^4}{h^2} \left(\frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right)$$

This energy he puts equal to $h\nu$, where ν is the frequency of radiation, so he obtains

$$(11) \quad \nu = \frac{2\pi^2 m e^4}{h^3} \left(\frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right)$$

From here Bohr's thought movement goes into two directions. One deals with the expression $\frac{2\pi^2 m e^4}{h^3}$ in (11) and the second with the expression in the brackets of (11). The former expression is used by Bohr for the recomputation of the constant given above as $c \cdot N_0$; for the values of e , m , and h it yields the value $3.10 \cdot 10^{15} \text{ s}^{-1}$. The expression in the brackets is used by

Bohr in the *explanation* of already known spectral series and in the *prediction* of the existence of spectral lines as yet undetected. By putting $\tau_2 = 2$ and for a varying τ_1 he explains Balmer's series and for $\tau_2 = 3$ he *explains* Paschen's series. He *predicts*, by putting $\tau_2 = 1$, or 4, or 5, the existence of spectral series both in the extreme ultra-violet and the extreme ultra-red, which "are not observed, but the existence of which may be expected" (Bohr 1913, 8). Already in 1914 the series corresponding to $\tau_2 = 1$ was detected; few years later the series for τ_2 equal to 4 and 5 were also experimentally detected.⁷

Finally, still in Part I, Bohr leaves the framework of reflections on hydrogen atom and its spectra, and gives an interpretation of Ritz combination principle as holding for all substances producing line spectra:

The circumstance that the frequency can be written as a difference between two functions of entire numbers suggests an origin of the lines in the spectra in question similar to the one we have suggested for hydrogen, i.e. that the lines correspond to a radiation emitted during the passing of the system between two stationary states. This may account for the different sets of series in the line spectra emitted from substances in question. (Bohr 1913, 11)

What *epistemological lessons* can be drawn here? Once Bohr succeeded in the above reconstructed two directions of thought movement, his *knowledge has grown*. What we mean by this can be understood when we put side by side the initial formulation (8) of the spectral formula for hydrogen and the spectral formula (11) derived by Bohr:

$$(12) \quad \nu = c \cdot N_0 \left(\frac{1}{n^2} - \frac{1}{m^2} \right), \nu = \frac{2\pi^2 m e^4}{h^3} \left(\frac{1}{\tau_2^2} - \frac{1}{\tau_1^2} \right)$$

By their comparison it becomes readily seen that the expression " $c \cdot N_0$ " is restated as $\frac{2\pi^2 m e^4}{h^3}$, and so the formula for the phenomenon of spectral emission is completely tied to its explanans, the latter being Bohr's unification of Rutherford's atom with Planck's views on the quanta of radiation. The symbols " e ", " h " and " m " together with their respective meanings ap-

⁷ On this see Lyman (1914), Brackett (1922) and Pfund (1924).

pear in the explananda and predicanda because there are shifted here in the process of explanation and prediction from Bohr's explanans – the understanding of the *ground's* mechanism producing the respective phenomena.⁸ By this comparison one finds out also that the interpretation of the symbols “*n*” and “*m*” is different from that of “ τ_2 ” and “ τ_1 .” While the former pair stands just for an ordered pair of numbers, the latter stands for the sequence of orbits of the electron.

The growth of knowledge becomes readily seen also when one reconstructs the respective scientific laws into which the formulas (12) are embedded. We reconstruct them as follows:

$$\forall x [H(x) \ \& \ I(x) \rightarrow v(x) = c \cdot N_0 \left(\frac{1}{n^2} - \frac{1}{m^2} \right)],$$

$$\forall x \forall y [El(x) \ \& \ N(y,x) \ \& \ H^*(z,x,y) \ \& \ D(z) \ \& \ \frac{v(y)}{c} = 0 \rightarrow v(x) =$$

$$\frac{2\pi^2 \ m(x)^{(1)} \ E(y)^2 \ e(x)^2}{h^3} \left(\frac{1}{\tau_2(x)^2} - \frac{1}{\tau_1(x)^2} \right)]$$

In the antecedent of the second law “*El(x)*” stands for “*x* is an electron,” “*N(y,x)*” for “*y* is the nucleus orbited by *x*,” “*H(z,x,y)*” for “*z* is hydrogen composed of *x* and *y*,” “*D(z)*” for “*z* is disturbed,” and “ $\frac{v(y)}{c}$ ” for “the ratio of the speed of *y* and speed of light.” In the consequent “*m*⁽¹⁾(*x*)” stands for “mass of *x* subjected to one idealization”.

By comparing these two laws, one finds out that what has changed is the very meaning of the term “hydrogen,” that is, the universe of discourse to which the law can be applied at all. While in the former law, symbolized as “*H*,” it is understood as, say, a substance with a certain atomic weight, reacting in certain proportion with other substances and with a certain ionization potential, in the latter, symbolized as “*H**,” it is understood as being set up by an electron orbiting the nucleus. What has changed also is the understanding when the hydrogen is at all radiating: in the former law it has to be incandesced, in the latter law it has to be disturbed from the outside by an input of energy, so that the electron is shifted from one orbit

⁸ In order not to overburden the text we use here only the category *explanans* as the basis for the thought derivation of both the explananda and predicanda; thus refraining from the use of the category *predicans*.

to another. Finally, Bohr's equation given in the law holds only when the idealization holds that the orbiting speed of the electron inside the hydrogen must be much smaller than that of light.⁹

Based on these analyses, we come to the conclusion that while it is possible to assign both to the knowledge which is the point of departure of the thought movement *to* the ground's mechanism as well as to the knowledge which is obtained by explanation and prediction *from* the knowledge about the ground's mechanism the same category pair *phenomena* and *law of phenomena*, this pair of philosophical categories does not capture adequately the above reconstructed cases of growth of knowledge. The phenomena and laws of phenomena as the epistemic points of departure for the movement to ground's mechanism and as the latter's epistemic consequences differ mutually in some fundamental aspects. This difference, as can be seen from Diagram 1, was not taken into account in the article Bogen – Woodward (1988). As a consequence of the absence of this account, the whole cycle *data* → *phenomena* → ... → *phenomena*, reconstructed for the first time in Bogen – Woodward (1988) fails to express the extension of knowledge. This failure could be turned into an argument arguing in favor of a modified Hempelian view from the fifties, namely, that *even if the phenomena are not observable, still it does not make any sense to make a detour through theory and the knowledge of mechanism given in it*; one starts and ends with the same knowledge expressed by the category phenomena. A remedy to this deficit will be given now.

6. The epistemological lessons

Based on the above given reconstructions of the growth of knowledge, we propose the following, more subtle differentiation between philosophical categories. Knowledge which was till now labeled by the category *phenomena* is now labeled either by the category *appearance* or by the category *manifestation*, and knowledge which was till now labeled by the category *law of phenomena* is now labeled either by the category *law of appearance* or by the category *law of manifestation*. The philosophical category *appearance* is assigned to that level of knowledge and thinking where

⁹ On the reconstruction of the structure of scientific laws containing idealizations see Nowak (1972).

the phenomena are thought as yet independently and prior to their unification in scientific laws, where the latter have the status of *laws of appearance*; it holds for them the characterizations of the category *law of phenomena* given above. The category *law of appearance* characterizes knowledge and thinking before the knowledge about the ground's mechanism is derived from them. Once the very mechanism of the ground is understood, the laws of phenomena and the phenomena derived (explained and/or predicted) have already the status of *laws of manifestation* and of *manifestations*. The category *law of manifestation* can be characterized as follow: *it unifies in mind a number of different phenomena pertaining to entities of the same kind, and where this unification stands for the derivation based on the understanding how the phenomena are produced by the working of the mechanism of the ground of these phenomena.*

How can the difference between categories *appearance* and *law of appearance*, on the one hand, and *manifestation* and *law of manifestation*, on the other hand, be epistemologically explicated? As shown above, the laws of spectra and thus also the frequencies computed on their basis, once derived from Bohr's understanding of the working of the mechanism of the atom, contain already the symbols e , m , h , τ_1 , and τ_2 . These symbols, as well as their respective meanings, are part of the explananda/predicanda and have their origin in their common explanans – Bohr's understanding of the working of the mechanism, which is based on the meaning of such terms as “electron moving on in orbit above the nucleus,” “quantum of radiation emitted,” “stationary orbits,” etc. This can be stated in more general, epistemological terms as follows. Concepts introduced at the level of knowledge and thinking characterized by the category *working of the mechanism of the ground* – as the explanans – are shifted in the process of explanation/prediction to the explananda/predicanda. So, while at the level of knowledge and thinking characterized by the categories *appearances* and *laws of appearances*, concepts standing for the knowledge about the working of the mechanism are as yet not given, they are already given at the level knowledge characterized by the categories *laws of manifestations* and *manifestations*. This is the difference between knowledge characterized by the categories *appearance and laws of appearance* and knowledge characterized by the categories *law of manifestation and manifestation*.

Based on this differentiation of categories, we propose – as an alternative to S. Schindler's Diagram 1 – the following diagrammatical representation of the sequence of categories reconstructed till now.

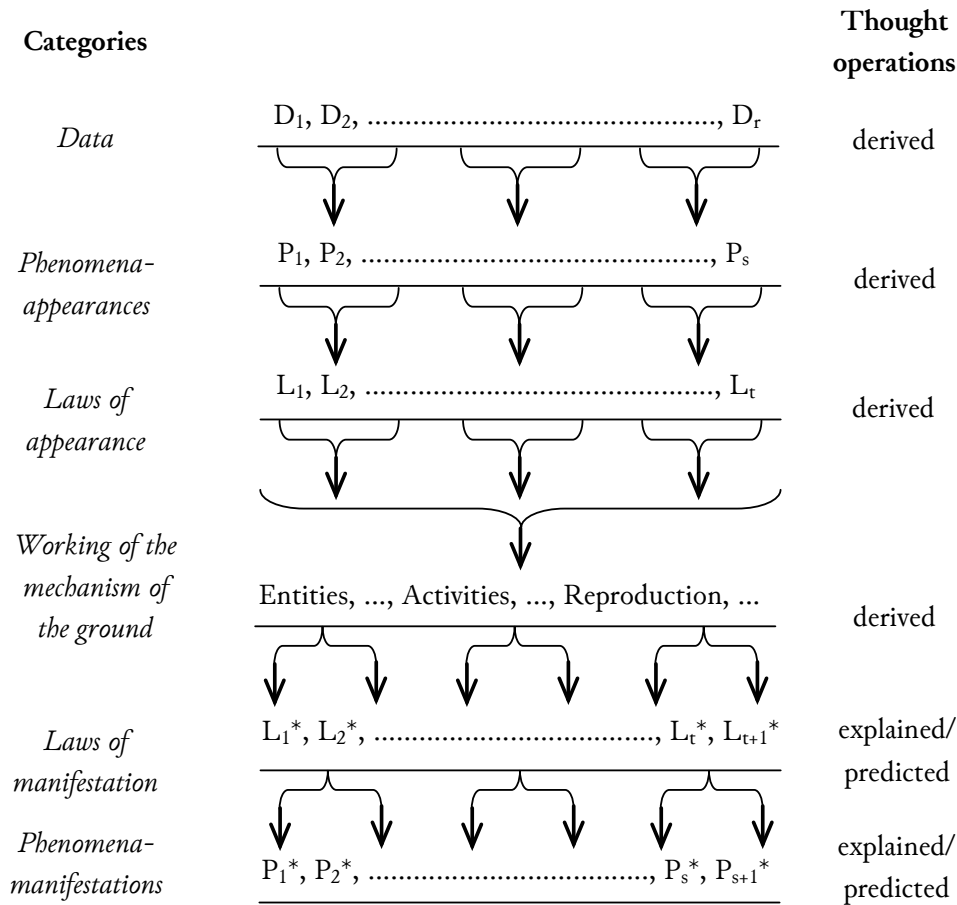


Diagram 2: Sequence of categories reconstructing the growth of scientific knowledge

Here the scientific laws $L_1^*, L_2^*, \dots, L_t^*$ stand for a reinterpretation by means of explanation of the scientific laws L_1, L_2, \dots, L_t , while the laws L_{t+1}^*, \dots , stand for laws of manifestation which do not have their counterpart in the laws of appearance; they are *predicanda* laws and not explananda laws. The same holds for manifestations expressed as $P_1^*, P_2^*, \dots, P_s^*$; they are the reinterpretations, based on explanation from the laws of manifestation, of the appearances P_1, P_2, \dots, P_s , while the manifestations P_{s+1}^*, \dots , are *predicanda phenomena* derived from the laws of manifestations.

References

ÅNGSTRÖM, A.-J. (1868): *Recherches sur le spectre solaire*. Upsal: W. Schultz.
 BALMER, J. J. (1885a): Notiz über die Spectrallinien des Wasserstoffs. *Verhandlungen der Naturforschenden Gemeinschaft in Basel* 7, 548-560.

- BALMER, J. J. (1885b): Zweite Notiz über die Spectrallinien des Wasserstoffs. *Verhandlungen der Naturforschenden Gemeinschaft in Basel* 7, 750-752.
- BANET, L. (1966): Evolution of Balmer Series. *American Journal of Physics* 34, 496-503.
- BANET, L. (1970): Balmer's Manuscripts and the Construction of His Series. *American Journal of Physics* 38, 821-828.
- BOGEN, J. (2011): 'Saving the Phenomena' and Saving the Phenomena. *Synthese* 182, 7-22.
- BOGEN, J. – WOODWARD, J. (1988): Saving the Phenomena. *Philosophical Review* 97, 303-352.
- BOGEN, J. – WOODWARD, J. (1992): Observations, Theories, and the Evolution of Human Spirit. *Philosophy of Science* 59, 590-611.
- BOHR, N. (1913): On the Constitution of Atoms and Molecules. *Philosophical Magazine* 26, 1-25, 476-502, 857-875.
- BOHR, N. (1922): *The Theory of Spectra and Atomic Constitution*. Cambridge: Cambridge University Press.
- BRACKETT, F. S. (1922): A New Series of Spectral Lines. *Nature* 109, 206.
- DARDEN, L. (2008): Thinking Again about Biological Mechanisms. *Philosophy of Science* 75, 958-969.
- HEILBRON, J. L. – KUHN, T. D. (1969): The Genesis of the Bohr Atom. *Historical Studies in the Physical Sciences* 1, 211-290.
- LYMAN, T. (1914): An Extension of the Spectrum in the Extreme-Violet. *Physical Review* 3, 504-505.
- MACHAMER, P. – DARDEN, L. – CRAVER, C. F. (2000): Thinking about Mechanism. *Philosophy of Science* 67, 1-25.
- NOWAK L. (1972): Laws of Science, Theories, Measurement. *Philosophy of Science* 39, 533-548.
- PASCHEN, F. (1908): Zur Kenntnis ultraroter Linienspektren. I. (Normalwellenlängen bis 27 000 Å.-E.). *Annalen der Physik* 27, 537-570.
- PFUND, A. H. (1924): The Emission of Nitrogen and Hydrogen in the Infrared. *Journal of the Optical Society of America* 9, 193-196.
- RITZ, W. (1903): Zur Theorie der Serienspektren. *Annalen der Physik* 12, 264-310.
- RITZ, W. (1908): Über ein neues Gesetz der Serienspektren. *Physikalische Zeitschrift* 9, 521-529. Reprinted in Ritz (1911, 141-162).
- RITZ, W. (1911): *Gesammelte Werke*. Paris: Gauthier-Villars.
- RYDBERG, J. R. (1890): Recherches sur la constitution des spectres d'émission des éléments chimiques. *Kungliga Svenska Vetenskaps-Akademiens Handlingar* 23, 1-151.
- SCHINDLER, S. (2007): Rehabilitating Theory. *Studies in the History and Philosophy of Science* 38, 160-184.
- SOMMERFELD, A. (1921): *Atombau und Spektrallinien*. Braunschweig: Vieweg und Sohn.
- SOMMERFELD, A. (1923): *Atomic Structure and Spectral Lines*. London: Methuen.
- WOODWARD, J. (1989): Data and Phenomena. *Synthese* 79, 393-472.
- WOODWARD, J. (2000): Data, Phenomena, and Reliability. *Philosophy of Science* 67, 163-179.
- WOODWARD, J. (2011): Data and Phenomena. *Synthese* 182, 165-179.

Consciousness and Peripheral Self-Awareness

SİNEM ELKATİP HATİPOĞLU

Department of Philosophy. College of Humanities and Social Sciences. İstanbul Şehir University
İstanbul 34660. Turkey
sinemhatipoglu@sehir.edu.tr

RECEIVED: 10-12-2012 • ACCEPTED: 11-04-2013

ABSTRACT: It has been argued that consciousness is not possible without peripheral self-consciousness; i.e., without an implicit awareness of oneself as the subject of one's mental state. My purpose is to undermine this view. I contend that a subject's first-personal access to her conscious mental states and an awareness of them as *hers*, along with a particular approach to consciousness according to which a subject cannot be unconsciously conscious of things motivate the view that consciousness is not possible without self-consciousness. In order to undermine this view I argue that not all conscious states are accompanied by a sense of *mineness*. I also reject the reasons for endorsing an approach to consciousness according to which a subject cannot be unconsciously conscious of things. Then I critically examine Kriegel's arguments for the dependence of consciousness on self-consciousness based on the first-personal access a subject has to her conscious mental states and discuss the difficulties involved.

KEYWORDS: Consciousness – higher-order theories of consciousness – peripheral self-awareness – self-consciousness.

1. Introduction

Is self-consciousness required for consciousness? It seems that the answer is no. We do not think that for a person to have a conscious perception of a tree, she needs to be aware of herself perceiving the tree. In fact, when one is intensely engaged in some mental activity, we often say that he's lost himself in it suggesting that his center of attention is not on him-

self. However, Kriegel (2004) has argued that a failure to take note of the distinction between two forms of self-consciousness is what gives way to the idea that consciousness is possible without self-consciousness. He further contends that all forms of consciousness depend on one particular form of self-consciousness, viz. intransitive state self-consciousness which consists in the peripheral awareness of oneself as the subject of one's mental state.¹

Henceforth I refer to the view that self-consciousness is required for consciousness, in the sense that the latter is not possible in the absence of the former, as the *requirement thesis* and my purpose is to undermine it.² I contend that the following considerations motivate the requirement thesis. The first is the special first-personal access one has to one's inner life. After Kriegel (2004, 197) I call this *the first-person knowable property* of conscious mental states. The second consists in the mistaken view that self-consciousness is ubiquitous. The third consists in the endorsement of a particular approach to consciousness according to which a subject cannot be unconsciously conscious of things. In what follows, in order to undermine the requirement thesis, I argue that self-consciousness is not as ubiquitous as is assumed; in other words, it does not have to accompany all conscious mental states. I also reject the reasons for endorsing an approach to consciousness according to which a subject cannot be unconsciously conscious of things. Then I critically examine Kriegel's (2004) arguments for the requirement thesis based on the first-person knowable property of mental states and discuss the difficulties involved.

¹ Henceforth by self-consciousness I mean the peripheral awareness of oneself and not the reflective, focal, introspective self-consciousness unless noted otherwise. Also I'm using 'awareness' and 'consciousness' interchangeably with one stylistic difference. When I talk about subjects being conscious of things, whether they are mental states, objects or oneself, I use 'conscious' or 'aware', but when I talk about mental states being conscious, I only use 'conscious'. This reflects the distinction Rosenthal (2005, 4) draws between transitive and state consciousness discussed later. For the moment, it is sufficient to say that the former represents subjects being conscious *of* things and the latter represents mental states being conscious where it is understood that mental states are not conscious *of* things.

² There have been stronger attacks on the requirement thesis, denying even the possibility that outer focal consciousness is *sometimes* accompanied by such peripheral self-awareness. See Gennaro (2008) for a discussion of this and Kriegel (2009) for replies. Also see Schear's (2009) critique of Zahavi's (2006) arguments in favor of the requirement thesis.

2. Self-consciousness and the requirement thesis

It is possible to trace the idea of no consciousness without self-consciousness – although not in so many words – as far back as to Aristotle (*On the Soul* 3.2) and his view that it is impossible to perceive something and not be aware that one is perceiving it, to which Sorabji refers as Aristotle's 'most Cartesian remark'.³ The idea can also be found in Locke (1694/1975, 115) when he says 'thinking consists in being conscious that one thinks'. Gallagher – Zahavi's (2008) survey of the phenomenological literature⁴ reveal considerable consensus that consciousness calls for self-consciousness and this self-consciousness is described as 'an intrinsic feature of primary experience ... it is not thematic or attentive or voluntarily brought about; rather it is tacit, and very importantly, thoroughly non-observational (that is, it is not a kind of introspective observation of myself) and non-objectifying (that is, it does not turn my experience into a perceived or observed object)' (Gallagher – Zahavi 2008, 46). In addition, Zahavi (2006) himself argues that consciousness essentially involves self-consciousness.

There are similar remarks in the analytical tradition too. Goldman (1970, 96) says that the process of thinking about something carries with it a non-reflective self-awareness. Flanagan (1992, 194) speaks of a 'low-level self-consciousness involved in experiencing my experiences as *mine*'. And Kriegel (2004) argues that all forms of consciousness depend upon a peripheral awareness of oneself as the subject of one's mental state. So there has been some growing consensus with regards to the requirement thesis.

As can be seen from the claims above, the force of the idea that consciousness is not possible without self-consciousness seems to rely on the simple yet powerful contention that we are always aware of our conscious mental states and aware of them as *ours*. However, a subject's awareness of her conscious mental states as *hers* is not necessarily implied by the subject's mere awareness of those conscious mental states. There is no doubt that

³ See Caston (2002) for a discussion of this. Caston cites Sorabji's remark in (2002, 759).

⁴ Gallagher – Zahavi (2008, 46) cite Husserl who says that consciousness always involves a self-appearance, Heidegger who says every consciousness is also self-consciousness, and Henry who claims that experience is always self-manifesting. They also cite Sartre who says "This self-consciousness we ought to consider not as a new consciousness, but as the only mode of existence which is possible for a consciousness of something" (Gallagher – Zahavi 2008, 47).

through introspection, I'd consider the conscious states that I am in as mine, but that is through introspection and awareness of one's conscious mental states is not the same as introspection as will be discussed later.⁵ With regular conscious mental states that we do not introspect upon, self-consciousness does not come as naturally as it sounds. When I'm aware of a cup on my desk in front of me next to my laptop, I don't really think about the perception (of the cup) as mine. Awareness of myself as the owner of the perception seems to be lacking from the phenomenology of my experience, and that is usually the case with ordinary, non-reflective experiences.⁶

However the lack of self-consciousness in the phenomenology of one's experience may be due to the peripheral quality of self-consciousness. In other words one may argue that self-consciousness is ubiquitous but remains tacit and therefore is not noticed in ordinary, non-reflective experiences. Kriegel (2004) talks about the methodological difficulties of defending peripheral phenomena in general and using the example of peripheral vision, he says that such phenomena may be denied because 'whenever we want to direct our attention to the periphery of our visual field, the latter thereby becomes the focal center' (Kriegel 2004, 193). While this may be true of various peripheral phenomena, it must also be noted that Hume (1739-40/1978, 252) for instance found no self to be aware of even when he has painstakingly tried to direct his attention to himself. Hence the analogy Kriegel draws between peripheral awareness of oneself as the subject of one's mental states and other types of peripheral phenomena may be a little overreaching and the temptation to deny peripheral self-awareness may need to be taken more seriously than the temptation to deny other peripheral phenomena.

Nevertheless, self-consciousness is well and alive in other cases like when there is introspection. When I introspect on the perception I had of the cup, consciousness of the perception of the cup as mine in addition to the mere perception of the cup may simultaneously take part in my mental life. Hence one might argue that self-consciousness is present in all con-

⁵ For the moment I can say that introspection has to do with the awareness of the awareness of a mental state.

⁶ Such experiences would make up what Gennaro (2008) calls outer focal consciousness and he argues that such consciousness is never accompanied by inner peripheral self-awareness.

scious states but indeed peripheral and becomes focal through introspection. Later I suggest that by endorsing a higher-order theory of consciousness, it is possible to attribute self-consciousness to the presence of a higher-order mental state, e.g. the introspective state and not to the conscious lower-order mental state, e.g. the perception of the cup, in which case one does not need to assert that self-consciousness is ubiquitous.

One way of undermining the requirement thesis then is to suggest that the self-consciousness is a property of only certain kinds of mental states and not all conscious states. To that end, I now turn to a discussion of two different approaches to consciousness.

3. Two approaches to consciousness

First and foremost it is important to take note of the two different ways the term “conscious” can be used. One refers to the individual, a person or an organism, becoming conscious of something, and the other refers to a mental state’s being conscious. This distinction is noted and explained in detail by Rosenthal (2005, 4) and he calls the former transitive consciousness and the latter state consciousness. There is also creature consciousness that basically amounts to an organism’s being awake and mentally responsive (cf. Rosenthal 2005, 41). Rosenthal also talks about the transitivity principle according to which a mental state is conscious when its subject is first-personally aware of it, i.e., in a non-inferential manner (cf. Rosenthal 2005, 4). In accordance with the transitivity principle Rosenthal also says ‘A creature can be in mental states without being in conscious mental states, and can be conscious of external or bodily events without also being aware of its own mental states’ (Rosenthal 2005, 41). He explicitly states that transitive consciousness can occur without intransitive state consciousness (1997, 737).

According to Kriegel (2004, 184) however, ‘the concept of consciousness-of presupposes the concept of consciousness’ and ‘... a person exhibits transitive creature consciousness only if she has a mental state that exhibits transitive state consciousness’. In other words, an individual is conscious of something by virtue of being in a *conscious* mental state. So while both Kriegel and Rosenthal seem to agree on the fact that a person is conscious of something in virtue of being in a mental state, Kriegel requires that the relevant mental state is conscious, where Rosenthal does not. To emphasize the difference I shall henceforth refer to Rosenthal’s view as the *transi-*

tivity principle. I will refer to Kriegel's view, as the *state principle* since according to him no subject is conscious of anything unless the relevant mental *state* she is in is conscious.

If one endorses the transitivity principle, the question of state consciousness turns on a question of how an awake and a mentally responsive organism gets to be aware of something, or rather how it gets to be aware of its mental states to be precise and the answer is likely to call for a cognitive mechanism by which this process of awareness can be explained. In other words it would be possible to explain state consciousness in virtue of transitive consciousness. However, for the defender of the state principle, this explanation would be circular since the subject's awareness of things in general including mental states is not possible without an already existing state consciousness. Hence, given the allegedly ubiquitous nature of self-consciousness, the defender of the state principle uses self-consciousness to explain state consciousness. In other words, it seems that Kriegel's endorsement of the state principle gives way to his endorsement of the requirement thesis.

Another way of undermining the requirement thesis then is to undermine the state principle. But that is not enough since there is no immediate reason to think that an approach to consciousness that endorses the transitivity principle is not going to in turn depend upon self-consciousness. Hence my purpose is twofold; to argue that there are reasons to reject the state principle and to argue that transitivity principle can indeed accommodate an approach to consciousness that does not rely on self-consciousness. I start with the latter and in doing so I also show that self-consciousness does not have to be a property of all conscious states.

4. Transitivity principle

As mentioned before, according to the transitivity principle, it is in virtue of the creature's awareness of *its mental states* that a particular mental state in question is conscious. According to some theories of consciousness, a subject gets to be aware of its mental state x when there is another mental state y about it. Mental state y could be a thought or a perception⁷ and is usually described as a higher-order state because it is about another men-

⁷ See Rosenthal (2005) for the former and Lycan (1996) for the latter.

tal state, viz., the target state. Hence it is in virtue of there being a mental state about another mental state that the latter state, viz., the target state is conscious. The question then is whether this account of state consciousness involves self-consciousness.

According to one version of higher-order theories, viz. the higher-order thought (HOT) theory, the higher-order state is described as a thought to the effect that *I* am in this (target) state (Rosenthal 2005). Thus the content of the higher-order state involves self-consciousness. There are two remarks that can be made with regards to this.

Firstly, although the higher-order state involves self-consciousness, the higher-order state does not need to be conscious for the target state to be conscious.⁸ However this does not immediately solve the problem. In favor of the requirement thesis, one might argue that when the target state is conscious, self-consciousness is nevertheless present in the accompanying higher-order state since it is contained in the content of the higher-order state, waiting as it were, to be phenomenologically activated when the higher-order state becomes conscious. Hence one might argue that consciousness is indeed not possible without self-consciousness since each conscious state is accompanied by self-consciousness – albeit possibly just potential – contained in the higher-order state. But it should be noted that with regards to arguments in favor of the requirement thesis there is a lot of appeal to a sense of *mineness* with regard to experiences and such *potential* self-consciousness in the higher-order state would be phenomenologically no different from there being no self-consciousness. Besides, whatever implications the consciousness of the higher-order state might have, since the higher-order state is numerically distinct from the target state, self-

⁸ The suggestion that the higher-order state does not need to be conscious for the target state to be conscious comes from Rosenthal as a reply to an objection to his higher-order thought theory. The objection is that if the higher-order state in turn needs to be conscious so that the target state is conscious, that would mean there is yet another higher-order state, viz. a third-order state about the second-order state so that the second-order state is conscious. And this can go on infinitely. But if the higher-order state does not need to be conscious for the target state to be conscious, then the infinite regress stops. That the higher-order state is not necessarily conscious would also explain why self-awareness is usually missing from the phenomenology of ordinary, non-reflective experiences but present in experiences of a reflective and introspective kind since it is likely that in the latter case relevant higher-order states are also conscious.

consciousness would be separated from the target conscious state, as a result of which the allegedly ubiquitous nature of self-consciousness would be undermined. In other words, self-consciousness would not be taken to accompany all conscious states.

Still, it is worth considering if the content of the higher-order thought can be described in another way. For instance, there is no reason to think that the higher-order state could not be an indexical kind of thought with the content “that state” rather than “*I* am in that state.” As such it would be possible to re-describe the higher-order state in a more Humean or Parfit’s reductionist way. The purpose of the indexical type of higher-order state may be to assign the target state to a group of other psychologically continuous states – or if one favors bodily continuity theories of identity, to a group of bodily states that fall on the same spatio-temporal path – including the higher-order state itself. Hence, self-conscious friendly content of the higher-order state which is “I’m in so and so state” may be replaced by the following self-consciousness free content of the higher-order state: “this target state goes with this body and this series of psychophysical states.” And when the higher-order state is conscious, one may inferentially form the conception of oneself as the subject of a group of psychologically continuous states, but that is far from the kind of self-consciousness denoted by the requirement thesis.⁹

I now turn to Kriegel’s arguments for the state principle and dependence of consciousness on self-consciousness. My purpose is to show the various difficulties involved with his arguments including his reasons for the state principle in order to deny what might have otherwise been positive and independent reasons to think that consciousness depends on self-consciousness.

5. State principle

Kriegel (2004) talks about four kinds of consciousness viz. transitive creature consciousness, intransitive creature consciousness, transitive state consciousness and intransitive state consciousness. In relation to the differ-

⁹ This is also compatible with the fact that a conception of oneself comes much later in the development of mental lives, definitely after consciousness. See Damasio (1999) for a discussion of this.

ent forms of consciousness introduced earlier by Rosenthal viz., state and transitive consciousness, one can say that intransitive state consciousness is the same as state consciousness, which consists in the mental state's being conscious. Transitive creature consciousness is the same as transitive consciousness, which consists in the creature's being conscious *of* something. Intransitive creature consciousness is the same as creature consciousness that consists in the organism's being awake and mentally responsive. That leaves one with transitive state consciousness.

The difficulty with transitive state consciousness is that while it is relatively easy to grasp what it means for a creature, an individual, a person or an animal to be conscious *of* something it is not so when it comes to understanding what it means for a mental state to be conscious *of* something. Kriegel (2004) says that this is merely because of 'the way the phrase "conscious of" works that mental states cannot be conscious of anything' and he defines transitive state consciousness as the *property* a mental state *M* has when *M*'s subject is conscious of something in virtue of being in *M* (cf. Kriegel 2004, 185). He says:

the occurrence of transitive creature consciousness implies the occurrence of a mental state. When Smith is conscious of her new car, she is in a certain mental state, and it is in virtue of being in this mental state that she is conscious of the car (as opposed to being conscious of a table, or not being conscious of anything). We may label the property the mental state in question has transitive state consciousness. A mental state exhibits transitive state consciousness iff it is a mental state in virtue of which its subject exemplifies transitive creature consciousness. Thus, Smith's thought of her new car is transitively state-conscious, because in virtue of having it, Smith is conscious of her new car. (Kriegel 2004, 183-184)

Kriegel then argues for the dependence of transitive state consciousness on intransitive state consciousness. The relation between transitive and intransitive state consciousness is described by Kriegel as a question of whether a subject can be conscious of something in virtue of being in an unconscious mental state (cf. Kriegel 2004, 185). In other words, he drops the use of the phrase "transitive state consciousness" and refers to a property of the subject; hence it can be understood as the argument for the dependence of, in Rosenthal's terms, transitive consciousness on state con-

sciousness, in other words, the argument for the state principle. I will now consider whether these arguments are tenable.

First Kriegel (2004, 185) uses the example of a tacit belief that 17.3 is greater than 13.9 and says that we would not want to say that someone is conscious of the fact that 17.3 is greater than 13.9 just because he has a tacit belief to that effect. Hence the mental state involving this mathematical proposition should be conscious so that one can talk about the individual being aware of such a proposition. However, this example is a bit biased. It is much more likely that a person's mental states regarding *mathematical propositions* are conscious when he is doing calculations and is aware of how numbers relate to one another. But one cannot thereby make a generalization and assert that this is how it'll be for all mental states across the board. In other words one cannot infer that for everything that an individual is aware of, there will be an accompanying relevant conscious mental state. Consider Armstrong's (1968, 93) famous example of the long-distance truck driver who has absolutely no recollection of his driving whatsoever. As he is driving on the highway, not causing any accidents, he must be conscious of some things, but at the same time it is obvious that he is not in as many conscious mental states as his awareness of the road would call for if the state principle were true.

The second reason Kriegel in (2004, 185) gives for why transitive creature consciousness depends on intransitive state consciousness is that 'as far as the normal usage of the terms is concerned, it is all but incoherent to describe a person as unconsciously conscious of the fact that p '. In other words, one cannot be in a non-conscious mental state about x and at the same time be conscious of x . However, it is important to notice that the term "conscious" does not have the same referent here. In the supposedly incoherent description of a person being unconsciously conscious of something, the former qualifying term "unconsciously" refers to state consciousness and the latter "conscious" or rather "conscious of" refers to a person's being conscious of something, so transitive consciousness. Being unconsciously conscious of something amounts to a person's being aware of something but not being in the relevant conscious mental state and this is not immediately incoherent unless one endorses the state principle and thereby cannot be used to show the truth of the state principle.

One prime example of subjects being unconsciously conscious of things involves the case of blindsighted persons. People with blindsight have suffered damage to the visual center of the higher brain and when they are

presented with an object in their visual field that is related to the damaged region, they report that they don't see anything. But it's been shown that when they are asked to make guesses that involve the objects presented to them, their guesses are correct considerably above chance. This suggests that they perceive the object but is not aware of doing so. In other words, they are conscious of the object but the relevant mental state is not conscious, i.e., they are *unconsciously conscious* of it.¹⁰

The third reason Kriegel in (2004, 186) gives for the state principle is that persons are conscious of only a few things at a time and that if persons were conscious of things in virtue of being in non-conscious mental states too, they'd be conscious of innumerable many things. I contend that this argument reveals a misunderstanding of the transitivity principle. According to the transitivity principle, those mental states that an individual gets to be aware of are classified as conscious mental states. An individual's being conscious of things (other than mental states) does not imply the presence of conscious mental states. But it does not thereby claim that an individual gets to be aware of things in virtue of being in non-conscious mental states. It just endorses the possibility of transitive consciousness without state consciousness. That does not mean every non-conscious mental state one is in will make him conscious of something. Hence an overcrowded transitive consciousness is not an inevitable consequence of the transitivity principle.

After examining Kriegel's (2004) reasons for the state principle and discussing the difficulties involved, I'd like to turn now to his arguments for why consciousness depends on self-consciousness, or rather why intransitive state consciousness in turn depends on intransitive state self-consciousness, rendering all forms of consciousness to be dependent on intransitive state self-consciousness. To do that, I will first consider Kriegel's different forms of self-consciousness.

¹⁰ This phenomenon is thoroughly discussed by Weiskrantz (1986). The case of the long distance truck driver would also constitute an example for being unconsciously conscious of things. As mentioned above, even after a brief time of his driving, the driver does not recall when he made a turn, or stopped at the traffic lights etc. However since he is accident-free, he must be conscious of the turns, and the traffic lights etc. The driver can be said to be unconsciously conscious of the road, meaning he was conscious of the road but the relevant mental states were not conscious.

6. Forms of self-consciousness

Kriegel (2004) talks about four forms of self-consciousness but only two of them are relevant to the discussion here and those are transitive creature self-consciousness and intransitive state self-consciousness. Transitive creature self-consciousness is described by Kriegel (2004, 187) as ‘the property a creature has when she is conscious of herself and her mental states’ and is exemplified by a statement like, ‘Smith is *consciously* thinking that her car is new.’ The kind of self-consciousness present here is focal and is distinct from the thought about the car (cf. Kriegel 2004, 188).

In contrast to transitive creature self-consciousness, there is intransitive state self-consciousness exemplified by a sentence like ‘Smith is *self-consciously* thinking that her car is new’ (Kriegel 2004, 187). The contrast between Smith *self-consciously* thinking about the car and Smith *consciously* thinking about the car is articulated by Kriegel as the difference between two ways a person can be aware of being in mental state M; peripherally or focally (cf. Kriegel 2004, 189). In consciously thinking that her car is new, the thought about the car and the state of self-consciousness are two distinct states and Smith is focally aware of being in mental state M. In self-consciously thinking that her car is new, there is peripheral awareness of being in a mental state and the state of self-consciousness is an integral part of the thought about the car. Kriegel adds:

When one has a mental state self-consciously, one is aware primarily of the (external) object of the mental state in question. But there is also a more subtle, peripheral awareness of oneself implicit in that state. One is aware of oneself precisely as that state’s *owner*, or *subject*. To say that Smith is thinking that her car is new in a self-conscious sort of way, then, is to say that Smith is implicitly, or peripherally, aware of her *having* the thought, or of the thought being *her own*. (Kriegel 2004, 189)

Consider the famous example of the long-distance truck driver again. For anyone who endorses the state principle, the driver cannot be unconsciously conscious of the road. The driver’s awareness of the road would be possible in virtue of a conscious mental state, viz. his conscious perception. But the driver’s awareness of the road needs to be qualified in a certain way such that the long-distance driver’s awareness of the road can be distinguished from the awareness of a more careful driver, one that can report back on his experience of driving. Given Kriegel’s infrastructure of con-

consciousness, one would say that the difference consists in the long-distance truck driver's being less self-consciously aware of the road compared to the careful driver. This would be supported by Kriegel's (2004, 202, fn. 12) comment that peripheral awareness and focal awareness need not be two exclusive states and that there 'probably is a wide spectrum of awareness, ranging from the very focal to the very peripheral'. But it is not clear how one's awareness of oneself as the owner of a mental state varies in degree, and *when* a more self-conscious way of having an experience finally gives way to a distinct state of focal self-consciousness, thus a conscious way of having an experience. Such issues require explanation in order to make more sense of Kriegel's infrastructure of self-consciousness. More importantly though is Kriegel's argument for the requirement thesis and that's what I'd like to consider now.

7. Critique of the arguments for the requirement thesis

Kriegel's (2004, 198) main argument for the dependence of consciousness on self-consciousness is:

Premise 1: Conscious states are first-person knowable.

Premise 2: First-person knowable mental states must be intransitively self-conscious.

Conclusion: Conscious states are intransitively self-conscious.

This is a valid argument and I do agree with the first premise. In order to critically assess the truth of the second premise it is worth considering what first-person knowable property of conscious mental states amounts to. Kriegel describes the first-person knowable property in the following way:

Suppose I now imagine a camel. The way I know that a camel is what I am right now imagining and the way you know this are very different. There are certain steps you have to go through to attain this knowledge that I can skip. The rule that guides us in applying the predicate "conscious" is that a phenomenon is conscious just in case our knowledge of it is the sort of knowledge I do, and you do not, have of the fact that a camel is what I am right now imagining. (Kriegel 2004, 198)

Basically the first person knowable property has to do with the immediate and non-inferential access one has to her own (conscious) mental

states. The kind of access one has to another's mental states is bound to be inferential. Also one's awareness of another's mental states does not make them conscious and one's inferential awareness of one's own mental states does not necessarily make them conscious either.¹¹ The argument Kriegel (2004, 198) gives for the second premise is that if one were to have a mental state un-self-consciously, that is without awareness of it whatsoever, he says, then one would need to infer its existence, rendering her knowledge of the mental state non-first-personal, in which case the mental state in question would not be conscious. But since first-person knowable mental states are conscious, he concludes that a subject could not have such states un-self-consciously.

In this argument for the second premise, Kriegel equates having a mental state un-self-consciously with having a mental state without any awareness of it, but equating them as such assumes the truth of the requirement thesis and cannot be used to support it. In having a mental state un-self-consciously, peripheral awareness of oneself as the owner of the mental state is missing. But having a mental state without *any* awareness of it – and thus having a non-conscious mental state – is different from having a mental state without peripheral awareness of oneself as the owner of a mental state. One could be aware of the content of a mental state, its intentional or qualitative character without being aware of oneself as the owner of that mental state. Hence awareness of a mental state and peripheral awareness of oneself as the subject of one's mental state cannot be equated. Equating them as such leads to a circular argument for the truth of the conclusion that conscious mental states must be intransitively self-conscious.

It is impossible for a mental state to be conscious and for there to be absolutely no awareness of it. Surely some awareness of the mental state is needed to classify it as a conscious mental state. But the question is whether that awareness necessarily involves self-consciousness. And my answer is no. With regards to perception, Gurwitsch (1941, 330) says that 'In ... dealing with the [perceived] object I am aware of this very dealing.' So some

¹¹ To further illustrate the point, consider Rosenthal's (2005, 310) example of a person who might be aware *that* she is angry but not actually experience it. So, say Sarah is angry but she is not aware of it. Sarah's friend notices that Sarah is angry from the way she behaves and tells her. Sarah is surprised. She claims that she was not experiencing anger, but at the same time counting on her friend's observation, she infers and believes that she is angry. But that doesn't necessarily make her experience anger. In other words, insofar as Sarah's awareness of her anger is inferential, the experience is missing.

awareness of the perception is present but what that awareness holds is open to discussion. Strawson (2009) comments that Gurwitsch's earlier choice of words, viz. 'Being confronted with an object, I am at once conscious of the object and aware of my being conscious of it', is misleading since it leads one to think that awareness of mental goings-on involves awareness of oneself but according to Strawson (2009, 440) '... the phenomenon in question doesn't require any thought of oneself as such, i.e. as subject considered as such.' Strawson refers to the earlier pages of Gurwitsch's work to make things more clear, where Gurwitsch (1941, 327) says 'the subject in his dealing with the [perceived] object, aware as he is of this dealing, is nevertheless in no way aware of his ego, much less of his ego's involvement in his dealing.' In other words, it is possible to conceive the awareness of a mental state and the awareness of oneself as the subject of a mental state as two distinct mental phenomena. But Kriegel makes the mistake of equating the two phenomena, which results in a circular reasoning for the requirement thesis.

The second argument Kriegel (2004, 199) gives for the requirement thesis is due to Rosenthal. In his discussion of Rosenthal's argument, Kriegel raises the question of whether our awareness of conscious mental states is also self-awareness. He refers to Rosenthal who argues that it is not possible to think about a particular mental-state token, as opposed to thinking simply about a type of mental state, without thinking about the subject whose mental state it is. And since conscious mental states are states *we* are aware of, then *our* awareness of them should also include an awareness of the subject whose mental state it is, which is ourselves.

It is a conceptual truth that where there is an experience, there is a subject of experience. After Strawson (2003, 280) I call this the subject thesis. I agree that it is not possible to think about a mental state token in abstraction from its subject. However it is possible that this restriction applies only to the way we think about others' and not about our own mental states. Consequently I suggest that there is an asymmetry between the way I think about *my* mental states and the way I think about another person's mental states. I contend that given this asymmetry, Rosenthal's view cannot be used to support the requirement thesis.

Thinking about a particular mental state token implies thinking about what it would be like for someone to have it. In order to think about a particular mental state token of say the olfactory sensation of recently mowed grass, one needs to think of the olfactory sensation as somebody's

sensation with its subjective character and phenomenology since that's what makes it a particular mental state *token* and not a *type*. But when I think about a particular mental state of *mine*, its subjective character is immediately given to me. I don't need to further think about the subject whose mental state it is in order to understand that it has a unique phenomenology since in having the mental state, its unique phenomenology is already given. So perhaps the special access a subject has to *her* mental states consists precisely in the fact that she does not need to further think of them as belonging to her.

Given the asymmetry described above, I contend that the way we think about a particular mental state token does not necessarily provide the right insight into the way we think about *our* mental state tokens. This is most likely a consequence of the gap between the immediate and non-inferential knowledge one has concerning her own (conscious) mental states as opposed to the inferential knowledge another has concerning those conscious mental states.

One might be curious about the nature of the gap involved. For my friend to know about the olfactory sensation that I'm having, she either needs to be told by me or she needs to observe me take in deep breaths with a pleasant look on my face and also realize that the grass has been recently mowed etc. I, on the other hand know it immediately. As mentioned before by Kriegel (2004, 198) too, there are no intermediate steps between my having the olfactory sensation and my knowing that I'm having such a sensation. This constitutes essentially an epistemological gap between my knowledge of my conscious mental states and someone else's knowledge of them. There is yet another gap that seems to cut deeper than the mere epistemological one. Even if my friend knows, by whatever means, that I'm having an olfactory sensation, there is no way she will know what it is like for me to have that sensation. What-it-is-likeness of a conscious mental state is private to the subject who has it.¹² While the epistemological gap mentioned above is about how my friend knows that I'm having e.g. an olfactory sensation, this gap is about the knowledge of what it is like for me to have that olfactory sensation which is never fully acquired by my friend.¹³

¹² And as mentioned before, it is what makes a mental state a particular mental state token and not a type.

¹³ Since the gap concerns knowledge of something, viz. knowledge of what it is like to have a particular olfactory sensation, it can still be said to be of an epistemological kind,

8. Conclusion

To conclude, the requirement thesis fails for the following reasons. Firstly, granted that there is such a thing as peripheral awareness of oneself as the owner of one's mental state, even though I discussed its dubitability, such awareness does not have to be an integral part of all conscious states. I suggested that by endorsing a higher-order theory of consciousness, it is possible to separate self-consciousness from the content of the conscious target state and attribute it to the higher-order state. Secondly, I suggested that Kriegel does not have the option of explaining state consciousness in virtue of transitive consciousness since he denies the possibility of being unconsciously conscious of things and he explains state consciousness in virtue of self-consciousness instead. Then granted that the reasons I gave for rejecting the state principle are tenable, there may be no reason to think that consciousness depends on self-consciousness. Thirdly, the argument Kriegel gives for the requirement thesis is not sound since the argument he gives for the truth of the second premise that all first-person knowable states must be intransitively self-conscious assumes its truth rather than show it. And finally Kriegel's appeal to Rosenthal's argument with regards to the subject thesis does not succeed either because there is an asymmetry in the way a subject thinks about a particular mental state token when that mental state is hers as opposed to someone else's, which is expected given the first-person knowable property of conscious mental states.

References

- ARMSTRONG, D. (1968): *A Materialist Theory of the Mind*. London: Routledge.
- BLOCK, N. – FLANAGAN, O. – GÜZELDERE, G. (eds.) (1997): *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- BYRNE, A. (2004): What Phenomenal Consciousness is Like. In: Gennaro, R. (ed.): *Higher Order Theories of Consciousness: An Anthology*. Philadelphia: John Benjamins Publishing Co., 203-226.
- CASTON, V. (2002): Aristotle on Consciousness. *Mind* 111, 751-815.

but since it is the kind of knowledge that can never be fully acquired by the third person, I think the gap cuts deeper than an epistemological kind but I don't think it qualifies as an ontological one.

- DAMASIO, A. (1999): *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harvest Books.
- GALLAGHER, S. – ZAHAVI, D. (2008): *The Phenomenological Mind*. New York: Routledge.
- GENNARO, R. J. (2008): Representationalism, Peripheral Awareness, and the Transparency of Experience. *Philosophical Studies* 139, 39-56.
- GURWITSCH, A. (1941): A Non-Egological Conception of Consciousness. *Philosophy and Phenomenological Research* 1, 325-338.
- FRANKFURT, H. G. (1971): Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68, 5-20.
- FLANAGAN, O. (1992): *Consciousness Reconsidered*. Cambridge, MA: MIT Press.
- GOLDMAN, A. (1970): *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.
- HUME, D. (1739-40/1978): *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge and P. H. Nidditch, 2nd ed. Oxford, UK: Oxford University Press.
- KRIEGEL, U. (2004): Consciousness and Self-Consciousness. *The Monist* 87, 182-205.
- KRIEGEL, U. (2009): Self-Representationalism and Phenomenology. *Philosophical Studies* 143, 357-381.
- LOCKE, J. (1694/1975): *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.
- LYCAN, W.G. (1996): *Consciousness and Experience*. Cambridge, MA: MIT Press.
- PARFIT, D. (1984): *Reasons and Persons*. New York: Oxford University Press.
- ROSENTHAL, D. (1997): A Theory of Consciousness. In: Block, N. – Flanagan O. – Güzeldere, G. (eds.): *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, 729-754.
- ROSENTHAL, D. (2005): *Consciousness and Mind*. New York: Oxford University Press.
- SCHEAR, J.K. (2009): Experience and Self-Consciousness. *Philosophical Studies* 144, 95-105.
- STRAWSON, G. (2003): What's the Relation between an Experience, the Subject of the Experience and the Content of the Experience. *Philosophical Issues* 13, 279-315.
- STRAWSON, G. (2009): Against 'Corporism': The Two Uses of 'I'. *Organon F* 16, 428-448.
- WEISKRANTZ, L. (1986): *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.
- ZAHAVI, D. (2006): *Subjectivity and Selfhood: Investigating the First Person*. Cambridge, MA: MIT Press.

Infallibility about the Self

GLEN HOFFMANN

Department of Philosophy. Faculty of Liberal Arts and Professional Studies York University
416 Ross South. Canada
ghoffman@yorku.ca

RECEIVED: 08-02-2013 • ACCEPTED: 16-05-2013

ABSTRACT: On a common formulation, rationalist infallibilism is committed to two main theses: (i) ‘analytic a priori infallibilism’ and (ii) ‘synthetic a priori infallibilism’. According to thesis (i), a relatively wide range of analytic a priori propositions can be infallibly justified. According to thesis (ii), a relatively wide range of synthetic a priori propositions can be infallibly justified. In this paper, I focus on rationalist infallibilism’s second main thesis, what is being called ‘synthetic a priori infallibilism’. I argue that synthetic a priori infallibilism, and by extension rationalist infallibilism, is untenable. In particular, exploring what seems to be the only potentially plausible species of synthetic a priori infallibility, I reject the infallible justification of propositions about the self.

KEYWORDS: Cogito – dubito – infallibility – privileged access – rationalism – self-knowledge.

1. Introduction: rationalist infallibilism and propositions about the self

On a traditional formulation, rationalist infallibilism is a doctrine committed to two main theses: (i) ‘analytic a priori infallibilism’ and (ii) ‘synthetic a priori infallibilism’. According to thesis (i), a relatively wide range of analytic a priori propositions can be infallibly justified. According to thesis (ii), a relatively wide range of synthetic a priori propositions can be infallibly justified. In sum, then, for the rationalist infallibilist two main categories of propositions are susceptible of infallible justification: (i) analytic propositions such as logical, mathematical, and ‘conceptual’ propositions, and (ii), synthetic propositions such as those concerning the external world

and the self. Though rationalist infallibilism is increasingly falling out of favor, there is no shortage of notable 20th-21st century adherents of one or more of the two species of infallible a priori justification. Among others, Fumerton (2001), McGrew (2003), and McGrew et al. (2007) still profess the a priori infallibility of some analytic propositions, while Burge (1986; 1988; 1996) and Lewis (1996) have recently defended the a priori infallibility of some synthetic propositions (see Hoffmann 2011, 241).

In this paper, I examine rationalist infallibilism's second thesis, so-called 'synthetic a priori infallibilism'.¹ More specifically, focus will be restricted to what are likely the only potentially viable candidates for synthetic a priori infallible justification, propositions about the self. In the course of this exploration, I reject the possibility of the infallible a priori justification of propositions about the self. Moreover, on the seemingly reasonable assumption that propositional knowledge involves propositional justification, that for a proposition to be known it must in some sense be justified, the implication is that infallible a priori *knowledge* about the self is also out of reach.² On the other hand, this anti-infallibilist argument does not eliminate the possibility of a priori justification and or knowledge (about the self or otherwise) *in general*, since it leaves open that some species of *fallible* a priori justification/knowledge might be within reach.

2. Infallibility about the self

One primary candidate for a priori infallibility, perhaps the most promising one, is in the domain of syntheticity.³ To name just a few, proposals

¹ I undertake a similar project in Hoffmann (2012).

² In what follows, though, discussion will primarily focus on *justification* instead of *knowledge*, in order to avoid questions surrounding the definition or analysis of the concept of *knowledge*.

³ I assume here that there is (i) an analytic/synthetic distinction and (ii) an a priori/a posteriori distinction that can be drawn along relatively traditional lines: (i) propositions whose truth depends exclusively on meaning/propositions whose truth doesn't depend exclusively on meaning, and (ii) propositions justifiable independently of experience/propositions not justifiable independently of experience. If it turns out (i) and/or (ii) is false, this is of little concern, since surely skepticism that there is a special domain of analytic and/or a priori propositions is far more likely to be harmful to infallibilist than anti-infallibilist aspirations.

for synthetic a priori infallible justification have included propositions about material objects, substances, properties, and the self (see Hoffmann (2011, 246). For example, the first three proposals for synthetic a priori infallible justification have been canvassed in some form or another by both Price (1953) and Lewis (1996). There seems to be a consensus, though, that these kinds of proposals for synthetic a priori infallibility have not survived sustained review. In any event, whatever the merits of the case for synthetic a priori infallibility about various external world phenomena, discussion here is confined to a class of synthetic propositions whose a priori infallibility still seems to be a live issue – propositions about the self.⁴

Traditionally, propositions about the self have been at the heart of the infallibilist program in epistemology. For the infallibilist, (some) propositions about the self are supposed to be self-justifying or self-verifying in the basic sense that their mere belief establishes their truth. In specifically rationalist terms, a proposition about the self is supposed to be self-justifying in the sense that its belief is *a priori* sufficient to establish its truth. Descartes' proposals about first-person mental states (1996/1641, 80ff) are the *locus classicus* for a priori infallibility about the self (see Hoffmann 2011, 47; 2012, 57f):

(Cogito) I am thinking (therefore I exist as a thinking thing)

and along the same lines

(Dubito) I am doubting (therefore I exist as a doubting thing).

Independent of whether (Cogito) and (Dubito) are self-justifying at all, i.e., whether belief in them establishes their truth (in some potentially *fallible* sense), is the question of whether they are susceptible of *infallible* justification, i.e., whether belief in them establishes their truth infallibly or indefeasibly. For the infallibilist the pertinent question is whether Cogito-like propositions, or any other proposition about the self of significant cognitive interest, can be justified in a way that is truth-entailing? More spe-

⁴ Other proposals for synthetic a priori infallibility include putative ontologically necessary truths to the effect that ⟨something exists⟩ or ⟨at least one thing exists⟩, and weakened Cogito-like propositions such as ⟨thought exists⟩ or ⟨thinking exists⟩. I leave open that one or more of these synthetic propositions is an infallibly justified a priori truth, while bearing in mind their indisputably negligible cognitive import in comparison with classical Cogito-like propositions (to be considered next in the text).

cifically, can these propositions be justified in a way that confers on them absolute warrant, or a probability of 1, where ‘probability’ is read in putatively *objective* terms, and ‘truth’ is read in standard realist terms?⁵

One rationale for rejecting infallible a priori self-justifying propositions stems from a firmly entrenched epistemological principle. The received view at present seems to be that justification for propositions about the self can never be *determinately* reflexive or self-referring (e.g., McDonald 1998). For example, if I have the thought ⟨I am currently thinking⟩, the warrant for the truth of the proposition expressed by this thought cannot be obtained by the mere having of this thought. Instead, it is reasoned, it must be secured via some kind of reflection on the thought. Moreover, insofar as reflection implies the existence of two, numerically distinct thoughts present in this case (a first and a second-order thought), it is unclear how the second-order thought ⟨I think that I am thinking⟩ can infallibly justify its first-order constituent ⟨I am thinking⟩. The upshot of this line of argument is that Cogito-like propositions such as ⟨I am thinking⟩, or even iterated Cogito-like propositions such as ⟨I think that I am thinking⟩, do not appear susceptible of infallible justification.

But this nascent anti-infallibilist line of argument is less than decisive as presently formulated since the Cartesian infallibilist about the self has at her disposal a steadfast reply to it. For Descartes (Cogito) and (Dubito) are infallible a priori self-justifying propositions since they are instances of *direct* ratiocination that supply unmediated acquaintance with the reality of the thinking self (Descartes 1996/1641, 80–81). David Lewis (1996, 564ff) similarly seems to construe (Cogito) and (Dubito) as pure rational intuitions, minus the Cartesian metaphysics (about subjects) and commitment to any form of internalism, intuitions that furnish unmediated access to the reality of the subject (see Hoffmann 2011, 248; and 2012, 57ff for more on this).⁶

⁵ More on this construal of *infallibility* and its bearing on the possibility of a priori infallibility about the self in §5.

⁶ One might object to direct, time-sensitive rational intuitions as a source of justification/knowledge on *pro tanto* empirical grounds. If one’s direct ‘acquaintance’ with subjective reality is restricted to the specious present and vanishes on second-order reflection, it might be argued, Cartesian rational intuitions cannot confer rational justification on propositions about first person mental states (or any other content-bearing state). Assuming judgment is diachronic and rational intuitions are synchronic, intuitions indexed to time fragments lack normative force since they disappear just as one registers

Needless to say, no claim to the unlimited scope of infallible a priori justification about the self is made here.⁷ For Descartes and Lewis the possibility of infallibility about the self is conditional upon the satisfaction of explicit desiderata, respectively, Descartes' clearness and distinctness criterion and Lewis' numerous criteria for ignoring counterfactual possibilities—states of affair that would undermine the truth of the proposition at issue. But in situations where the relevant criteria are met one can be acquainted with one's mental states directly and immediately,⁸ in a way that eliminates the need for second-order reflection. For Descartes, although Lewis is more circumspect on this point, this includes mental states with propositional content, such as beliefs about the self (e.g., ⟨I am now thinking⟩). It follows, on this general outlook, that there is at least a small class of propositions about the self that are determinately reflexive, and consequently, whose belief guarantees their truth (see also Burge 1986; 1988).

a judgment about them. I reserve judgment on this kind of objection since it enlists controversial premises concerning the nature of memorial cognition and knowledge, e.g., that they are temporal tokens, indexed to their time of conception. For discussion of the nature of memorial cognition and knowledge and its bearing on the viability of first person a priori judgments, see Boghossian (1989), Ludlow (1999), Brueckner (1997), and Goldberg (1997).

⁷ For ease of exposition, 'infallibility about the self' will hereafter largely be used as shorthand for 'infallible a priori justification about the self' and 'infallible a priori knowledge about the self'.

⁸ The general notion of direct a priori acquaintance demands close inspection. In particular, beyond indistinct formulations of the phenomenon in terms such as 'the light of reason', 'intuitive obviousness' or the like, important issues arise concerning the modal status, empirical status, and truth-conduciveness of pure rational intuitions. Typically, rational intuitions are deemed necessary, impervious to empirical data, and infallibly truth-conducive. For example, Bonjour (2005, 99), in deference to the first two criteria, declares rational intuitions provide "direct or immediate insight into the truth, indeed the necessary truth, of the relevant claim They are thus putative insights into the essential nature of things or situations of the relevant kind, into the way that reality in the respect in question *must* be". Turri (2011), contra the traditional view, defends the possibility of some kinds of *contingent* a priori rational intuitive knowledge. The bearing of such issues on the possibility of infallibility about the self will become especially evident in §4.

3. Infallibilism and direct ratiocination

In consequence, a more direct confrontation of Cartesian infallibilism about the self is required. As a first attempt, it might be argued against the Cartesian infallibilist that Cogito-like self-justifying propositions cannot be rationally justified *in esse*. At bottom, direct acquaintance with the mental is a notoriously perplexing epistemic phenomenon (see Bonjour 2003, 17f for a discussion of some of the main concerns): it is an enigma how brute mental states (about the self or otherwise) can be vehicles of rational justification. A pivotal concern is mental states of the form ⟨I'm experiencing like *this*⟩, where *this* designates a primitive feature of experience, look devoid of cognitive content—the kind of content that can figure in judgments.

To be sure, mental states of this kind lack *descriptive* content since they take the form ⟨I'm experiencing like *that*⟩ instead of ⟨I'm experiencing like *that* and not like *that*⟩. Such mental states, it can scarcely be denied, do not encode either definite or indefinite descriptions, on any remotely plausible construal of 'description'.⁹

The crucial question, then, is whether the mental states under consideration possess some kind of 'minimal', non-descriptive cognitive content, content that can play a justificatory role. One reason for being skeptical about this possibility is there doesn't seem to be an intelligible account of the reference of ostensibly primitive mental states. On the surface, the referent of the demonstrative *this* in such mental states is intractable. For starters, since it is fundamentally deictic, *this* obviously does not designate a context-independent state of affairs. More importantly, the referent of *this*, since it is necessarily a primitive feature of reality, looks to be completely *subject*-dependent, meaning it doesn't even seem to have a *context-relative* semantic value—a semantic value stable across some set of contextually relevant variables. If this is correct, the mental states in question can be viewed as purely phenomenal, and consequently, as lacking cognitive significance. It would follow, in this case, that they cannot furnish support for particular hypotheses or claims, that they cannot raise the credibility of

⁹ Some infallibilists (e.g., McGrew 2003; McGrew et al. 2007), typically of an internalist bent, explicitly reject this seemingly innocuous thesis. A more direct defense of this thesis is provided in §4, in our discussion of the infallibilist's appeal to the phenomenon of *direct reference* or *acquaintance* in application to beliefs about mental states.

a proposition, belief, or utterance in comparison with any other proposition, belief, or utterance.

Naturally, in response, the infallibilist will contend that the referent of the demonstrative *this* in primitive mental states is, in a fundamental sense, non-discursive: i.e., it lacks conceptual constituents, complex structure, discrete components, etc. In this connection, the infallibilist might appeal to a special kind of rudimentary cognition (the likes of which has been articulated by Husserl and Brentano, among others), a kind of cognition that is prior to language or possibly even to conceptualization, as the source of self-justifying infallible belief about the self. But the view that there can be pre-conceptual, or what has been called ‘semi-judgmental’ mental states that infallibly justify propositions about the self looks patently untenable. A pivotal concern is it seems virtually undeniable that mental states that are vehicles of justification must in *some* sense be representational (Bonjour defends this contention in Bonjour – Sosa 2003, 20). This, of course, is not to say that the representational content of justificatory mental states need be anything like a truth-evaluable propositional thesis or assertion, or that it be explicitly formulatable in terms of something like ‘comparison’ or ‘contrast’ classes. At a minimum, though, such mental states must have some kind of representational content (or perhaps *informational* content, if one prefers), i.e., they must depict things as being one way rather than another, in order to function as justifiers. But since the mental states in question lack such content, they cannot supply warrant for propositions about the self, let alone *absolute* warrant. The irresistible inference is that direct ratiocination about the self cannot be the spring of a priori infallibility about the self.¹⁰

4. Direct ratiocination and conceptual acquaintance

Another reply to our dismissal of infallibilism about the self involves the attempt to develop a more theoretically motivated account of direct ratioci-

¹⁰ The argument I advance here has some resemblance to a widespread anti-foundationalist argument according to which a mental state cannot both (i) have cognitive import and (ii) provide a foundation for propositional knowledge. Here I take no stance on the anti-foundationalist polemic. Since we haven’t ruled out the possibility that there are defeasible foundations for propositional knowledge that can be secured in a manner that does not involve direct acquaintance with mental states, our argument has no direct bearing on fallibilist variants of foundationalism.

nation. In brief, in this vein, one might defend the possibility of a kind of direct conceptual acquaintance on the basis of an account of first person rational intuitions, an account that might be considered in some sense phenomenologically faithful. For example, one might exploit a view of first person rational intuitions along the lines of Bonjour's (1996; 2003) and Bealer's (2000) *appearance* model to this effect, where such intuitions involve a kind of *seeing* or *visualizing* of subjective reality, rather than a conceptual or propositional grasp of it (n.b.: neither Bonjour nor Bealer are infallibilists about the self). Since on this model Cogito-like propositions take the general form $\langle I'm \text{ experiencing/feeling/being-appeared-to like } that, \text{ therefore } that \rangle$, where *that* does not refer to a propositional object, they might be thought to secure direct (non-discursive) acquaintance with the reality of the thinking/experiencing self. The infallibility of Cogito-like propositions in this case is thus supposed to stem from the fact that they enjoin a kind of direct reference or acquaintance with mental states (McGrew et al. 2007, 132; Burge 1986; Burge 1988), analogous to the way some have claimed one might be directly acquainted with features of the external environment through perception.

Importantly, this variant of infallibilism about the self is not tied to internalism (as in McGrew et al. 2007) – where the content of and/or the justification conditions for one's mental states are in some sense 'subject-internal' (e.g., introspectable, open to consciousness, transparent, luminous, etc.). A number of self-proclaimed externalists, including Burge (1986; 1988), Lewis (1996), Heil (1988), and Davidson (1984; 1987), have defended an account of a priori self-knowledge along the lines of the account currently being sketched (though seemingly only the former two can rightly be viewed as infallibilists). Burge (1988, 659-660), most notably, has developed an influential account of infallible a priori self-knowledge that is non-discursive in the sense under consideration, one in which there is a non-inferential link between conception and apperception of the self (and wherein the infallibility of certain beliefs (or thoughts) about the self is guaranteed by their reflexive character – by the fact that they 'logically lock' onto their first-order constituents).

In my view, though, there are powerful reasons for rejecting the refined formulation of direct mental acquaintance, whether internalist or externalist in orientation. But even if it is possible to directly apprehend the content of a mental state, or to have indefectible 'privileged access' to certain kinds of mental phenomena, this looks to be small consolation to the infal-

libilist. On our central line of argument, mental states with which one is directly acquainted *ipso facto* lack cognitive import since they are devoid of representational or informational content. Since mental states of the form ⟨it seems, appears, or feels to me that *such-and-such*⟩ concern “those conditions of the subject, whatever they are, which are accessible to the subject, whenever they obtain...” (Williamson 2000, 15), such states are veridical *by fiat*. Primitive mental states must, on this score, be considered strictly phenomenal, and quite possibly, beyond the limits of what is expressible.

Williamson (2000, 14ff), among others, has also argued for the epistemic triviality of primitive or what he calls ‘luminous’ mental states, on the grounds that they lack truth-tracking conditions ensuring their ‘safety’ or ‘secure’ application. Bypassing Williamson’s externalist-motivated *universal* anti-luminosity argument,¹¹ my principal contention is that primitive mental states, insofar as they concern the putative content of *immediate* experience or the character of *bare* sentiment, are cognitively vacuous. Whatever import they might have, such states are effectively nebulous from the standpoint of epistemology: they can neither figure in judgments nor confer rational justification on beliefs, statements, utterances, or any other bearer of propositional content.

Now obviously there are fundamentally divergent views about the nature, scope, and limits of cognitively significant mental content. While the argument currently being developed leaves open that there can be cognitively significant mental states (though not via a direct or unmediated grasp of their content), no substantive position about the nature of cognitively significant mental content is being adopted here. Instead, my specific contention is that there is no satisfactory construal of direct mental acquaint-

¹¹ Williamson employs the semi-technical notion *luminosity* to refine a conception of *transparency* in application to mental states. Williamson (2000, 95) defines *luminosity* as follows: A condition C is defined to be *luminous* if and only if (L) holds:

(L) For every case α , if in α C obtains, then in α one is in a position to know that C obtains.

He then advances an argument to show that no non-trivial condition is luminous. While the strategy is an interesting one, I take a different route (albeit one that is consistent with and intersects with Williamson’s strategy). Since luminosity is an explicitly *reliabilist* property, by Williamson’s own admission (Williamson 2000, 95f), it is not clear his anti-luminosity argument directly confronts any *internalist* infallibilist position that jettisons reliability constraints on rational justification/knowledge (e.g., see McGrew et al. 2007, 132ff).

ance where a mental state one is *directly* acquainted with can serve as the *infallible justifier* of a Cogito-like proposition or any other proposition about the self of significant cognitive interest.

5. Semantic infelicity

A final reply to our dismissal of infallibilism about the self calls into question our position via a revisionary maneuver.¹² This proposal tries in effect to *intrinsically* guarantee infallibility about the self, typically, by enlisting some kind of semantic-error theory of the general kind recommended by Wittgenstein (1969, 10ff). On this proposal, the putative fallibility of Cogito-like propositions is the result of some variety of semantic or conceptual infelicity. The apparent fallibility of such propositions, it is urged, is a product of linguistic misappropriation that involves the violation of conventions governing concepts about the mental. Based on the way expressions such as *I*, *think*, *doubt*, *exist*, *thinking thing* and *doubting thing* are used, the semantic error theorist reasons, it is rationally impossible to dissent from (Cogito) or (Dubito). Such dissent, it is maintained, constitutes an infringement of de facto rules concerning the proper use of language for rational subjects (where ‘rational’ itself is interpreted in some pragmatic manner). For instance, for Wittgenstein (1969) it would involve breaching implicit rules constraining the use of discrete fragments of language, i.e., making incorrect moves in what are dubbed ‘language games’. For others such as Davidson (1984; 1987) it would involve assenting to propositions (e.g., $\sim\langle I \text{ exist as a thinking thing} \rangle$) that conflict with one’s overall rational commitments,¹³ thereby generating some kind of doxastic contradiction.¹⁴ For those with Kantian inclinations it would involve endorsing proposi-

¹² This section draws on previous work of mine (see Hoffmann 2011, 245-246). The difference is that in the earlier article I consider a revisionary reply to *analytic a priori infallibilism* instead of *synthetic a priori infallibilism*.

¹³ Davidson, though, is not a proponent of semantic-error theory, nor does he profess the infallibility (as it has been defined here) of (Cogito), (Dubito), or any other proposition about the self. So it is doubly misleading to characterize him as a champion of the semantic infelicity defense of infallibilism about the self.

¹⁴ In recent work, Burge (1996) adopts a similar strategy to Davidson, one that significantly departs from Burge’s earlier defense of infallibilism about the self (1986; 1988).

tions that conflict with the postulates of Kant's transcendental deduction of the unity of apperception in consciousness, formulated in the first *Critique*.

Whatever the merits of semantic-error theory and similar outlooks, it looks to be an unpromising candidate for resurrecting infallibilism about the self. Such a viewpoint is borne out by disinterring some basic implications of the semantic infelicity proposal. Suppose it turns out the apparent fallibility of Cogito-like propositions stems from semantic impropriety of the general kind under consideration. In this instance what has been shown is that it is rationally impossible, incoherent, or inconsistent (or some combination of these things) to doubt such propositions. On the surface, though, it would not have been established that (Cogito) and (Dubito) are *logically* indefeasible in the sense at issue (see also Hoffmann 2011, 245). As I've argued elsewhere (see Hoffmann 2011, 245f), the impossibility of *doubting* a proposition is equivalent to the impossibility of a proposition's *falsity* only on the condition that logical possibility is a species of some kind of epistemic possibility, i.e., where (roughly) what is logically possible is identical to the set of propositions that is, at least in principle, knowable. Now while the logical possibility \leftrightarrow epistemic possibility thesis cannot be dismissed outright, and is typically endorsed in conjunction with some kind of semantic-error theory, this scarcely seems to threaten the anti-infallibilist argument currently on offer.

As it happens, granting the apparent fallibility of (Cogito) and (Dubito) is the product of semantic impropriety, and the concomitant logical possibility \leftrightarrow epistemic possibility thesis, in actuality involves recasting the infallibilism/fallibilism debate in a way that leaves fallibilism about propositions about the self unscathed. As I've argued elsewhere (see Hoffmann 2011, 246), this basic outlook, defensible or not, forecloses on the possibility of infallibility about the self since it involves redefining the alethic concepts of truth and falsity in explicitly anti-realist terms. The contention that (Cogito) and (Dubito) can be infallibly justified becomes essentially insensible on theories that dispute realist construals of expressions whose subject terms refer to mental phenomena, and that accordingly, contest the reality of the mental. If there is no fact of the matter about our mental states independent of the way we employ mental concepts, the infallibility/fallibility of Cogito-like propositions is a question proscribed at the outset. Davidson, for his part, explicitly acknowledges this point since, as mentioned, his account of first-person authority and privileged access does not aspire to anything in the vicinity of infallibility, indefeasibility, incorrigibility, or the like. Wittgenstein similarly

recognizes this consequence of the basic outlook under consideration. The semantic view articulated in *On Certainty*, Wittgenstein concedes, rules out the possibility of certainty as the notion is traditionally understood, i.e., from something like a language-independent standpoint.

Field, on the other hand, obliquely suggests an alternative interpretation of matters. Field (2005) has recently argued in Quinean fashion that all ‘high level’ debates about the status of purported a priori knowledge, e.g., in the areas of mathematics and logic, are best treated pragmatically – via the inferential role, illocutionary force, doxastic entitlement, or some comparable pragmatic dimension of the propositions in question. One might argue, along these general lines, that if there is no interesting question concerning the objective, theory-neutral infallibility of propositions about the self (or any other a priori proposition), and if there is a compelling pragmatic defense of infallibility about the self (as some have maintained), then infallibility about the self par excellence, the only a priori infallibility to aspire to, *is* within reach. This is correct as far as it goes but doesn’t directly confront the anti-infallibilist argument on offer. Granting the question of a priori infallibilism is a pragmatic one, and that within this framework certain kinds of a priori propositions such as those about the self are infallibly justified, only bolsters the conviction that a priori propositions about the self are not amenable to *objective* infallible justification, where *objective* is read in standard realist terms. Similar to the other semantic infelicity proposals, this one preemptively outlaws exactly the kind of infallibility about the self presently under consideration, the kind advocated by the garden variety rationalist infallibilist. Field, as it happens, is seemingly in agreement on this point (as Wittgenstein and Davidson are).

6. Concluding remarks

If I am correct, contra Descartes (1996/1641), Burge (1986; 1988), Lewis (1996) and McGrew et al. (2007), there is no proposition about the self that can be infallibly justified a priori. On the view articulated here, reason cannot furnish absolute warrant for (Cogito), (Dubito), or any other proposition about the self of significant cognitive interest. In the end, the principal defect of infallibilism about the self is found in its account of direct mental acquaintance, and in particular, in its failure to supply an account of direct mental acquaintance that enables primitive mental states to infallibly justify propositions about the self. Moreover, the class of revisionary ma-

neuers typically proposed also fail to deliver infallibilism about the self. Construing dissent from Cogito-like propositions as a form of semantic/conceptual infelicity resurrects infallibility about the self only at the unacceptable cost of redefining the terms of the infallibilist/fallibilist debate (as traditionally conceived).

The upshot of the foregoing discussion is that rationalist infallibilism and one of its two main sub-theses, synthetic a priori infallibilism, are indefensible doctrines.¹⁵ On the other hand, the arguments advanced in this paper have no direct bearing on any *non-rationalist* infallibilist doctrine, nor, as far as I can see, do they provide anything resembling a template for constructing a general argument against all forms of infallibilism. Moreover, the position defended here clearly leaves open the possibility that some species of *fallible* a priori justification may be within reach since no real reason has been given for rejecting any fallibilist model of a priori justification.

References

- BEALER, G. (2000): A Theory of the A Priori. *Pacific Philosophical Quarterly* 81, 1-30.
- BOGHOSSIAN, P. (1989): Content and Self-Knowledge. *Philosophical Topics* 17, 5-26.
- BONJOUR, L. (1996): Plantinga on Knowledge and Proper Function. In: Kvanvig, J. (ed.): *Warrant in Contemporary Epistemology: Essays in honor of Plantinga's Theory of Knowledge*. Lanham, MD: Rowman & Littlefield.
- BONJOUR, L. (2005): In Defense of the A Priori. In: Steup, M. – Sosa, E. (eds.): *Contemporary Debates in Epistemology*. Malden, Mass.: Blackwell Publishing, 98-105.
- BONJOUR, L. – SOSA, E. (2003): *Epistemic Justification*. Malden, Mass.: Blackwell Press.
- BRUECKNER, A. (1997): Externalism and Memory. *Pacific Philosophical Quarterly* 78, 1-12.
- BURGE, T. (1986): Individualism and Psychology. *Philosophical Review* 45, 3-45.
- BURGE, T. (1988): Individualism and Self-Knowledge. *Journal of Philosophy* 85, 649-663.
- BURGE, T. (1996): Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society* 96, 91-116.
- DAVIDSON, D. (1984): First Person Authority. *Dialectica* 38, 101-112.
- DAVIDSON, D. (1986): A Coherence Theory of Truth and Knowledge. In: Lepore, E. (ed.): *Truth and Interpretation, Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- DAVIDSON, D. (1987): Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association* 60, No. 3, 441-458.

¹⁵ In my (2011), I also argue for the implausibility of rationalist infallibilism, although along different lines. Moreover, in that work, I argue against *both* sub-theses of rationalist infallibilism, not merely synthetic a priori infallibilism.

- DESCARTES, R. (1996/1641): *Meditations on First Philosophy*. (Trans.) J. Cottingham. Cambridge: Cambridge University Press.
- FIELD, H. (2005): Recent Debates about the A Priori. In: Gendler, T. – Hawthorne, J. (eds.): *Oxford Studies in Epistemology*. New York: Oxford University Press, 69-88.
- FUMERTON, R. (2001): Classical Foundationalism. In: Depaul, M. (ed.): *Resurrecting Old-Fashioned Foundationalism*. Lanham: Rowman and Littlefield.
- GOLDBERG, S. (1997): Self-Ascription, Self-Knowledge, and the Memory Argument. *Analysis* 57, 211-219.
- HEIL, J. (1988): Privileged Access. *Mind* 98, 238-251.
- HOFFMANN, G. (2011): Two Kinds of A Priori Infallibility. *Synthese* 181, 241-253.
- HOFFMANN, G. (2012): Infallible A Priori Self-Justifying Propositions. *Croatian Journal of Philosophy* 12, No. 1, 55-68.
- LEWIS, D. (1996): Elusive Knowledge. *Australasian Journal of Philosophy* 74, No. 4, 549-567.
- LUDLOW, P. (1999): First Person Authority and Memory. In: De Caro, M. (ed.): *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy*. Dordrecht: Kluwer Academic Publishers.
- MACDONALD, C. (1998): Externalism and Authoritative Self-Knowledge. In: Wright, C. – Smith, B. – MacDonald, C. (eds.): *Knowing Our Own Minds*. Oxford: Oxford University Press.
- MCGREW, T. (2003): A Defense of Classical Foundationalism. In: Pojman, L. (ed.): *The Theory of Knowledge*. 3rd Edition. Belmont: Wadsworth.
- MCGREW, T. – MCGREW, L. (2007): *Internalism and Epistemology: The Architecture of Reason*. New York: Routledge.
- PRICE, H. H. (1953): *Thinking and Experience*. Cambridge, Mass.: Harvard University Press.
- TURRI, J. (2011): Contingent A Priori Knowledge. *Philosophy and Phenomenological Research* 83, No. 2, 327-344.
- WILLIAMSON, T. (2000): *Knowledge and its Limits*. New York: Clarendon Press.
- WITTGENSTEIN, L. (1969): *On Certainty*. Anscombe, E. – Von Wright, G. H. (eds.). Oxford: Basil Blackwell.

Elipsa¹

MILOŠ KOSTEREC

Katedra logiky a metodológie vied. Filozofická fakulta. Univerzita Komenského v Bratislave
Šafárikovo nám. 6. 814 99 Bratislava. Slovenská republika
milos.kosterec@gmail.com

ZASLANÝ: 23-02-2013 • AKCEPTOVANÝ: 25-03-2013

ABSTRACT: The paper deals with semantic content of elliptic sentences and its relation to semantic content of the corresponding non-elliptic sentences. On the basis of certain kinds of examples it is shown that syntactic theories of ellipsis have serious limits. It is also demonstrated that the so-called Property Theory, which is an example of a semantic theory of ellipsis, bears serious limitations. Another semantic theory, namely that of Minimal Indexicalism, is analyzed thereafter. Theoretical tools of the theory – in particular, its criteria of linguistic expressions identity and three layers of content – that are vital to its handling of ellipsis are discussed in some detail. Finally, a new theory of ellipsis based on Transparent Intensional Logic is proposed and argued for.

KEYWORDS: Antecedent – ellipsis – linguistic expression – substitution – variable.

1. Úvod

Predstavme si jednoduchý rozhovor:

A: Kto s kým včera sedel v kine?

B: Peter sedel s Máriou a Jožo s Anetou.

¹ Táto štúdia vznikla na Katedre logiky a metodológie vied Filozofickej fakulty UK v Bratislave v rámci projektu podporeného grantom VEGA č. 1/0046/11 *Sémantické modely, ich explanačná sila a aplikácia*. Chcem poďakovať Marii Duží, Pavlovi Maternovi a Mariánovi Zouharovi za kritické pripomienky k predchádzajúcim verziám článku.

Každý zrejme porozumel, že Peter sedel s Máriou a Jožo sedel s Anetou. Zdá sa, že keby hovorca B odpovedal

(1) Peter sedel s Máriou a Jožo sedel s Anetou,

povedal by to isté čo v predchádzajúcom príklade, len niektoré slová vynechal. Takéto vynechanie sa nazýva *elipsa*. V tomto článku sa pokúsim obhájiť nasledovné tvrdenia: Rozdiel medzi dvomi odpoveďami hovorcu B nespočíva iba vo vynechaní nejakého slova. Zároveň tieto dve odpovede hovorcu B na položenú otázku nemajú ten istý význam. Keďže termín „elipsa“ má viac ako jedno chápanie, najprv spresním, o čom budem hovoriť.

Stephen Neale vo svojej štúdii o neúplných deskripciách upozornil na to, že výrazy „elipsa“ či „eliptický“ sa používajú minimálne tromi odlišnými spôsobmi:

Domnievam sa, že spomedzi viacerých vzájomne sa prekrývajúcich a prepojených použití slov „eliptický“ a „elipsa“ možno rozlíšiť a spresniť tri použitia, ktoré zohrávajú kľúčovú úlohu vo filozofii jazyka. Týkajú sa a) toho, čo A implikoval v danej situácii, b) toho, čo A povedal, a c) vety, ktorú A použil. Nazvime ich, v tomto poradí, a) *konverzačná elipsa*, b) *eliptická výpoveď* a c) *vetná elipsa*. (Neale 2004, 97)

V tomto článku sa nezaobráme konverzačným typom elipsy, ktorý patrí do pragmatických skúmaní. Eliptickú výpoveď vysvetlíme pomocou príkladu. Predstavme si rozhovor:

A: Dajme si nejaké pivo.

B: Všetky fľaše sú prázdne.

Hovorca B zjavne netvrdí, že všetky fľaše na svete sú prázdne. Zároveň vidíme, že obe vyjadrenia z rozhovoru sú gramaticky úplnými vetami. Podľa Neala nemôžeme tvrdiť, že hovorca B vyslovil eliptickú (nekompletnú) vetu, keďže veta

(2) Všetky fľaše sú prázdne

je úplná. Ako teda vysvetlíme, že hovorca B nehovorí o všetkých fľašiach na svete? Neale netvrdí, že veta (2) je v tomto rozhovore eliptická, ale to, že hovorca B sa elipticky vyjadril. Informácia, že hovorca B hovorí iba o určitých fľašiach, podľa Neala nie je obsiahnutá vo význame vety (2).

Prejdime k eliptickým vetám.² V úvodnom príklade hovorca B použil vetu:

(3) Peter sedel s Máriou a Jožo s Anetou.

Povrchová forma³ tejto vety nie je úplná. Podľa rôznych syntaktických teórií, napríklad z oblasti generatívnych gramatík, vznikla syntaktickou transformáciou z pôvodnej úplnej vety (1). Na rozdiel od konverzačnej elipsy a eliptickej výpovede sa teda eliptické vety vyznačujú neúplnosťou na úrovni povrchovej formy. V tomto článku sa venujem eliptickým vetám a ich významu.

Najprv sa však vyjadrím k Nealovmu rozlíšeniu troch druhov elipsy. Podľa mňa vychádza z nerešpektovania určitých základných rozlíšení. Pavol Cmorej zaviedol rozlíšenie medzi jazykovým výrazom a semivýrazom (pozri Cmorej 2005). Identitu jazykového výrazu určuje usporiadaná dvojica ⟨semivýraz, význam⟩. Semivýraz je grafický znak zvyčajne zložený z písmen (príp. hlások). Každý jazykový výraz má význam nevyhnutne. Semivýraz môže mať význam, no nemá ho nevyhnutne. Cmorejovo kritérium pritom funguje bez ohľadu na prijatú sémantiku. Vďaka rozlíšeniu medzi jazykovým výrazom a semivýrazom napríklad jednoducho uchopíme rozdiel medzi výrazmi „koruna“ (keď hovoríme o peniazoch) a „koruna“ (keď hovoríme o stromoch). Ide o dva jazykové výrazy, ktoré sú síce nerozlíšiteľné z fonetického a grafického hľadiska, no napriek tomu to nie je jeden a ten istý jazykový výraz, pretože majú odlišné významy. Zouhar navrhuje doplnenie tohto kritéria identity jazykových výrazov (pozri Zouhar 2011, 144–150): Dve vety, ktoré sú foneticky a graficky rovnaké, sa napriek tomu môžu líšiť v syntaktickej stavbe semivýrazu. Preto navrhuje doplnenie kritéria o identifikáciu štruktúry semivýrazu.

Neale zjavne nepoužíva takéto kritérium identity jazykového výrazu. Fakt, že sa v príklade nehovorilo o všetkých fľašiach na svete, vysvetľuje kontextualisticky: „Naopak výpoveď o tom, že (ten jediný) stôl je veľký,⁴ nie je prípadom vetnej elipsy. Hovorcovi B napomáha samozrejmy mimoja-

² Mali by sme rozlišovať eliptické vety a vetnú elipsu. Eliptické vety sú vety, ktoré obsahujú vynechanie. Vetná elipsa vzniká vynechaním celej vety v rámci inej vety. Napríklad: „Peter vie, že $2 + 2 = 4$, a Jano tiež.“

³ Pre rýchly a zrozumiteľný úvod do spôsobu reprezentácie lingvistických štruktúr odporúčam kapitolu *Lingvistické minimum* z monografie Zouhar (2009).

⁴ Neale používa príklad *The table is large*. Parafráza pre náš príklad: V prípade výpovede *Všetky fľaše sú prázdne* sa nevyskytuje eliptická veta. Hovorcovi B napomáha samozrejmy mimojazykový fakt, že vo svete je viac fliaš ako tie, o ktorých hovorí.

zykový fakt: Svet obsahuje viac ako jeden stôl“ (Neale 2004, 156). Prijatie kontextualistických pozícií nám však sťažil Zouhar vo svojej knihe *Význam v kontexte* (pozri Zouhar 2011, 311-332). Namiesto toho využime kritérium identity jazykového výrazu v prípade eliptickej výpovede. Veta (2), ktorá sa objavila v uvedenom rozhovore, nemusí byť úplnou vetou. Inými slovami, môžeme rozlíšiť dva zložené výrazy, ktoré sa graficky a foneticky zhodujú s vetou (2). Veta (2) mohla vzniknúť minimálne dvomi spôsobmi. Mohla vzniknúť aj eliptickým vynechaním vedľajšej vety z vety (4):

(4) Všetky fľaše, ktoré sme doniesli, sú prázdne.

Ak vynecháme vedľajšiu vetu, vznikne eliptická veta, ktorá je graficky a foneticky nerozlišiteľná od vety (2). Rozdiel medzi dvomi vetami, ktoré sú na povrchovej úrovni zhodné s vetou (2), uchopíme pomocou prislúchajúcich syntaktických stromov. Tieto vety vieme rozlíšiť pomocou navrhnutého kritéria identity jazykových výrazov. Nealovo rozlíšenie eliptickej výpovede od eliptických viet však predpokladá, že veta (2) nevznikla z vety (4). Ak na druhej strane prijmem toto rozlíšenie, vysvetlíme pôvod zdieľaného významu z použitia eliptickej vety, ktorá vznikla z vety (4). Aj Neale prijíma rozdiel medzi fonetickou rovinou a logickou rovinou vety. Vďaka tomu existuje možnosť, že sa na význame nejakej vety podieľa aj význam takých syntaktických prvkov vety, ktoré sa neobjavili v jej fonetickej forme. Zouhar o afonetických prvkoch hovorí ako o implicitných syntaktických prvkoch vety. Ak teda prijmem navrhované kritérium, vysvetlíme eliptickú výpoveď ako prípad eliptickej vety a nemusíme sa opierať o kontextualistické vysvetlenie.⁵

2. Syntaktické teórie elipsy

May vo svojej stati o elipse definuje eliptickú vetu ako jazykový fenomén:

Vo všeobecnosti je jazyková elipsa skrátená alebo neúplná jazyková forma. Konštituenty, ktoré sa bežne vyskytujú vo vete, v tejto jazykovej forme na jej povrchovej úrovni absentujú, čo je umožnené štruktúrnou prítomnosťou predchádzajúceho antecedentu. (May 2002, 1094)

⁵ Tento prístup však ešte nevyvracia kontextualizmus. Ide skôr o alternatívu k Nealovmu rozlišovaniu rôznych druhov elipsy. Iba ukazujem, že nemusíme rozlišovať rôzne druhy elipsy, ak rozlíšime rôzne druhy výrazov.

Rôzne syntaktické teórie, ktoré sa pokúšajú o vysvetlenie vzniku skrátených jazykových foriem, zahrniem pod hlavičku syntaktické teórie elipsy.⁶ Antecedentom je taký syntaktický výraz, ktorého pomocou získame vypustenú časť eliptickej vety. Napríklad v prípade vety (3) je antecedentom výraz „sedel“ v prvej časti vety. Práve vďaka nemu vieme, že v druhej časti vety (3) bol vynechaný výraz „sedel“. Scheffler tvrdí: „Vo všeobecnosti predpokladáme, že môžeme vynechať časť výpovede, ak sa nejako ‚podobá‘ explicitne prítomnému antecedentu“ (Scheffler 2013, 3). Ďalší spoločný predpoklad syntaktických teórií elipsy formuluje May: „...už pozorujeme základnú vlastnosť elipsy, ktorú by podľa očakávania malo zachytiť akékoľvek vysvetlenie vzťahu medzi eliptickými a neeliptickými formami: význam sa pri použití elipsy nemení“ (May 2002, 1094). Zásadným obmedzením týchto teórií je to, že syntaktické transformácie sú prípustné len v rámci vety. Pre tieto teórie bude teda problémom, ak sa antecedent nachádza mimo eliptickej vety.

Syntaktické teórie elipsy predpokladajú, že na nejakej úrovni syntaktickej analýzy je eliptická veta totožná s neeliptickou vetou. Podľa teórií vymazania vznikol rozdiel medzi vetami (1) a (3) vymazaním výrazu „sedel“ pri prechode z hĺbkovej k povrchovej štruktúre. Pri analýze hĺbkovej syntaktickej štruktúry vety (3) by sme objavili vymazané sloveso „sedel“. Pre tieto teórie je predpokladom vymazania nejakého výrazu z vety splnenie tzv. podmienky návratnosti. Ak teda niečo vymažeme, musíme mať možnosť to do vety vrátiť. Nebudeme sa hlbšie venovať podmienke návratnosti, iba spomeniem to, že ide o syntaktickú podmienku, napríklad to, že antecedent je tzv. hlavou frázy,⁷ v ktorej sme niečo vynechali. Radšej upozorním na obmedzenia teórií vymazania. Pozrime sa na nasledujúci semivýraz:

(5) Martin videl svoju matku a aj Edo.

Podľa jedného čítania každý z týchto pánov videl svoju matku: Martin videl Martinovu matku a Edo videl Edovu matku. Podľa druhého možného čítania Martin videl Martinovu matku a aj Edo videl Martinovu matku. Podľa teórie vymazania je antecedent syntakticky totožný s vynechanou časťou eliptickej vety, a preto táto teória uchopí iba čítanie, podľa ktorého každý z pánov videl svoju matku.

⁶ Prehľad syntaktických teórií elipsy ako aj ich obmedzení pozri v May (2002), resp. Scheffler (2013).

⁷ Termín „hlava frázy“ je odborný termín v rámci Government and Binding theory. Pozri napríklad Haegeman (1994).

Prejdime k druhému príkladu (May 2002, 1099) a pozrime sa na vetu

(6) Dr. Watson podozrieval všetkých, ktorých Sherlock Holmes.

Doteraz sme predpokladali, že antecedent a vynechané miesto sú syntakticky nezávislé.⁸ Vynechaná časť nebola súčasťou antecedentu. Vo vete (6) sa však vynechaná časť nachádza v podrad'ovacom súvetí. V syntaktickom stromovom diagrame je podrad'ovacie súvetie syntakticky podradené slovu „podozrieval“. Ak kopírujeme antecedent, mali by sme kopírovať aj jeho syntaktické vlastnosti. Ak mu teda prislúcha nejaká podradená veta, mali by sme kopírovať aj ju. Podľa teórií, ktoré predpokladajú syntaktickú totožnosť antecedentu a vynechaného miesta, by veta (6) mala mať ten istý význam ako výraz:

(7) Dr. Watson podozrieval všetkých, ktorých Sherlock Holmes podozrieval všetkých, ktorých Sherlock Holmes.

Výraz (7) vznikol doplnením vynechanej časti vo vete (6) pomocou kópie antecedentu s podrad'ovacou vetou „podozrieval všetkých, ktorých Sherlock Holmes“. Vidíme, že ak antecedent a vynechané miesto nie sú syntakticky nezávislé, tak sa pomocou predpokladu ich syntaktickej totožnosti vynechaného miesta nezbavíme.

Predpoklad syntaktickej totožnosti antecedentu a vynechaného miesta v eliptickej vete spôsobuje problémy. Ak ho však odmietneme, ako vysvetlíme, že vynechanému miestu v eliptickej vete rozumíme vďaka antecedentu? Čo keby sme tvrdili, že antecedent a vynechané miesto v eliptickej vete nie sú syntakticky totožné, ale vzťahujú sa na tú istú vlastnosť? Dôsledkom by bolo to, že vzťah medzi antecedentom a vynechanou časťou eliptickej vety nebude daný syntaktickým, ale sémantickým kritériom. Tým sa dostávame do oblasti sémantických teórií elipsy.

3. Sémantické teórie elipsy

3.1. Teória vlastnosti (*Property theory*)

Vráťme sa k príkladu semivýrazu (5). Syntaktické teórie elipsy neuchopili obe čítania tejto vety. Nerozlišovali totiž medzi dvomi rôznymi vlast-

⁸ V rámci hĺbkovej syntactickej analýzy viet vystupujú ako rovnocenné syntaktické členy. Jeden nie je nadradený druhému.

nostíami, ktoré môžu byť významom antecedentu. Tento rozdiel lepšie zobrazíme pomocou jazyka λ -kalkulu:⁹

- a) «*vidieť svoju matku*»:¹⁰ $\lambda x(x \text{ videl matku } x)$
 b) «*vidieť Martinovu matku*»: $\lambda x(x \text{ videl Martinovu matku})$

Rozdiel medzi oboma vlastnosťami spočíva v počte výskytov viazaných premenných v λ -výraze. Vidíme, že λ -výraz pre vlastnosť a) obsahuje dva viazané výskyt x a výraz pre vlastnosť b) obsahuje iba jeden viazaný výskyt x . Pri analýze výskytu vlastností a) a b) vo významoch priradených semivýrazu (5) sa uplatní pravidlo β -redukcie, ktoré uchopuje aplikáciu funkcie na argument. Symbol \gg predstavuje krok redukcie:

$$\lambda x(Fx)(a) \gg Fa$$

Pozrime sa teraz na jeho použitie v λ -zápisoch¹¹ zodpovedajúcich dvom významom semivýrazu (5):

- (5a) $\lambda x(x \text{ videl matku } x)(\text{Martin}) \ \& \ \lambda x(x \text{ videl matku } x)(\text{Edo}) \gg$
 Martin videl Martinovu matku & Edo videl Edovu matku
 (5b) $\lambda x(x \text{ videl Martinovu matku})(\text{Martin}) \ \& \ \lambda x(x \text{ videl Martinovu matku})(\text{Edo}) \gg$
 Martin videl Martinovu matku & Edo videl Martinovu matku

Po redukcii pomocou argumentov «*Martin*» a «*Edo*» sa λ -výrazy zodpovedajúce rôznym vlastnostiam nezhodujú. Vďaka rozlíšeniu dvoch vlastností rozlíšime dva rôzne významy semivýrazu (5).

Zástancovia teórie vlastnosti (ďalej len TV)¹² stále prijímajú podmienku totožnosti významu eliptickej a neeliptickej vety. Túto významovú totožnosť vysvetľujú pomocou sémantickej totožnosti významu antecedentu a významu vynechanej časti eliptickej vety. TV nepredpokladá syntaktické vzťahy alebo transformácie medzi antecedentom a vynechanou časťou eliptickej vety. Preto sa na ňu nevzťahuje obmedzenie, ktoré vymedzuje použi-

⁹ Pre bližšie oboznámenie s teóriou λ -kalkulu pozri napríklad Zlatuška (1993).

¹⁰ Ak hovorím o významoch výrazov, používam francúzske úvodzovky: «*a*» odkazuje na význam výrazu „*a*“.

¹¹ Tieto formulácie iba používajú λ -zápisy, nie sú napísané v jazyku λ -kalkulu.

¹² Pozri napr. Sag (1976) či Williams (1977).

tie syntaktických transformácií. Antecedent sa teda už nemusí nachádzať priamo v eliptickej vete.

Identickým λ -zápisom zodpovedá tá istá vlastnosť. Dva λ -zápisy sú podľa TV identické, ak sa zhodujú v štruktúre. Môžu sa síce líšiť písmená predstavujúce viazanú premennú, no vďaka β -redukcii je tento rozdiel nepodstatný. Ak λ -zápisy navyše obsahujú λ -voľnú premennú, sú takéto výrazy identické len vtedy, keď sú výskyty danej λ -voľnej premennej viazané tým istým logickým operátorom, napríklad kvantifikátorom. Všimnime si funkciu tohto kritéria identity na konkrétnom prípade:

- (8) Martin videl svoju matku a Edo verí, že aj Jana.

Semivýrazu (8) prislúcha viacero rôznych významov. Napríklad:

- (8a) «Martin videl Martinovu matku a Edo verí, že Jana videla Martinovu matku.»
 (8b) «Martin videl svoju matku a Edo verí, že Jana videla svoju matku.»
 (8c) «Martin videl svoju matku a Edo verí, že Jana videla Edovu matku.»

Pozrime sa teraz na vlastnosti prislúchajúce antecedentu a vynechaným častiam eliptických viet v týchto príkladoch:

	Antecedent	Vynechaná časť
(8a)	$\lambda x(x \text{ videl Martinovu matku})$	$\lambda x(x \text{ videl Martinovu matku})$
(8b)	$\lambda x(x \text{ videl matku } x)$	$\lambda x(x \text{ videl matku } x)$
(8c)	$\lambda x(x \text{ videl matku } x)$	$\lambda x(x \text{ videl matku } z)$

V prípadoch (8a) a (8b) je podmienka identity antecedentu a vypustenej časti eliptickej vety splnená. V prípade (8c) sa však štruktúra λ -výrazov prislúchajúcich jednotlivým vlastnostiam nezhoduje. λ -zápis zodpovedajúci významu antecedentu obsahuje dva viazané výskyty premennej x . λ -zápis pre vynechanú časť však obsahuje iba jeden viazaný výskyt tejto premennej. Okrem toho obsahuje výskyt premennej z , ktorý je viazaný v rámci širšieho λ -zápisu pre celú druhú vetu súvetia (8c). Podľa kritéria identity z TV teda antecedent a vynechaná časť pre čítanie (8c) nie sú identické. Z hľadiska tejto teórie preto veta (8) nepredstavuje eliptickú podobu neeliptickej vety pre čítanie (8c). To je však zjavne proti jazykovým intuíciam, podľa ktorých môžeme vete (8) takto rozumieť. TV vysvetľuje elipsu na základe predpokladu o sémantickej identite antecedentu a vynechanej časti eliptickej vety. Vidíme, že takto neuchopí TV čítanie (8c), a teda neposkytuje dostatočne všeobecné vysvetlenia.

3.2. Teória elipsy podľa minimálneho indexikalizmu

Doteraz preskúmané teórie analyzovali elipsu tak, že predpokladali syntaktické, resp. sémantické vzťahy medzi antecedentom a vynechanou časťou eliptickej vety. Na príkladoch sme videli, že boli úspešné iba do istej miery. Dôvodom postulovania týchto vzťahov bol predpoklad totožnosti sémantických obsahov eliptickej a prislúchajúcej neeliptickej vety. Čo keby sme však tento predpoklad odmietli? Ako potom vysvetlíme, že rozumieme, ktoré výrazy boli vypustené z eliptickej vety, resp. rozumieme významu eliptickej vety? A aký bude vzťah významov antecedentu a vynechanej časti eliptickej vety?

Teória minimálneho indexikalizmu (ďalej len TMI) má ambíciu vysvetliť širšie množstvo javov, medzi ktoré patrí aj elipsa. Pri skúmaní eliptických viet táto teória v princípe pripúšťa, že význam vypustenej časti vety získame vďaka sémantickému obsahu antecedentu. Na druhej strane však odmieta predpoklad totožnosti významu eliptickej a jej prislúchajúcej neeliptickej vety. Čo je to TMI a prečo neprijíma tento predpoklad? V článku sa obmedzím na tie časti teórie, ktoré potrebujeme v súvislosti s teóriou elipsy.

Marián Zouhar predstavuje podstatu minimálneho indexikalizmu pomocou tejto tézy:

Všetky vplyvy kontextu, ktoré sa (vzhľadom na daný kontext použitia) podieľajú na determinácii vyjadreného sémantického obsahu (propozície) výrazu, ktorý hovorca použil na uskutočnenie svojej výpovede, musia byť sémantickým obsahom lexikálnych či gramatických jednotiek vyskytujúcich sa v danom výraze; ak nejaká zložka vyjadreného sémantického obsahu nemôže byť obsahom explicitných lexikálnych, resp. gramatických jednotiek vety, musí byť hodnotou implicitnej premennej – v takom prípade hovorca vo svojej výpovedi použil vetný radikál, v ktorom sa táto premenná vyskytuje na úrovni logickej formy, nie kompletnú vetu. (Zouhar 2011, 160)

Pozrime sa na niektoré základné termíny, ktoré sa vyskytli v citovanej téze. Začnime termínom „vyjadrený sémantický obsah“. Zouhar rozlišuje tri druhy významu (pozri Zouhar 2011, 27-66), ktoré môže mať použitie vety vzhľadom na určitý kontext. Prvým druhom je doslovný význam: „Doslovný význam vety je daný sémantickými konvenciami pre jednoduché výrazy a jej syntaktickou stavbou“ (Zouhar 2011, 47). Druhou rovínou významu vety je vyjadrená propozícia: „Ide o rovínu toho, čo veta vyjadruje vzhľadom na určitý kontext použitia“, propozíciu, „ktorú veta vyjadruje vzhľadom na kontext použitia“ (Zouhar 2011, 62-63). Tretia rovina významu je tzv. impliko-

vaný význam: „je tým, čo hovorca mieni komunikovať použitím vety v určitom kontexte“ (Zouhar 2011, 62). Rozdiel medzi doslovným významom vety a jej vyjadrenou propozíciou vznikne, ak veta obsahuje kontextovo citlivé výrazy (napríklad indexické výrazy ako „ten“). Ak neobsahuje žiadnu kontextovo citlivú lexikálnu či gramatickú jednotku, tak sa jej doslovný význam zhoduje s vyjadrenou propozíciou. V záujme korektnosti upozorním na to, že termín „propozícia“ používa Zouhar aj s niektorými problematickými vlastnosťami pojmu, na ktorý sa vzťahuje.¹³ Sám však upozorňuje na to, že vo formuláciách jeho stanovísk môžeme nahradiť tento termín slovným spojením „propozičná konštrukcia“. Termín „konštrukcia“ má zrejme podobné použitie ako v Transparentnej intenzionálnej logike (ďalej len TIL). Nateraz nám postačí porozumieť rozdielu medzi propozíciou a propozičnou konštrukciou. Propozícia je nositeľom pravdivostnej hodnoty a nie je štruktúrovaná. Propozičná konštrukcia je štruktúrovaná entita. Vyjadrenej propozícii teda môžeme rozumieť ako štruktúrovanému objektu.

Vyjadrenú propozíciu ako návod na určenie pravdivostnej hodnoty vety získame, ak priradíme sémantický obsah jednotlivým lexikálnym či syntaktickým zložkám výrazu použitého v hovorcovej výpovedi. Ak obsahuje kontextovo citlivé prvky, vyjadrenú propozíciu získame až po získaní ich sémantického obsahu. Zouhar však má vlastné kritérium identity jazykového výrazu. V skratke, kritérium identity je dané usporiadanou trojicou ⟨semivýraz, štruktúra semivýrazu, význam⟩. Zouhar navyše rozlišuje medzi explicitnými a implicitnými syntaktickými prvkami výrazu. Identitu výrazu určíme až vtedy, keď určíme všetky prvky jeho semivýrazu bez ohľadu na to, či sú alebo nie sú explicitné.

Význam vynechanej časti eliptickej vety sa podieľa na vyjadrenej propozičnej konštrukcii. Na druhej strane však tento význam nie je sémantickým obsahom žiadnej explicitnej lexikálnej či syntactickej zložky vyjadreného výrazu. Podľa tézy minimálneho indexikalizmu teda sémantický obsah vynechanej časti eliptickej vety musí byť hodnotou implicitnej premennej.

Eliptická veta podľa TMI obsahuje implicitné syntaktické jednotky. To však má závažný dôsledok pre eliptickú vetu – nie je totiž kompletnou vetou, ale iba tzv. „vetným radikálom“:

¹³ Myslím tým napríklad problematickú otázku štruktúrovanosti propozície. Zouhar otvorene hovorí o problematickom pojme propozície, ktorý používa. Pozri napr. Zouhar (2011, 13–15).

Na označenie nekompletných viet, ktorých pomocou možno vyjadriť kompletne propozície (vzhľadom na daný kontext použitia fixujúci hodnoty premenných), budem používať termín *vetný radikál*... Vetný radikál možno z vety dostať tak, že niektorý výraz, resp. niektoré výrazy sa nahradia premennými vhodných druhov (Zouhar 2011, 158).

Na jednoduchom príklade si ukážeme, ako TMI vysvetľuje elipsu. Vráťme sa k výrazu (3) (t. j. „Peter sedel s Máriou a Jožo s Anetou“). Na to, aby sme podľa TMI určili vyjadrenú propozíciu prislúchajúcu tomuto výrazu, musíme priradiť sémantický obsah všetkým syntaktickým jednotkám, ktoré sú v ňom obsiahnuté. Vo výraze (3) však zjavne nie sú všetky plnovýznamové syntaktické zložky explicitné. Implicitnú premennú zodpovedajúcu vypustenému výrazu zviditeľníme napríklad pomocou zápisu

(3I) Peter sedel s Máriou a Jožo R^2 s Anetou.

Výraz R^2 predstavuje zviditeľnenú implicitnú premennú. Index ² znamená, že v rámci vetného radikálu (3I) ide o premennú, ktorej oborom sú binárne relácie. Vetný radikál (3) však zjavne vyjadruje propozíciu (vzhľadom na kontextovo podmienené priradenie hodnoty premennej, hoci sám osebe nevyjadruje kompletnú propozíciu). Implicitná premenná R^2 v rámci tohto vetného radikálu teda vystupuje ako ohodnotená. Ak totiž rozumieme propozícii vyjadrenej týmto radikálom, tak podľa TMI má každá (implicitná alebo explicitná) syntaktická jednotka priradený sémantický obsah. Premenná nadobúda hodnoty vďaka ohodnoteniu. Ohodnotenie však musíme vykonať. Nie je to samozrejماً operácia na premennej. Pozrime sa na nasledujúci príklad:

(9) Jožo s Anetou.

(9I) Jožo R^2 s Anetou.

Výraz (9), resp. (9I) zjavne nevyjadruje úplný návod na priradenie pravdivostnej hodnoty. Nepoznáme totiž hodnotu premennej R^2 . Táto premenná ešte nenadobudla hodnotu na základe ohodnotenia nezávislého od kontextu jej použitia. Ak na rozdiel od toho rozumieme výrazu (3), resp. (3I) tak, že vyjadruje propozičnú konštrukciu, tak kontext umožnil ohodnotenie tejto premennej. Jediný relevantný vplyv kontextu v tomto prípade predstavuje výskyt výrazu „sedel“ v prvej časti výrazu (3).

Ohodnotenie premennej nie je syntaktický vzťah. V prípade elipsy sú hodnoty implicitnej premennej podľa TMI aj obsahom niektorého výrazu z predchádzajúcej časti textu či komunikácie. Tento vzťah sémantických

obsahov antecedentu a implicitnej premennej reprezentujúcej vynechanú časť eliptickej vety sa nápadne podobá kritériu sémantickej identity sémantických obsahov v TV. Je tu však zásadný rozdiel. V TV boli výskyty premenných neoddeliteľnou súčasťou λ -zápisov, ktoré zodpovedali sémantickým obsahom antecedentu a vynechanej časti eliptickej vety. TMI nepostuluje túto spätosť a uvažuje binárne relácie a individuové premenné ako sémanticky nezávislé objekty.

Pozrime sa teraz na to, ako si TMI poradí s príkladmi, ktoré boli problematické pre predchádzajúce teórie. Zjavne pomocou tejto teórie uchopíme obe možné čítania semivýrazu (5) (t. j. „Martin videl svoju matku a aj Edo“). Vieme, že vlastnosť «*vidieť*» je binárna. Semivýraz (5) preto okrem implicitnej premennej prislúchajúcej tejto binárnej relácii podľa TMI obsahuje aj implicitnú individuovú premennú. Zobrazíme to nasledovne:

(5I) Martin videl svoju matku a aj Edo R^2 i .

R^2 predstavuje implicitnú premennú pre binárne relácie a i implicitnú individuovú premennú. Dve možné čítania semivýrazu (5) teda uchopíme pomocou nasledujúcich dvoch rôznych ohodnotení implicitných premenných:

(5Ia) Martin videl svoju matku a aj Edo R^2 i .

Ohodnotenie a: $R^2 = \langle \text{videl matku} \rangle$; $i = \langle \text{Edo} \rangle$.

(5Ib) Martin videl svoju matku a aj Edo R^2 i .

Ohodnotenie b: $R^2 = \langle \text{videl matku} \rangle$; $i = \langle \text{Martin} \rangle$.

Sémantický obsah implicitných premenných sme určili vzhľadom na predchádzajúci jazykový kontext. Pomocou týchto ohodnotení zároveň uchopíme vyjadrené propozičné konštrukcie, ktoré predstavujú dve rôzne čítania semivýrazu (5).

Prejdime k druhému problematickému príkladu (6) (t. j. „Dr. Watson podozrieval všetkých, ktorých Sherlock Holmes“). Opäť zviditeľníme implicitnú premennú:

(6I) Dr. Watson podozrieval všetkých, ktorých R^2 Sherlock Holmes.

Hodnotu implicitnej relácie R^2 získame z predchádzajúceho jazykového kontextu. V (6I) predstavuje túto hodnotu binárna relácia «*podozrieval*». Pri takomto ohodnotení sa vyhneme nekonečnému regresu, ktorý predstavoval problém pre syntaktické teórie. TMI rieši problémy, na ktoré sme narazili pri syntaktických teóriách.

Príklad (8) (t. j. „Martin videl svoju matku a Edo verí, že aj Jana“) predstavoval problém pre sémantickú TV. Okrem implicitnej premennej zodpovedajúcej binárnej relácii obsahuje semivýraz (8) aj implicitnú individuovú premennú. Po zviditeľnení implicitných premenných má semivýraz (8) nasledujúcu podobu:

(8I) Martin videl svoju matku a Edo verí, že aj Jana R^2 i .

Prejdime si teraz tri čítania tohto semivýrazu spolu s ohodnoteniami implicitných premenných:

(8Ia) Martin videl svoju matku a Edo verí, že aj Jana R^2 i .
Ohodnotenie: $R^2 = \langle \text{videla matku} \rangle$, $i = \langle \text{Martin} \rangle$

(8Ib) Martin videl svoju matku a Edo verí, že aj Jana R^2 i .
Ohodnotenie: $R^2 = \langle \text{videla matku} \rangle$, $i = \langle \text{Jana} \rangle$

(8Ic) Martin videl svoju matku a Edo verí, že aj Jana R^2 i .
Ohodnotenie: $R^2 = \langle \text{videla matku} \rangle$, $i = \langle \text{Edo} \rangle$

TMI dovoľí uchopiť všetky tri deklarované čítania, a to vďaka sémantickému obsahu adekvátnych antecedentov a vynechaných častí eliptických viet. Rieši tak problémy, ktoré neriešila TV.

TMI vysvetľuje, ako eliptické vety nadobúdajú sémantický obsah predstavujúci úplnú propozičnú konštrukciu. Na otázky zo začiatku tejto kapitoly odpovedá TMI nasledovne:

Q: Ako rozumieme, aké časti boli vypustené z eliptickej vety?

A: Vieme to vďaka kontextu jej použitia, ktorého vplyv sa prejaví ohodnotením príslušných implicitných premenných.

Q: Aký je vzťah významov antecedentu a vynechanej časti eliptickej vety?

A: Hodnoty implicitných premenných sú zároveň sémantickým obsahom niektorej predchádzajúcej časti textu alebo komunikácie.

Povedzme si ešte, prečo TMI neprijíma postulát totožnosti sémantických obsahov eliptickej a jej prislúchajúcej neeliptickej vety. Eliptická veta podľa TMI obsahuje implicitné premenné a predstavuje vetný radikál. Neeliptická veta žiadne implicitné premenné neobsahuje. Hoci v konkrétnom použití vetného radikálu je jeho vyjadrený význam rovnaký ako význam prislúchajúcej neeliptickej vety, neplatí to vo všeobecnosti. Ten istý vetný radikál totiž môžeme použiť v inom kontexte, v ktorom bude vyjadrovať iný význam. To súhlasí s tvrdením TMI, že eliptická veta mimo kontext jej

použitia predstavuje otvorenú propozičnú konštrukciu, ktorá samostatne nie je úplným návodom na priradenie pravdivostnej hodnoty nejakej propozícii.

3.3. *Elipsa ako prípad anaforickej väzby v rámci TIL*

Teraz ponúknem alternatívne¹⁴ riešenie sémantickej analýzy významu vynechaných častí eliptických viet v rámci TIL.¹⁵ Tvrdím, že vzťah medzi významom antecedentu a vynechanej časti eliptickej vety sa zakladá na rovnakých princípoch ako vzťah významu anaforicke¹⁶ použitých výrazov a ich antecedentu. Svoje tvrdenie zakladám na jazykovom pozorovaní, podľa ktorého rozumieme významu vynechaných častí eliptickej vety vďaka predchádzajúcemu jazykovému kontextu. Anaforické výrazy rovnako nadobúdajú svoj význam vďaka väzbe na predchádzajúci jazykový kontext. Toto pozorovanie viedlo k pracovnej hypotéze, podľa ktorej môžeme analyzovať význam eliptických viet rovnako ako význam viet s anaforickým obsahom: Vysvetľujeme ho pomocou substitúcie. Aké sú teda princípy, na ktorých stojí táto sémantická analýza eliptických viet?

Naša analýza elipsy je dvojstupňová. Najprv nahradíme eliptickú vetu anaforickou vetou. Tú ďalej nahradíme vetou obsahujúcou prístavok, ktorá čo najvernejšie zachytáva význam vety obsahujúcej anaforický výraz. Sémantickú analýzu aplikujeme na vetu s prístavkom. Naša analýza viet obsahujúcich anaforický výraz totiž predpokladá formuláciu prístavkových viet, ktorých sémantický obsah potom skúmame. Z tohto pohľadu je rozdiel medzi eliptickými vetami a vetami obsahujúcimi anaforický výraz iba v tom, že pri analýze eliptických viet najprv nájdeme im prislúchajúce anaforické vety. Elipsa je teda podľa našej analýzy v podstate komplikovanejšia forma anafory. Jednoduchšie to ukážeme pomocou príkladov.

Príkladom anaforického výrazu je výraz „On“ v nasledujúcom kontexte:

(10) Pavel je logik. On je rozumný.

Rozumieme, že výraz „On“ sa vzťahuje na Pavla. Zároveň je to iný spôsob odkazu na Pavla ako výraz „Pavel“. Nesmieme si však zamieňať druhú vetu v kontexte (10) za vetu (11).

¹⁴ Tento návrh je alternatívny k analýze elipsy v rámci TIL predstavenej v článku Duží – Jespersen (2013).

¹⁵ Učebnicu TIL predstavuje Duží – Materna (2012).

¹⁶ Diskusia k anafore sa viedla vo svete aj v našich končinách. Pozri Gahér (2002), Zouhar (2004), Duží (2006).

(11) On je rozumný.

Zásadné rozdiely medzi týmito vetami spočívajú v ich sémantickom obsahu.¹⁷ Tento rozdiel ukážeme, keď druhú vetu v kontexte (10) obsahujúcu anaforický výraz nahradíme vetou

(12) On, teda Pavel, je rozumný.

Zároveň predpokladáme, že veta (12) neobsahuje vetu (11) ako samostatný podvýraz. Prístavok „teda Pavel“, ktorý sa dopĺňa k zámenu „On“, nahrádza vo vete (12) (resp., presnejšie, vo vete, ktorú dostaneme z (10), keď druhú vetu nahradíme vetou (12)) anaforickú väzbu, ktorá v (10) existuje medzi „On“ a „Pavel“. Ak budeme dôslední, musíme pripustiť, že kým druhá veta v (10) obsahovala anaforický výraz, tak veta (12) už anaforický výraz neobsahuje. Anafora sa totiž nahradila prístavkom.

Sémantická analýza prvej vety v príklade (10) vyzerá v TIL nasledovne:

(10T) $[\lambda w \lambda t [{}^0 \text{Logik}_{wt} {}^0 \text{Pavel}]]$

Ako táto analýza zachytáva sémantický obsah vety?¹⁸ Výraz „Pavel“ referuje na konkrétne individuum. Konštrukcia ${}^0 \text{Pavel}$, ktorá je významom tohto výrazu, konštruuje individuum Pavel. Významom výrazu „logik“ je konštrukcia vlastnosti, v tomto prípade trivializácia ${}^0 \text{Logik}$. Extenzionalizáciou vlastnosti «*logik*», teda jej aplikáciou na svet w a čas t (${}^0 \text{Logik}_{wt}$), konštruujeme populáciu logikov v tomto svete a čase. Uchopíme ju pomocou funkcie, ktorá priradí logikom pravdivostnú hodnotu Pravda a nelogikom Nepravda. Trivializácia objektu X , symbolicky ${}^0 X$, vracia objekt X bezo zmeny. Konštrukcia ${}^0 \text{Pavel}$ v analýze (10T) identifikuje objekt (individuum) Pavel a vráti ho bezo zmeny,

¹⁷ Musím upozorniť na to, že táto analýza syntaktického a sémantického obsahu druhej vety v kontexte (10) sa nezhoduje so syntaktickou a sémantickou analýzou tohto príkladu od Marie Duží. Podľa syntaktickej analýzy druhej vety kontextu (10), ktorú predpokladá moja analýza elipsy pomocou aparátu TIL, neobsahuje táto veta ako svoj samostatný podvýraz vetu (11). V tom tkvie zásadný rozdiel medzi týmto návrhom sémantickej analýzy anaforického výrazu a pôvodným všeobecným návrhom od M. Duží (pozri napr. Duží 2006). Tento rozdiel vo výsledkoch syntaktickej analýzy sa odráža aj v navrhovaných sémantických analýzach, pretože dôsledne uplatňujeme tzv. Parmenidov princíp (pozri napr. Duží – Materna 2012, 49-50).

¹⁸ Pri analýzach v TIL sa opierame o termín „konštrukcia“. Exaktnú definíciu tohto termínu pozri napríklad v Duží – Materna (2012, 46).

t. j. identifikuje ho priamo, bez identifikácie akýchkoľvek iných objektov. Je to teda najjednoduchšia konštrukcia individua Pavel. Oba výrazy pre konštrukcie 0Pavel a ${}^0Logik_{wt}$ sa nachádzajú uprostred hranatých zátvoriek. Hranaté zátvorky reprezentujú aplikáciu, resp. zloženie významov výrazov, ktoré sú v zátvorkách. Zápis $[{}^0Logik_{wt} {}^0Pavel]$ čítame nasledovne: Vezmi funkciu konštruovanú pomocou ${}^0Logik_{wt}$ (je to funkcia, ktorú aplikujeme na individua); vezmi objekt konštruovaný pomocou 0Pavel (táto konštrukcia konštruuje individuum); aplikuj danú funkciu na tento objekt. Ak aplikujeme funkciu na objekt, získame pravdivostnú hodnotu v závislosti od toho, či v danom svete a čase je Pavel logikom. Podľa TIL sú propozície funkcie, ktoré nadobúdajú pravdivostnú hodnotu vzhľadom na svet a čas. Keďže konštrukcia $[{}^0Logik_{wt} {}^0Pavel]$ predstavuje návod na získanie pravdivostnej hodnoty v konkrétnom svete a čase, nekonštruuje ešte propozíciu. Lenže podľa TIL sa empirické vety prirodzeného jazyka vzťahujú na propozície. Na to, aby sme zachytili túto vlastnosť, abstrahujeme od konkrétneho sveta a času. Tento krok sa dosiahne pomocou konštrukcie uzáveru. Všimnime si, že sémantická analýza (10T) vlastne predstavuje takúto abstrakciu od návrhu na získanie pravdivostnej hodnoty v konkrétnom svete a čase: $[\lambda w \lambda t [{}^0Logik_{wt} {}^0Pavel]]$.

Prejdime teraz k sémantickej analýze významu anaforických výrazov.¹⁹ Navrhujem, aby sme význam anaforických výrazov uchopovali pomocou nasledovnej schémy:²⁰

$${}^2[{}^0SUB [{}^0Tr ???] {}^0A {}^0A]$$

Konštrukcia 0A je trivializáciou objektu A a hranaté zátvorky predstavujú aplikáciu funkcie na argument, teda kompozíciu. Táto schéma sémantickej analýzy sa teda skladá z dvoch aplikácií konkrétnych funkcií na argument. Prvá kompozícia $[{}^0Tr ???]$ obsahuje ???. Namiesto otáznikov v konkrétnej analýze umiestnime význam antecedentu. Funkcia Tr nám po aplikácii na akýkoľvek argument vráti ako hodnotu trivializáciu tohto argumentu. Ak ju napríklad aplikujeme na argument $Pavel$, vráti nám ako hodnotu 0Pavel . Pozrime sa teraz na druhú kompozíciu: $[{}^0SUB [{}^0Tr ???] {}^0A {}^0A]$. Opäť predstavuje aplikáciu funkcie na argument. Funkcia SUB pracuje nasledovne: Výsledkom aplikácie funkcie SUB na argumenty A, B, C je nahradenie (inými slovami substitúcia) argumentu B pomocou argumentu A v argumente C . Objekt A v schéme

¹⁹ Pôvodný všeobecný návrh sémantickej analýzy viet obsahujúcich anaforické výrazy pozri v Duží (2006).

²⁰ Podrobnejšie vysvetlenie tejto schémy pozri v Kosterec (2013).

predstavuje premennú vhodného typu. Pomocou tejto schémy teda substituujeme za výskyt premennej význam antecedentu. Následne musíme získanú konštrukciu ešte vykonať, keďže obsahuje dvojité exekúciu.²¹

Použijeme teraz uvedenú schému na sémantickú analýzu druhej vety (12):

$$(12T) [\lambda\omega\lambda t [{}^0\text{Rozumný}_{wt} {}^2[{}^0\text{SUB} [{}^0T {}^0\text{Pavel}] {}^0\text{On} {}^0\text{On}]]]$$

Významom zloženého výrazu „On, teda Pavel“ je konštrukcia ${}^2[{}^0\text{SUB} [{}^0T r {}^0\text{Pavel}] {}^0\text{On} {}^0\text{On}]$. Táto konštrukcia konštruuje individuum Pavel. Je teda ekvivalentná konštrukcii ${}^0\text{Pavel}$, nie je s ňou však totožná. Ani význam výrazu „On, teda Pavel“, ktorým simulujeme²² význam anaforického výrazu „On“, nie je totiž totožný s významom výrazu „Pavel“. Použitie navrhutej schémy analýzy som zvýraznil kurzívou. Význam antecedentu (teda význam výrazu „Pavel“) ${}^0\text{Pavel}$ sa nachádza v schéme na mieste otáznikov. Celý výraz v kurzíve predstavuje postup substitúcie významu výrazu „Pavel“ za premennú *On*, ktorá je významom výrazu „On“ vo vete (12). Schéma hovorí: Vezmi premennú, substituuj za túto premennú význam antecedentu a potom danú konštrukciu vykonaj. Významom antecedentu je konštrukcia individua a konštrukcia ${}^0\text{Rozumný}_{wt}$ konštruuje charakteristickú funkciu pre množinu rozumných individuí. Zvyšný postup uchopený v analýze (12T) je rovnaký ako v analýze (10T).

Po tomto zhustenom výklade sa teraz vráťme k elipse. Na rozdiel od TMI tvrdím, že v rámci eliptických viet boli vynechané anaforické výrazy. Moja teória eliptických viet predpokladá, že dané vety obsahujú vo svojej logickej forme iné výrazy ako podľa TMI. Hoci súhlasím s postulovaním vynechaných častí eliptických viet v ich logickej forme, rozdiel je v charaktere daných výrazov. Podľa TMI sa tam vyskytujú implicitné premenné. Podľa predkladanej analýzy v TIL sú to anaforické výrazy. Navyše zdôrazňujem, že pri mojej analýze najprv vytiahneme štruktúry skryté na úrovni lo-

²¹ Konštrukcia 2X pracuje takto (pri zjednodušenom pohľade): Ak konštrukcia X konštruje konštrukciu Y a konštrukcia Y konštruje objekt Z , tak 2X konštruje Z . 2X nič nekonštruje v prípadoch, ak X nie je konštrukcia alebo ak konštrukcia X nekonštruje konštrukciu, alebo konštruje nevlastnú konštrukciu. Napríklad ak X je 02 , tak 2X nič nekonštruje, pretože objekt konštruovaný konštrukciou X , teda 2 , už nie je konštrukciou a nedá sa teda vykonať. Definíciu dvojitej exekúcie pozri napr. v Duží – Materna (2012, 46).

²² Pomocou termínu „simulácia“ sa pokúšam vyjadriť vzťah významu vety obsahujúcej anaforický výraz a jej prislúchajúcej prístavkovej vety, pomocou ktorej čo najvernejšie zachytávame význam vety s anaforickým obsahom.

gickej formy a analyzujeme výsledok, ktorý takto dosiahneme, teda vetu, kde sú anaforické výrazy explicitne prítomné. V konečnom dôsledku teda obe teórie nepredstavujú alternatívne teórie elipsy, pretože sa týkajú rôznych viet. V zásade predpokladajú inú syntaktickú analýzu skúmaných viet. Na základe prijímaného kritéria identity jazykového výrazu teda ide o mimobežné teórie, pretože skúmajú rozdielne vety. Tieto fakty si musíme uvedomiť pri korektnej kritike týchto teórií.

Prejdime teraz ku konkrétnym prípadom. Pozrime sa najprv na analýzu významov prislúchajúcich semivýrazu (5) (t. j. „Martin videl svoju matku a aj Edo“). V prvom kroku analýzy explicitne uvedieme zamlčané výrazy:

- (5a) Martin videl svoju matku a Edo je v tom istom vzťahu so svojou (matkou).

Výrazy „v tom istom vzťahu“ a „svojou (matkou)“ vo vete (5a) sú anaforické výrazy. Druhý krok našej analýzy predpokladá kompletizáciu vety (5a) pomocou prístavkov:

- (5pa) Martin videl svoju, teda Martinovu, matku a v tom istom vzťahu, teda vidieť, je Edo, s niekým, teda s matkou niekoho, teda Edovou.

Túto vetu následne analyzujeme prostriedkami TIL nasledovne:

- (5pTa) $[\lambda w \lambda t \ [^0 a \ [^0 \text{Videl}_{wt} \ ^0 \text{Martin} \ [^0 \text{Matka}_{wt} \ ^2 [^0 \text{SUB} \ [^0 \text{Tr} \ ^0 \text{Martin}] \ ^0 \text{Svoju} \ ^0 \text{Svoju}]]] \ [^2 [^0 \text{SUB} \ [^0 \text{Tr} \ ^0 \text{Videl}_{wt}] \ ^0 R \ ^0 R] \ ^0 \text{Edo} \ [^2 [^0 \text{SUB} \ [^0 \text{Tr} \ ^0 \text{Matka}_{wt}] \ ^0 S \ ^0 S] \ ^2 [^0 \text{SUB} \ [^0 \text{Tr} \ ^0 \text{Edo}] \ ^0 x \ ^0 x]]]]]$

V tomto zápise sme použili premenné R , S , x . Premenná R zodpovedá výrazu „v tom istom vzťahu ako“. Premenná S zodpovedá výrazu „niekým“ a premenná x zodpovedá výrazu „niekoho“.

Rovnaký postup uplatníme pri druhom čítaní semivýrazu (5):

Prvý krok:

- (5b) Martin videl svoju matku a Edo je v tom istom vzťahu s Martinovou matkou.

Druhý krok:

- (5pb) Martin videl svoju, teda Martinovu, matku, a v tom istom vzťahu, teda vidieť, je Edo, s niekým, teda matkou niekoho, teda Martinovou.

$$(5pTb) \quad [\lambda w \lambda t \ [^0 a \ [^0 Videl_{wt} \ ^0 Martin \ [^0 Matka_{wt} \ ^2 [^0 SUB \ [^0 Tr \ ^0 Martin] \ ^0 Svoju \ ^0 Svoju]]] \ [^2 [^0 SUB [^0 Tr \ ^0 Videl_{wt}] \ ^0 R \ ^0 R] \ ^0 Edo \ [^2 [^0 SUB \ [^0 Tr \ ^0 Matka_{wt}] \ ^0 S \ ^0 S] \ ^2 [^0 SUB \ [^0 Tr \ ^0 Martin] \ ^0 x \ ^0 x]]]]]$$

Premenné R , S , x zodpovedajú rovnakým výrazom ako v predchádzajúcej analýze.

Pri našej analýze teda rozlišujeme a) eliptickú vetu, b) vetu, v ktorej sa vyskytujú anaforické výrazy namiesto elipsy, c) vetu, v ktorej sa anaforické výrazy nahrádzajú prístavkovými zvratmi. V prvom kroku analýzy elipsy nahradíme eliptickú vetu anaforickou vetou, ktorú v druhom kroku nahradíme prístavkovou vetou. Až túto vetu analyzujeme prostriedkami TIL. Elipsa aj anafora, čo sú rozšírené jazykové fenomény, o ktoré sa často opiera komunikácia medzi používateľmi jazyka, sú podľa tejto koncepcie len implicitné prípady prístavku. TIL sa dá aplikovať až na vety, ktoré sú kompletne, a teda aj v našom postupe musíme najprv nájsť kompletnú formuláciu, v ktorej sú explicitne uvedené všetky implicitné zložky, a až takýto výsledok môžeme analyzovať prostriedkami TIL.

Uznávam, že ponúkané prepisy v jazyku TIL nie sú najprehľadnejšie.²³ Ich zdanlivá zložitost' však pramení z jemnosti nástrojov sémantickej analýzy systému TIL. So všetkými operátormi použitými v prepisoch (5pTa) a (5pTb) sme sa už stretli. Dĺžku prepisov spôsobuje fakt, že analyzovanými vetami sú súvetia. Pre nás stačí, ak si všimneme analýzu významu viet obsahujúcich prístavky v týchto súvetiach. V uvedených prepisoch sú zvýraznené a líšia sa iba na jednom jedinom mieste, pri poslednom použití Trivializácie.

Vynechanie sa podľa navrhovanej analýzy v eliptickej vete (5) vykonalo trikrát. Výraz, ktorý je z eliptickej vety vynechaný, doplníme pomocou anaforického výrazu. Význam anaforického výrazu následne simulujeme pomocou výrazu s prístavkom, ktorého význam zachytávame pomocou schémy $^2 [^0 SUB \ [^0 Tr \ ???] \ ^0 A \ ^0 A]$. Všimnime si teraz časti analýz, ktoré simulujú význam anaforického výrazu. Navrhnutú schému používame trikrát, raz pre každý anaforický výraz: Konštrukcia $^2 [^0 SUB \ [^0 Tr \ ^0 Videl_{wt}] \ ^0 R \ ^0 R]$ zachytáva substitúciu významu antecedentu „videl“ za premennú pre binárne relácie R , ktorá sa vzťahuje na výraz „v tom istom vzťahu ako“; konštrukcia $^2 [^0 SUB \ [^0 Tr \ ^0 Matka_{wt}] \ ^0 S \ ^0 S]$ zachytáva substitúciu významu antecedentu „matka“ za premennú S , ktorá zodpovedá výrazu „niekým“; konštrukcia $^2 [^0 SUB \ [^0 Tr$

²³ Čitateľa oboznámeného s TIL však vyzývam na kontrolu pomocou typovej analýzy a syntézy.

${}^0\text{Edo}] \text{ } {}^0x \text{ } {}^0x]$ zachytáva substitúciu významu antecedentu „Edo“ za individuovú premennú x , prislúchajúcej výrazu „niekoho“. Pomocou týchto troch substitúcií vybudujeme konštrukciu propozície «*Edo videl Edovu matku*» pre čítanie (5a) a «*Edo videl Martinovu matku*» pre čítanie (5b). Takto pre prípad (5) uchopujeme význam vynechaných častí daných eliptických viet.

Nasleduje sémantická analýza vety (6) (t. j. „Watson podozrieval všetkých, ktorých Sherlock Holmes“):

Prvý krok:

(6a) Watson podozrieval všetkých, s ktorými bol v tom istom vzťahu Sherlock Holmes.

Druhý krok:

(6pa) Watson podozrieval všetkých, s ktorými bol v tom istom vzťahu, teda vzťahu podozrievania, Sherlock Holmes.

(6pTa) $[\lambda\omega\lambda t \ [\ [{}^0\text{All} \ [\lambda y \ [{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Podozrieval}_{wt}] \ {}^0R \ {}^0R]$
 ${}^0\text{Sherlock } y]]] \ [\lambda x[{}^0\text{Podozrieval}_{wt} \ {}^0\text{Watson } x]] \] \]$

Sémantická analýza vety (6pa) je jednoduchšia ako v prípade (5). Vo vete (6) sme totiž vypustili iba jeden výraz. Preto schému sémantickej analýzy prístavkového výrazu nahrádzajúceho anaforický výraz používame iba raz. V (6pTa) som toto použitie zvýraznil. Substituujeme v ňom význam antecedentu „podozrieval“ za premennú R , ktorá zodpovedá výrazu „v tom istom vzťahu“.

Posledným problematickým príkladom bol semivýraz (8) (t. j. „Martin videl svoju matku a Edo verí, že aj Jana“). Dostaneme:

(8a) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Martinovou.

(8Ta) $[\lambda\omega\lambda t \ [{}^0a \ [{}^0\text{Videl}_{wt} \ {}^0\text{Martin} \ [{}^0\text{Matka}_{wt} \ {}^0\text{Martin}]] \ [{}^0\text{Verí}_{wt} \ {}^0\text{Edo}$
 ${}^0[{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Videl}_{wt}] \ {}^0R \ {}^0R] \ {}^0\text{Jana} \ [{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Matka}_{wt}] \ {}^0S$
 ${}^0S] \ [{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Martin}] \ {}^0x \ {}^0x]] \] \] \]$

(8b) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Janinou.

(8Tb) $[\lambda\omega\lambda t \ [{}^0a \ [{}^0\text{Videl}_{wt} \ {}^0\text{Martin} \ [{}^0\text{Matka}_{wt} \ {}^0\text{Martin}]] \ [{}^0\text{Verí}_{wt} \ {}^0\text{Edo}$
 ${}^0[{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Videl}_{wt}] \ {}^0R \ {}^0R] \ {}^0\text{Jana} \ [{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Matka}_{wt}] \ {}^0S$
 ${}^0S] \ [{}^2[{}^0\text{SUB} \ [{}^0\text{Tr} \ {}^0\text{Jana}] \ {}^0x \ {}^0x]] \] \] \]$

- (8c) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Edovou.
- (8Tc) [$\lambda\omega\lambda t$ [${}^0 a$ [${}^0 \text{Videl}_{wt}$ ${}^0 \text{Martin}$ [${}^0 \text{Matka}_{wt}$ ${}^0 \text{Martin}$]]] [${}^0 \text{Veri}_{wt}$ ${}^0 \text{Edo}$ ${}^0 [{}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Videl}_{wt}] {}^0 R {}^0 R]$ ${}^0 \text{Jana}$ [${}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Matka}_{wt}] {}^0 S {}^0 S]$ ${}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Edo}] {}^0 x {}^0 x]$]]]]]

V týchto analýzach premenné R , S , x zodpovedajú po rade výrazom „v tom istom vzťahu“, „niekým“, „niekoho“. Sústreďme sa na časti prístavkových viet, ktoré zodpovedajú doplneným anaforickým výrazom vo vetách. Použité eliptické vety obsahovali rovnako ako v prípade (5) tri anaforické výrazy. Ak si pozorne všimneme sémantické analýzy viet, uvidíme, že sa v nich navrhnutá schéma objavuje trikrát.²⁴ Zároveň si všimneme, že sa tieto tri analýzy líšia iba na jednom mieste v závislosti od toho, o koho matke sa hovorí. Opäť je rozdiel iba pri poslednom použití Trivializácie. Konštrukcia ${}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Videl}_{wt}] {}^0 R {}^0 R]$ zachytáva substitúciu významu antecedentu „videl“ za premennú pre binárne relácie R . Konštrukcia ${}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Matka}_{wt}] {}^0 S {}^0 S]$ zachytáva substitúciu významu antecedentu „matka“ za premennú S . Konštrukcia ${}^2 [{}^0 \text{SUB} [{}^0 \text{Tr } {}^0 \text{Martin}] {}^0 x {}^0 x]$ zachytáva v prípade analýzy (8Ta) substitúciu významu antecedentu „Martin“ za individuovú premennú x . V zápisoch (8Tb) a (8Tc) funguje táto schéma podobne.

Navrhovaná schéma analýzy fungovala uniformne vo všetkých skúmaných prípadoch. Ak sme v eliptických vetách vynechali anaforické výrazy a ak pomocou prístavkových zvrátov môžeme simulovať význam anaforických výrazov, tak sémantický obsah vynechaných častí eliptických viet stojí na rovnakých sémantických princípoch ako v prípade anaforicky použitých výrazov. Teraz môžeme odpovedať na rovnaké otázky ako TMI:

Q: Ako rozumieme, aké časti boli vypustené z eliptickej vety?

A: Vieme to vďaka tomu, že pre každú eliptickú vetu dokáže kompetentný používateľ jazyka nájsť zodpovedajúcu prístavkovú vetu.

Q: Aký je vzťah významov antecedentu a vynechanej časti eliptickej vety?

A: Ak sú sémantické obsahy vynechaných častí eliptických viet totožné so sémantickými obsahmi anaforických výrazov, významy vynechaných častí eliptických viet a ich antecedentu nie sú totožné. Sémantické

²⁴ V týchto analýzach sme prvú vetu súvetí analyzovali zjednodušene, keďže obsahuje anaforický výraz „svoju“.

tický obsah antecedentu sa však spolu so sémantickým obsahom premennej podieľa v navrhovanej reprezentácii významu vynechanej časti eliptickej vety pomocou prístavkových viet.

Podľa tohto riešenia sú sémantické obsahy eliptickej a jej zodpovedajúcej neeliptickej, teda prístavkovej vety rozdielne.

Literatúra

- CMOREJ, P. (2005): Semivýrazy a výrazy. In: Sousedík, P. (ed.): *Jazyk, logika, věda*. Praha: Filosofia, 63-88.
- DUŽÍ, M. (2006): Anafora a význam. In: Zouhar, M. (ed.): *Jazyk z pohľadu sémantiky, pragmatiky a filozofie vedy*. Bratislava: Filozofický ústav SAV, 99-136.
- DUŽÍ, M. – JESPERSEN, B. (2013): Procedural Isomorphism, Analytic Information and β -conversion by Value. *Logic Journal of IGPL* 21, No. 2, 291-308.
- DUŽÍ, M. – MATERNA, P. (2012): *TIL jako procedurální logika. Průvodce zvědavého čtenáře Transparentní intensionální logikou*. Bratislava: aleph.
- GAHÉR, F. (2002): Anafora a pojmové postoje. In: Gáliková, S. (ed.): *Filozofia Ludwiga Wittgensteina*. Bratislava: Veda, 130-155.
- HAEGEMAN, L. (1994): *Government and Binding Theory*. Oxford: Blackwell.
- KOSTEREC, M. (2013): Anaforický reťazec. *Organon F* 20, mimoriadne číslo 2, 140-159.
- MAY, R. C. (2002): Ellipsis. In: Nadel, L. (ed.): *Encyclopedia of Cognitive Science*. Basingstoke: Palgrave Macmillan, 1094-1102.
- NEALE, S. (2004): This, That, and the Other. In: Reimer, M. – Bezuidenhout, A. (eds.): *Descriptions and Beyond*. Oxford: Clarendon Press, 68-182.
- SAG, I. (1976): *Deletion and Logical Form*. Ph.D. Thesis. Cambridge: Massachusetts Institute of Technology.
- SCHEFFLER, T. (2013): Ellipsis, Subsentential Speech, and the Contextualism Discussion. Free Book. Dostupné na: <http://books5.net/e/ellipsis-subsentential-speech-and-the-contextualism-discussion-w4335-book.html>. s.3.
- WILLIAMS, E. (1977): Discourse and Logical Form. *Linguistic Inquiry* 8, 101-139.
- ZLATUŠKA, J. (1993): *Lambda-kalkul*. Brno: Masarykova univerzita.
- ZOUHAR, M. (2004): Anafora a referencia. In: Zouhar, M. (ed.): *Používanie, interpretácia a význam jazykových výrazov*. Bratislava: Veda, 128-143.
- ZOUHAR, M. (2009): *Teória kvantifikácie a extenzionálna sémantika prirodzeného jazyka*. Bratislava: Filozofický ústav SAV.
- ZOUHAR, M. (2011): *Význam v kontexte*. Bratislava: aleph.

On Changing the Past¹

ALEX BLUM

Department of Philosophy, Bar-Ilan University
Ramat - Gan 52900, Israel
alex.blum@biu.ac.il

RECEIVED: 05-04-2013 • ACCEPTED: 02-05-2013

ABSTRACT: We confront two seemingly incompatible positions in regard to the past. One, the modal status of a proposition is unchangeable; and two, that omnipotence is trumped only by necessity.

KEYWORDS: Change – modal status – necessity – omnipotence – past.

The view that the past could have turned out differently and yet is unchangeable appears to challenge at least one of two ingrained beliefs; one, that the modal status of a proposition is not changeable;² and two, that omnipotence is trumped only by necessity. And yet the past is unchangeable.

We accept both beliefs. The contingency of the propositions that make up the past remain contingent. For what makes a true or false proposition contingent is that the proposition could have been false or true. Consequently the proposition that the house once built will be painted within a week is contingent and remains contingent even after the house is built

¹ I'm deeply grateful to Yehuda Gellman for stimulating the paper and for his insightful comments on the subject. I also want to thank Peter Genco, Laurence Goldstein, Charlotte Katzoff and David Widerker for the very helpful discussions. I also wish to acknowledge the helpful suggestions of the editors.

² That is, the truths that make up the past, if ever contingent, remain contingent.

and painted; for it remains true that the house could have been built and never painted.³ But if it is true that the house was built and painted within a week it cannot also be false that the house was built and painted within a week.⁴

The impossibility of changing the past that thus emerges stems from the impossibility of a proposition being both true and false. Hence the changelessness of the past is necessary and thus it supersedes omnipotence.

³ Or, if you like, it does not become true in every possible world.

⁴ But could it not have been the case that it was true that the house was built at time *t* and painted within a week and then became false that the house was built at time *t* and painted within a week. No. For then the house could have been both built and painted within a week at time *t* and not be both built and painted within a week at the same time *t*. Couldn't the past be erased, as though it never happened? But it did happen.

Externalismus a skeptická výzva

MAREK PICHA

Katedra filozofie. Filozofická fakulta. Masarykova univerzita
Arna Nováka 1/1. 602 00 Brno. Česká republika
picha@phil.muni.cz

Organon F 20 (1) 2013 uveřejnil text Martina Nuhlička *Prečo externalistické kritériá zdôvodnenia principiálne zlyhávajú?* (Nuhlíček 2013). Jde o srozumitelný text věnovaný kritice významné epistemologické koncepce. Obávám se ale, že jeho autor zvolil k dosažení subtilního cíle příliš robustní a banální argumenty. Ve své reakci budu nanejvýš stručný. Nejprve krátce představím externalistickou koncepci a důvody jejího vzniku, poté popíši Nuhličkovu argumentaci ve strukturované podobě a ukáži její slabinu.

1. Radikální skepticismus

Autor píše, že externalismus vznikl jako reakce na těžkosti, které měl internalismus se skeptickou výzvou reprezentovanou např. Klamajícím démonelem či Mozkem v kádi (srov. Nuhlíček 2013, 39-40). Jak taková skeptická výzva vlastně vypadá?

(S1) Není možné určit, která přesvědčení jsou pravdivá.

Tato radikální pozice tvrdí, že kterékoli naše přesvědčení – o světě kolem, o našem vnitřním životě, o minulosti, o matematice – může být nepravdivé. Všechna naše přesvědčení mohou být produkována postupy, na jejichž začátku není odpovídající fakt, nýbrž záludný protihráč.

Nuhličkem uvedené myšlenkové experimenty jsou příklady podporující S1, což je v kontextu další diskuse zavádějící. Proč? Radikální skeptické výzvě typu S1 totiž není schopna uspokojivě čelit žádná teorie znalosti, inter-

nalistická ani externalistická. Přesněji řečeno, všechna navrhovaná řešení této skeptické výzvy jsou buď neproduktivní, nebo předpokládají přijetí velmi kontroverzních principů.¹ I ty nejsebevědomější epistemologické teorie jsou založeny na předpokladu, že to či ono přesvědčení je pravdivé a může sloužit jako vztažný bod pro další úvahy. Až na základě takového výchozího materiálu lze zobecňovat a budovat tak teorii. Bez několika pravdivých přesvědčení zkrátka nevíme, kam se při teoretizování chceme dostat. Nevíme, s čím návrhy porovnávat; nevíme ani, jak je vlastně zdůvodňovat. Kdyby býval externalismus skutečně vznikl s ambicí čelit skepticizmu typu S1, šlo by o mrtvě narozené dítě.

Je-li závěrem Nuhličkova textu teze, že si externalismus nedokáže poradit s radikálním skepticizmem typu S1, pak je to teze vysoce přijatelná, přijímaná a banální. Jde o notoricky známou věc, takže věřím, že chtěl Martin Nuhlíček ve svém textu říci něco víc.

2. Konformní skepticizmus

Budu předpokládat, že jádrem autorovy kritiky externalismu není pochybnost týkající se *pravdivosti* přesvědčení obecně, nýbrž pochybnost týkající se *zdůvodňování* přesvědčení, jejichž pravdivost neznáme. Jinak řečeno, autora zajímá, zda je ta či ona teorie schopna určit přijatelná kritéria, při jejichž splnění bude přesvědčení „epistemicky dobré“, tedy zdůvodněné.

Jak při hledání těchto zdůvodňovacích kritérií postupují internalisté? Jak externalisté? V zásadě stejně. Všimají si vlastností, jež mají naše pravdivá přesvědčení, tyto vlastnosti pak zobecňují a činí z nich právě ono epistemické kritérium. Internalisté zohledňují vlastnosti související se vztahem přesvědčení a jeho nositele, externalisté berou v potaz vlastnosti související se vztahem přesvědčení a jeho původu. Všimněme si, že obě strategie vycházejí z pravdivých přesvědčení, obě předpokládají popření S1.

Konformní, méně radikální skeptik má pochybovačný postoj ke zdůvodnění. Jeho názor lze vyjádřit tezí, že *zdůvodnění není dostatečnou podmínkou pravdivosti přesvědčení*. Konformní skeptik tedy tvrdí, že byť přijmeme pravdivost některých přesvědčení jako výchozí bod úvahy, kritéria na nich postavená nebudou z nějakého důvodu vhodná. Ve většině koncepcí je vztah mezi zdůvodněním a pravdivostí ve skutečnosti o něco oslaben; zdůvodnění není

¹ Například přijetí karteziánského pravidla o vztahu mezi obsahem a původcem ideje nebo sémantického externalismu.

chápano jako dostatečná podmínka pravdivosti přesvědčení, nýbrž jako symptom, že dané přesvědčení je spíše pravdivé, tedy jako ukazatel tendence k pravdivosti. Konformní skeptik potom tvrdí následující tezi:

(S2) Zdůvodnění není ukazatelem tendence k pravdivosti.

Takový skeptik tradičně útočí na internalistická kritéria a ukazuje příklady, kdy i sebelepší povědomí o důvodech může systematicky chybovat. Snaží se například ukázat, že naše paměť se mýlí příliš často na to, abychom mohli považovat vlastní vzpomínky za spíše pravdivé. Až zde je bod, kdy do diskuse vstupuje externalismus.

3. Externalismus

Obhájci externalismu přijímají skeptickou výzvu typu S2 a poměrně jednoduše na ni reagují. Jsou si dobře vědomi skutečnosti, že pravdivá přesvědčení se mohou zásadně lišit svým obsahem i vnitřní strukturou, některých přesvědčení si dokonce ani nemusíme být vědomi. Lze však vystopovat pravdivost mezi pravdivými přesvědčeními a způsobem, jak tato přesvědčení souvisí s fakty, jichž se týkají. Přesněji, jde o souvislost *přesvědčení o propozici p s faktem, který propozici p činí pravdivou*. Různé externalistické koncepce tuto souvislost popisují různě, nejvlivnější koncepcí v daném směru je reliabilismus, a na ten se také Nuhlíček soustředí.

Od pravdivých přesvědčení ke zdůvodněným přesvědčením je daleká cesta a reliabilismus je koncepce, která se tuto cestu snaží pokud možno co nejvíc zkrátit. Cílem je vybudování epistemologické teorie, která bude lépe řešit problémy spojené nikoli s klamajícím démonem, nýbrž s dětmi, idioty a zvířaty – koncepce, která umožňuje připsat zdůvodněná přesvědčení i bytostem, které nereflektují vlastní mentální stavy, nesledují své zdůvodňovací řetězce, nedokáží konceptualizovat, zkrátka nejsou vybaveny schopnostmi, které internalismus v té či oné podobě předpokládá.

(Rel) Jakýkoli proces, který v případě, že p , vede spíše k přesvědčení, že p , a v případě, že $non-p$, spíše nevede k přesvědčení, že p , je epistemicky důvěryhodný proces.

Je jedno, jak přesně je tento proces realizován. Je jedno, jestli si tu nebo onu část procesu uvědomujeme. Roli hraje pouze to, zda budeme mít za přítomnosti p spíše přesvědčení, že p a vice versa.

Dovolím si krátké exposé shrnout. Externalismus nevznikl jako reakce na radikální skepticismus, nýbrž jako reakce na konformní skepticismus. Na ten reaguje vlastně triviálně: zdůvodnění je z *definice* ukazatelem tendence k pravdivosti. Myšlenkové experimenty Klamající démon a Mozek v kádi nejsou v kontextu polemiky mezi internalismem a externalismem relevantní. Tolik základní představení motivace a výchozích tezí, nyní se podívejme na Nuhličkovu kritiku. Jeho hlavní argument představený v Nuhliček (2013, 42-43) si dovolím rekonstruovat následovně:

Argument proti externalismu

1. V rámci externalismu nelze odlišit zdůvodněná přesvědčení od nezdůvodněných.
2. Uspokojivá teorie epistemického zdůvodnění musí umožňovat odlišení zdůvodněných a nezdůvodněných přesvědčení.
3. Tudíž externalismus není uspokojivou teorií epistemického zdůvodnění.
4. Uspokojivá teorie znalosti musí předpokládat uspokojivou teorii epistemického zdůvodnění.
5. Tudíž externalismus není uspokojivou teorií znalosti.

Struktura argumentu je nekomplikovaná. Není-li externalismus schopen předložit použitelná kritéria zdůvodnění, nelze s ohledem na klasickou definici považovat externalismus za přijatelnou teorii znalosti. Klíčovou roli v argumentaci hraje první premisa; ta vyjadřuje stanovisko, které bychom mohli označit za *reliabilistický skepticismus*:

- (S3) Není možné určit, která přesvědčení vznikla důvěryhodným procesem.

Tuto tezi podporuje Martin Nuhliček dvěma podargumenty:

- #1 *Argument regrese* („problém třetí osoby“): Nuhliček v (2013, 43-44) popisuje situaci, kdy požádáme třetí osobu, aby zhodnotila, zda je naše přesvědčení zdůvodněné, či nikoli. V reliabilistickém pojetí to znamená, že dotyčná osoba bude zkoumat, zda je naše přesvědčení stejného druhu jako jiná přesvědčení, jež přijímáme jakožto spíše pravdivá. Tento proces zkoumání přesvědčení je však sám o sobě procesem přijímání a zdůvodňování přesvědčení (o přesvědčení někoho jiného) a i na něj by měla být vztažena reliabilistická kritéria. Každé další zanoření jen požadavek zdůvodnění oddálí.

- #2 *Argument dogmatismem* („problém první osoby“): Tento druhý podargument je preventivním odmítnutím jedné možné odpovědi na předchozí, první podargument. Nuhlíček zde tvrdí, že epistemický regres nelze zastavit přijetím primitivního výchozího zdůvodnění. Popisuje situaci, kdy nějaký vlastní proces získávání přesvědčení sami označíme za důvěryhodný. Takový dogmatický způsob zastavení epistemického regresu není zdůvodněný a nesplňuje reliabilistická kritéria, což je podle Nuhlíčka chyba.

Moje výtku je následující. Domnívám se, že autor se v podargumentech obloukem vrací ke kritice reliabilismu opřené o radikální skepticismus typu S1. Jak jsem již psal výše, reliabilismus předpokládá jistá přesvědčení jako pravdivá a z nich odvozuje kritérium zdůvodnění. Tato pravdivá přesvědčení jsou tedy ona epistemická primitiva, u kterých se zdůvodňovací řetězec dříve nebo později zastaví. Reliabilismus tudíž nemá s epistemickým regresem problém, protože výchozí body jsou prostě a jednoduše zavedeny. Zní to jako laciné řešení, ale jiné bohužel v epistemologii nemáme.

Nuhlíček tak vlastně vyčítá reliabilistickému zdůvodnění nikoli neschopnost zastavit se u zdůvodněných primitiv, nýbrž fakt, že taková zdůvodněná primitiva nemusí být pravdivá. Argumenty #1 i #2 tak ve skutečnosti podporují tezi S1. Jak již bylo řečeno, to je sice skutečně radikální výtku, které externalismus čelit nedokáže, avšak takové zjištění je banální. Navíc k němu není třeba rozlišovat mezi internalismem a externalismem, stačí hovořit o regresu při zdůvodňování obecně.

Abych to shrnul, Martin Nuhlíček napsal text, který srozumitelně popisuje některá úskalí epistemologie. Domnívám se ale, že argumenty, které vůči externalismu předkládá, jsou příliš silného kalibru a jsou příliš obvyklé. Je to stejné, jako byste dlouze kritizovali konkrétní značku auta za to, že auta obecně znečišťují vzduch.

Literatura

- NUHLÍČEK, M. (2013): Prečo externalistické kritériá zdôvodnenia principiálne zlyhávajú? *Organon F* 20, č. 1, 37-49.

Pravdivosť alebo zdôvodnenie? (Odpoveď M. Nuhlíčkovi)

MILOŠ TALIGA

Katedra filozofie. Fakulta humanitných vied. Univerzita Mateja Bela v Banskej Bystrici
Tajovského 40. 974 01 Banská Bystrica. Slovenská republika
Milos.Taliga@umb.sk

Tento článok je odpoveďou na Nuhlíčkovu stať (2013a), dotýka sa však aj problémov, ktorým sa Nuhlíček venoval v (2013b). V 1. časti sa pokúsim objasniť, prečo relevantný typ zdôvodnenia musí súvisieť s pravdivosťou toho, čo je zdôvodňované. Vďaka tomu sa presuniem k diskusii o internalistických a externalistických teóriách zdôvodňovania, ktorá bude pokračovať aj v 2. časti. 3. časť je zhrnutím negatívnych dôsledkov pre podmienku zdôvodnenia a 4. časť pripomenutím kľúčového rozdielu medzi zdôvodnením a kritikou.

1. Otázka súvisu zdôvodnenia a pravdivosti

Ďakujem Nuhlíčkovi za jeho reakciu (2013a) a aj za memento, ktoré uvádza na jej konci. Nevieť však, či ho uspokojím, ak poviem, že mi porozumel správne, iba pokiaľ ide o otázku, či zdôvodnenie chápem ako zdôvodnenie pravdivosti toho, čo je zdôvodňované. Ďalšie jeho interpretácie mojich názorov musím už spresniť. Podľa prvej z nich sú „zdôvodnenie v Taligovom chápaní a nezávislý dôkaz pravdivosti daného tvrdenia, resp. presvedčenia ... ekvivalentné“ (Nuhlíček 2013a, 143). Problém práve citovanej interpretácie sa ukazuje, keď Nuhlíček komentuje rozlišovanie dostatočného (resp. definitívneho) a čiastočného zdôvodnenia, ktoré som uviedol v Taliga (2009, 353). Nuhlíček píše:

Aj toto delenie je zrozumiteľné vo svetle chápania zdôvodnenia ako dôkazu pravdivosti: „definitívne zdôvodnenie“ môžeme chápať ako

kompletný dôkaz pravdivosti daného výroku, zatiaľ čo pod „čiasočným zdôvodnením“ si môžeme predstaviť určitú podporu daného tvrdenia či presvedčenia, ktorá zvyšuje pravdepodobnosť, že je pravdivé (aj keď táto podpora nie je priamo dôkazom jeho pravdivosti). (Nuhlíček 2013a, 144)

Delenie na definitívne a čiasočné zdôvodnenie má byť teda podľa Nuhlíčka *zrozumiteľné* vo svetle chápania zdôvodnenia ako dôkazu pravdivosti, pričom však čiasočné zdôvodnenie určitého tvrdenia či presvedčenia „nie je priamo dôkazom jeho pravdivosti“. To nevyzerá ako konzistentná interpretácia. Pritom si však Nuhlíček správne všima, že zdôvodnenie v mojom chápaní „predstavuje nejaký druh podpory v prospech pravdivosti presvedčení“ (Nuhlíček 2013a, 143). Tento názor by som prerozprával tak, že zdôvodnením je ľubovoľné zret'azenie dvoch skupín výrokov (presvedčení, teórií a pod.), z ktorých jedna má buď definitívne, alebo čiasočne zdôvodniť, resp. podporiť, pravdivosť tej druhej (pozri Taliga 2009, 353).¹

Hneď však musím dodať dve poznámky. Po prvé, takémuto chápaniu zdôvodnenia sa nevyšmyknú ani rôzne návrhy (často externalistické), ktoré nehovoria o zdôvodnení priamo ako o vzťahu medzi dvomi skupinami výrokov, ale – napríklad – ako o zdroji (pôvode), ktorý má hovoriť čosi v prospech zdôvodňovaných tvrdení či presvedčení. (K tomuto bodu sa vrátim v 2. časti.) Po druhé, v súčasnosti by som bol pri vymedzení čiasočného zdôvodnenia ústretovejší a neobmedzoval by som ho na podmienku uvedennú v Taliga (2009, 353), aby bol stupeň podpory (vyjadrený v pravdepodobnostných termínoch) vyšší ako $\frac{1}{2}$. Ak totiž podobne ako Hempel v (1965) rozlíšime medzi absolútnou a relatívnou konfirmáciou, potom by sme mohli vkladať nádej do návrhu, že v prípade relatívnej konfirmácie stačí, ak sa stupeň nededuktívnej podpory určitého tvrdenia *b*, ktoré má byť (v pravdepodobnostnom zmysle) podporené vzhľadom k inému tvrdeniu *e*, zvýši čo i len o desatinu alebo o stotinu percenta (podľa toho, čo sme ochotní považovať za relevantný nárast). Túto nádej však sklame Popperova-Millerova teoréma, ktorú som parafrázoval v Taliga (2009, 368–371): Evidencia *e* nemôže v žiadnom zmysle slova zvýšiť pravdepodobnosť tej časti obsahu hypotézy *b*, ktorou *b* prekračuje *e*. To vyvracia ideu pravdepodobnostnej nededuktívnej podpory, pretože časť obsahu *b*, ktorou *b* prekračuje *e*, je práve

¹ Uvedená parafráza nemá byť definíciou zdôvodňovania (keďže by išlo o definíciu kruhovú), ale iba zdôraznením rozlišovania medzi definitívnym a čiasočným zdôvodnením namiesto chápania zdôvodnenia ako nezávislého dôkazu pravdivosti.

tou časťou obsahu h , ktorá nie je od e deduktívne závislá. K Hempelovej myšlienke absolútnej confirmácie stačí povedať len toľko, že ak by sme o evidencii e (ktorá má confirmovať hypotézu h v nededuktívnom zmysle) aj vedeli, že je (objektívne) pravdivá, nedokázali by sme čeliť známemu problému indukcie, ale by sme mu, naopak, spočívali priamo v náruči. Problém indukcie sa nedá riešiť pomocou tvrdenia, že hypotéza h je zdôvodnená evidenciou e , pretože e pravdivo opisuje výskyt tých udalostí, ktoré predpovedali predikcie odvodené z h . Také tvrdenie by takisto podliehalo problému indukcie, keďže od začiatku vieme, že hypotéza h evidenciu e presahuje.

Druhé spresnenie, ktoré musím dodať k Nuhličkovej interpretácii mojich názorov, sa týka jeho domnienky, že moje chápanie zdôvodnenia sa vymyká ako internalistickému tak aj externalistickému chápaniu (pozri Nuhlíček 2013a, 146-147). Najprv sa vyjadrím k Nuhličkovmu vykresleniu internalistického zdôvodnenia ako zdôvodnenia, ktoré „nemá nič spoločné s pravdivosťou presvedčení“ (Nuhlíček 2013a, 147). To je totiž základ, na ktorom Nuhlíček postavil svoje tvrdenie, že moje chápanie zdôvodnenia (ako zdôvodnenia pravdivosti) nie je internalistické. Až potom prejdem k zdôvodneniu externalistickému.

Ak Nuhličkovi správne rozumiem, tak za príklad internalistického zdôvodnenia, ktoré nemá nič spoločné s pravdivosťou presvedčení, považuje Lehrerovu ideu akceptačného systému „presvedčení konkrétneho subjektu“ (Nuhlíček 2013a, pozn. 15). Lehrer však na viacerých miestach zdôrazňuje, že pre poznanie je potrebná „akceptácia, ktorej cieľom je pravda“ (Lehrer 1999, 19), teda, že jeho zámerom je „predložiť teóriu poznania a zdôvodnenia, ktorou vysvetlíme, prečo máme zdôvodnenie akceptovať informáciu ako správnu, a tvrdiť, že to vieme“ (Lehrer 1999, 36). Na otázku „Ako akceptovaný systém generuje zdôvodnenie?“ Lehrer odpovedá: „Tak, že nám povie, či je racionálne akceptovať nejakú vec pri hľadaní pravdy“, t. j. vtedy, „keď chceme akceptovať len to, čo je pravdivé“ (1999, 94). Lehrer síce podrobne opisuje tzv. osobné zdôvodnenie v termínoch internej koherencie (1999, 6. kap.), napokon však propaguje myšlienku tzv. nevyvráteného, resp. neanulovaného zdôvodnenia (undefeated justification), o ktorom tvrdí, že „presahuje internú koherenciu a preniká k pravde“ (Lehrer 1999, 116). V kontraste k externalizmu trvá na tom, že „zdôvodnenie presvedčenia, ktoré človek pozná ako pravdivé, sa ... zakladá na určitej evidencii vtedy a len vtedy, keď to, že má evidenciu, vysvetľuje, ako vie, že presvedčenie je pravdivé“ (Lehrer 1999, 136).

Máme tu teda príklad internalistu, podľa ktorého zdôvodnenie súvisí s pravdivosťou zdôvodňovaných presvedčení úplne bytostne.² Téma, podľa ktorej internalistické zdôvodnenie „nemá nič spoločné s pravdivosťou presvedčení“ (por. Nuhlíček 2013a, 147), je tak vyvrátená. Moje chápanie zdôvodnenia je však zlučiteľné s Lehrerovou internalistickou teóriou. Chápem, že môže byť nezlučiteľné s *inou* internalistickou teóriou zdôvodnenia, podľa ktorej zdôvodnenie *nesúvisí* s pravdivosťou zdôvodňovaného presvedčenia, v takom prípade by som však súhlasil s Lehrerom a tvrdil by som, že ide o *epistemicky irelevantný* typ zdôvodňovania.

2. Epistemické zdôvodnenie a externalizmus

Problém súvisu zdôvodnenia a pravdivosti sa vynára aj pri Nuhlíčkovej diskusii o externalistických teóriách zdôvodnenia. Nuhlíček píše:

Ani externalisti ... neopúšťajú predpoklad, že predmetom zdôvodnenia je samotné presvedčenie, a nie jeho pravdivosť (ako u Taligu). Zdôvodnené presvedčenie je podľa nich presvedčenie, ktoré nadobúda sklon byť pravdivé ... na základe svojho *pôvodu*. (Nuhlíček 2013a, 147)

Nie som si istý, ako chápať práve citovaný názor, hoci poznám viacero externalistických teórií zdôvodnenia, vrátane tých, ktoré spomína Nuhlíček. Zdá sa mi totiž, akoby Nuhlíček tvrdil, že podľa externalistov má zdôvodnené presvedčenie (vďaka svojmu pôvodu) *sklon byť pravdivé*, no zdôvodnenie presvedčenia nemá mať *nič spoločné s pravdivosťou* presvedčenia. Ani to nevyzerá ako konzistentná interpretácia externalistického zdôvodnenia.³ Súhlasím s tým, že podmienky pravdivosti a zdôvodnenia sa „tradične“ v analytickej epistemológii 20. a 21. storočia uvádzajú osobitne, nemôžem však už súhlasiť s tým, že by tieto dve podmienky boli (alebo mali byť) „navzájom nezávislé“ (Nuhlíček 2013a, pozn. 9). Skôr si myslím, že väčšina súčasných epistemológov by tvrdila to, čo Hetherington, a to, že „na rozdiel od pragmatického zdôvodnenia má byť epistemické zdôvodnenie nejakým spôso-

² A pri diskusii nefalibilistického fundacionalizmu Lehrer píše: „Ak je zdôvodnenie irelevantné pre pravdivosť základných presvedčení, nie je potrebné pre zdôvodnenie akceptácie alebo na získanie poznania“ (Lehrer 1999, 41). Relevantné zdôvodnenie musí teda podľa neho súvisieť s pravdivosťou.

³ Ak som však Nuhlíčkovi iba zle porozumel, vopred sa mu ospravedlňujem.

bom úzko spojené s pravdou“, t. j. že „epistemické zdôvodnenie presvedčenia je zdôvodnením pravdivosti presvedčenia, nie jeho užitočnosti alebo spoločenskej úctyhodnosti“ (Hetherington 1996, 22).

Ak má byť podmienka zdôvodnenia *relevantnou* podmienkou poznania, musí byť *bud'* nutnou, *alebo* dostatočnou podmienkou poznania. V oboch prípadoch musí existovať spojenie medzi zdôvodnením a pravdivosťou. Ak by totiž bola podmienka zdôvodnenia (PZ) *dostatočnou* podmienkou poznania (P), platil by kondicionál $PZ \rightarrow P$. Splnenie PZ by teda *zaručovalo* splnenie P, čiže *aj splnenie podmienky pravdivosti* (keďže podľa tripartitnej definície nepravdivé presvedčenia poznatkami nie sú). V mojom chápaní by, v takom prípade, išlo o *definitívne* zdôvodnenie, z ktorého vyplýva pravdivosť zdôvodneného presvedčenia. Ak by bola podmienka zdôvodnenia (PZ) *nutnou* podmienkou poznania (P), platil by kondicionál $P \rightarrow PZ$. Hoci by teda *nemohlo dôjsť k poznaniu bez splnenia* PZ, splnenie PZ by nezaručilo poznanie. V takom prípade by, podľa môjho chápania zdôvodnenia, išlo o zdôvodnenie *čiastočné*, pri ktorom nie je pravdivosť zdôvodneného presvedčenia ničím zaručená, ak však platí kondicionál $P \rightarrow PZ$, *nejaké spojenie medzi zdôvodnením a pravdivosťou tu byť musí*.

Nemôžem teda súhlasiť s Nuhličkovým názorom, že „zdôvodnenie *pravdivosti* poznatkov je ... niečím, čo protirečí litere tripartitnej definície, podľa ktorej zdôvodňujeme samotné presvedčenia, a nie ich pravdivosť“ (Nuhlíček 2013a, pozn. 9). Napokon, vyvrátením tohto názoru je aj jedna z najstarších epistemologických doktrín – nefalibilistický fundacionalizmus, podľa ktorého je *vď' aka zdôvodneniu zaručená pravdivosť* toho, čo je zdôvodnené. Preto radšej opäť počítam s možnosťou, že som Nuhličkovi porozumel nesprávne. Zatiaľ to však vyzerá tak, že v pozadí nášho nesúhlasu stoja rozdielne odpovede na otázku, či je dôležité, aby zdôvodnenie presvedčenia, že *p* (ktoré má byť poznatkom), viedlo k pravde alebo ju aspoň, takpovediac, sprevádzalo.⁴ Tvrdím, že *ak zdôvodnenie nesúvisí s pravdivosťou zdôvodňovaného presvedčenia, potom je úplne zbytočné, aby bolo vyžadované*, pretože v prípade takého zdôvodnenia neplatia kondicionály $PZ \rightarrow P$ a $P \rightarrow PZ$. Ak je cieľom poznania (aj) pravda, potom je zdôvodnenie, ktoré pravdivosť nijakým spôsobom nikomu nenaznačuje, irelevantné.

⁴ V oboch prípadoch („viest' k“; „sprevádzať“) má byť zdôvodnenie presvedčenia príznakom (buď definitívnym alebo čiastočným) pravdivosti presvedčenia. Otázka súvislosti zdôvodnenia a pravdivosti sa zvykne označovať aj ako „truth-conducivity of justification“, pozri napríklad Lammenranta (2008, časti 8 a 9).

Ak sa však tvrdí, že zdôvodnenie nejakým spôsobom naznačuje pravdivosť, potom to už s internalizmom a externalizmom ide dolu vodou. V prípade internalizmu musí totiž poznávajúca osoba *S* *vedieť*, že dobré dôvody pre jej presvedčenie, že *p*, na ktoré sa odvoláva, jej naznačujú, že *p* je pravdivé. Podľa argumentov opísaných v Taliga (2009) však *S* čosi také *vedieť* *nemôže*. V prípade externalizmu je situácia rovnako zúfalá: Externalista síce tvrdí, že na to, aby *S* (prvá osoba) vedela, že *p*, *S* *nemusí* vedieť, že jej presvedčenie, že *p* je zdôvodnené (stačí, ak je fakticky zdôvodnené a *S* o tom nemusí ani tušiť), *musí* tu však byť iná osoba (tretia osoba), ktorá stanoví, čo je faktickým zdôvodnením pôvodného presvedčenia osoby *S*, že *p*. Tretia osoba musí teda sformulovať teóriu zdôvodnenia, ktorá – pokiaľ má byť pre poznanie relevantná – *musí* *postulovať* súvislosť medzi externalistickým zdôvodnením (označme ho neutrálne symbolom „*Ex*“) a pravdivosťou zdôvodňovaného presvedčenia. A tak sa opäť vynára tradičná otázka, odkiaľ tretia osoba *vie*, že ňou stanovená podmienka faktického zdôvodnenia presvedčenia prvej osoby je *správna* – otázka, ktorá (vo svojom tradičnom šate) vedie naspäť k požiadavke, aby tretia osoba (alebo ďalšia, iná osoba) poskytla *zdôvodnenie* pre to chápanie faktického zdôvodnenia, ktoré (podľa predpokladu) musí splniť prvá osoba *S*, aby mohla vedieť, že *p*.

Túto otázku *musí* externalizmus nejakým spôsobom zodpovedať, aspoň pokiaľ sa nechce stať prázdnu doktrínou, ktorá o spojení pravdivosti pôvodného presvedčenia *p* s jeho postulovaným (a údajne správnym či faktickým) zdôvodnením *Ex* nič nepovie. Ak sa o spojení pravdivosti *p* s *Ex* nič nehovorí, potom je návrh, podľa ktorého sa *p* stáva poznatkom osoby *S* až *vtedy*, keď splní podmienku *Ex*, *irrelevantný*. (Pre čo by malo byť potrebné splniť podmienku *Ex*? A *vzhľadom* k čomu je podmienka *Ex*₁, o ktorej súvislosti s pravdou sa nič nehovorí, lepšia ako podmienka *Ex*₂, o ktorej súvislosti s pravdou sa nehovorí takisto?) Ak sa však tvrdí, že spojenie medzi *Ex* a pravdivosťou *p* existuje, potom musí externalista čeliť klasickým vecným kritickým protiargumentom, ktoré Nuhliček veľmi dobre pozná, ako o tom svedčí jeho *stat'* (2013b). Nepovažujem preto za potrebné podrobne tieto argumenty rozoberať. Namiesto toho sa pokúsím objasniť, ako súvisí externalistické zdôvodňovanie s mojim chápaním zdôvodnenia. Presnejšie povedané, pokúsím sa vysvetliť, prečo sa problémy zdôvodnenia uvedené v Taliga (2009) týkajú aj externalizmu.

Dajme tomu, že nejaká externalistická koncepcia stanoví podmienku zdôvodnenia *Ex*, ktorá ma pomôcť presvedčeniu, že *p* (nejakej prvej osoby *S*) stať sa poznatkom. Nech je touto podmienkou napríklad Goldmanom

navrhnutý vhodný spôsob kauzálneho spojenia medzi faktom, že *p* a presvedčením, že *p*.⁵ Ak by mala byť KP *dostatočnou podmienkou* poznania (P), musel by platiť kondicionál $KP \rightarrow P$. Ak by mala byť KP *nutnou podmienkou* poznania (P), musel by platiť kondicionál $P \rightarrow KP$. V Taliga (2010a) som sa venoval otázke, či je kauzálna podmienka preferovaná Davidsonom nutnou alebo dostatočnou podmienkou poznania, tu sa však pokúsím zhodnotiť KP na všeobecnejšej rovine. Vieme, že podľa externalizmu osoba S nemusí vedieť, že jej presvedčenie spĺňa KP, no teraz sa treba pýtať, či to musí vedieť iná (tretia) osoba. Ak by externalista tvrdil, že nie, jeho koncepcia by stratila čaro: Platnosťou uvedených kondicionálov by sa nikto nemusel zaoberať. Ak by však tvrdil, že tretia osoba⁶ *musí byť* schopná v konkrétnych prípadoch určiť, či prvá osoba S splnila KP, potom čelí kľúčovému problému: *Ako* to tretia osoba zistí a *ak* to aj zistí, *ako* jej to pomôže rozhodnúť, či S disponuje *poznáním*? Pomohlo by to iba vtedy, ak by tretia osoba *vedela*, že platí kondicionál $KP \rightarrow P$. To by si však vyžadovalo, aby tretia osoba vedela, že *vždy*, keď je splnená kauzálna podmienka KP, dochádza k poznaniu P. Taká úloha musí byť zjavne nad sily a schopnosti tretej osoby, a to kvôli problému indukcie: Nikto nemôže byť svedkom *všetkých* prípadov splnenia kauzálnej podmienky KP, pretože evidencia, ktorá je k dispozícii, sa môže zaoberať iba *doteraz zaznamenanými* prípadmi.⁷ Nikto teda nemôže posúdiť, či splnenie KP *zaručuje* nadobudnutie poznania P, pretože to by si vyžadovalo *definitívne* zdôvodnenie kondicionálu $KP \rightarrow P$.

Môže tretia osoba vedieť aspoň to, že platí kondicionál $P \rightarrow KP$, teda to, že kauzálna podmienka KP je nutnou podmienkou poznania P? Najprv si všimnime, že touto otázkou sa tretia osoba vôbec nevenuje otázke, *ako zistí*, či prvá osoba splnila KP. Predpokladajme však, podobne ako predtým, že tretia osoba to zistiť môže.⁸ V takom prípade by zo splnenia KP v prípade

⁵ Na účely tohto článku nie je potrebná detailná analýza Goldmanovej podmienky. V záujme stručnosti ju nazvem *kauzálnou podmienkou* a budem na ňu odkazovať aj pomocou skratky „KP“.

⁶ Prípadne ďalšie osoby, či už filozofi, vedci alebo celé komunity príslušných odborníkov.

⁷ Navyše, ak by ktosi tvrdil, že nejaká teória (dajme tomu vedecká teória) vysvetľuje, prečo kondicionál $KP \rightarrow P$ platí, urobil by prvý krok v nekonečnom regrese, pretože by musel čeliť otázke, čím je vedecká teória zdôvodnená.

⁸ Hoci uvedený predpoklad považujem za neobhájiteľný (stroskotal by na podobných problémoch, ktoré opísal Nuhlíček v (2013b) na príklade reliabilizmu), mojou úlohou je ukázať, prečo externalizmu nepomôže ani jeho ústretové prijatie.

prvej osoby odvodzovala záver, že prvá osoba (ktorá KP splnila) disponuje poznaním. Tretia osoba by teda usudzovala podľa neplatného úsudkového pravidla:

$$\frac{P \rightarrow KP}{KP} \\ \hline P$$

Preto by aj v tomto prípade musela tretia osoba čeliť problému indukcie. V ceste by jej pritom stála Popperova-Millerova teoréma, podľa ktorej splnenie KP neposkytuje *žiadnu nededuktívnu podporu* (vyjadrenú v pravdepodobnostných termínoch) záveru P, ktorý premisy presahuje. Inak povedané, pravdivosť externalistickej teórie nemôže byť zdôvodnená *ani čiastočne*.

Prirodzene, tretia osoba by mohla z nejakých (momentálne nejasných) príčin od samého začiatku tvrdiť, že prvá osoba *má* poznanie a že *preto* aj splnila KP, t. j. tretia osoba by mohla usudzovať takto:

$$\frac{P \rightarrow KP}{P} \\ \hline KP$$

Hoci je uvedené úsudkové pravidlo platné, asi je zrejmé, v čom spočíva chyba. Ak niekto tvrdí prvú premisu (podľa ktorej je kauzálna podmienka KP nutnou podmienkou poznania P) a súčasne druhú premisu (podľa ktorej prvá osoba vie, že *p*), a z toho odvodzuje, že prvá osoba spĺňa KP, potom sa dopúšťa chyby *petitio principii*, pretože otázka, ktorá mala byť ešte len zodpovedaná („Je KP nutnou podmienkou poznania?“), je v skutočnosti už vopred považovaná za rozhodnutú a spolu s ňou aj otázka, či prvá osoba vie, že *p*.

Pre uvedené problémy nemôže tretia osoba zdôvodniť tvrdenie, že kauzálna podmienka je nutnou alebo dostatočnou podmienkou poznania. Tieto problémy, ktoré sú v jadre rovnaké ako problémy opísané v Taliga (2009), stoja v ceste *ľubovoľnej teórii epistemického zdôvodnenia* (označme ju písmenom „L“), pretože súvisia so zisťovaním adekvátnosti, t. j. *pravdivosti*, L'. Predstavme si, že máme teóriu L' a že chceme zistiť, či je pravdivá. Možno za týmto účelom vykonáme vedecký experiment alebo budeme argumentovať pomocou abstraktnejšej úvahy, v každom prípade sa však musíme odvolať na nejaký jazykový výraz T, či je ním už jednoduché tvrdenie (napr. v podobe záznamov o výsledkoch vykonaného experimentu) alebo komplex-

ná teória (napr. vedecká teória opisujúca kognitívne procesy). Ak má však odvolanie sa na T hovoriť niečo v prospech pravdivosti L' , musíme si vybrať *buď* cestu definitívneho zdôvodňovania, pri ktorom by pravdivosť L' vyplývala z pravdivosti T , *alebo* čiastočného zdôvodňovania, pri ktorom by T nededuktívnym spôsobom zvyšovalo stupeň pravdepodobnosti teórie L' . V oboch prípadoch budú naše ambície sklamané z objektívnych príčin opísaných v Taliga (2009).

Pre úplnosť dodám, že tu zostáva ešte jedna možnosť, ktorá sa však neteší príliš veľkej obľube: Môžeme sa *úplne vzdať* podmienky zdôvodnenia a tvrdiť, že teória L' nie je *ničím viac ako hypotézou* o tom, akým spôsobom nadobúdame poznanie. Kauzálnu podmienku alebo ľubovoľnú inú externalistickú podmienku Ex by sme teda prestali chápať ako podmienku zdôvodnenia a jej naturalizáciu by sme dovedli do dôsledkov: Tvrдили by sme, že k poznaniu *fakticky dochádza* vďaka nej, hoci by sme si boli vedomí toho, že hypotézu L' nie je možné zdôvodniť. Ak by L' bola vedeckou hypotézou testovateľnou pomocou experimentov, a ak by obstála v testoch, stále by sme *nemohli* tvrdiť, že bola *vďaka tomu* zdôvodnená. (Ak by sme to tvrdili, dostali by sme sa do nezávideniahodnej situácie: Okrem toho, že by sme si protirečili, by sme museli opäť čeliť problému indukcie.) Ak by sme sa však okrem podmienky zdôvodnenia vzdali aj tripartitnej definície poznania, mohli by sme s čistým svedomím tvrdiť, že L' je prípadom vedeckého poznania – falzifikovateľnej hypotézy. Kým by teda L' nebola vyvrátená, mohli by sme tvrdiť, že je poznatkom o tom, akým spôsobom nadobúdame poznatky, a že výsledkom tohto spôsobu poznávania je aj sama L' . Neprítomnosť protipríkladu k teórii L' , podobne ako neprítomnosť dobrých dôvodov v prospech jej pravdivosti, predsa nijakým spôsobom neohrozuje *pravdivosť* teórie L' . *Iba prítomnosť protipríkladu by pravdivosť teórie L' spochybnila.* Takto by mohol externalista čeliť otázke, ako sa dá overovať pravdivosť jeho teórie *bez toho*, aby o podmienke Ex tvrdil, že je podmienkou zdôvodnenia, a *bez toho*, aby sa venoval zdôvodňovaniu svojej teórie L' . Výsledkom je pozícia, ktorú som v Taliga (2010b) nazval *skepticizmom* a ktorú Nuhliček v (2013a, 140) nazýva „poznáním bez zdôvodnenia“. Diskusii o nej sa venuje nasledovná časť.

3. Neplodnosť zdôvodnenia a prínos skepticizmu

Po formálnej stránke sú argumenty proti externalistickým teóriám *zdôvodnenia*, ktoré som načrtnol vyššie, totožné s tými, ktoré uvádza Nuhliček

v (2013b): Externalizmus zlyháva na probléme indukcie, na chybe *petitio principii* (*begging the question*) a na probléme nekonečného regresu. To by však ani nemalo prekvapovať, keďže *skeptické argumenty* operujú s týmito problémami už od antiky (pozri napr. Lammenranta 2008). Máme tu teda dva obsahovo mierne odlišné spôsoby argumentácie proti externalizmu, ktoré sa stretávajú v rovnakom závere. Súhlasím s Nuhličkom, keď uzatvára, že „v rámci externalizmu principiálne nie je možné dôveryhodne identifikovať zdôvodnenie akýchkoľvek presvedčení“ (Nuhlíček 2013b, 47). Preto ma mierne prekvapilo, keď som sa v Nuhličkovom texte (2013a) dočítal, že moje chápanie zdôvodnenia považuje za „neplodné“, keďže ma údajne „ukracuje o rôzne iné chápania zdôvodnenia“, ktoré „môžu byť potenciálne zaujímavé a prínosné“ (Nuhlíček 2013a, 149). Po tom, čo som uviedol, je azda zrejmé, že moje chápanie zdôvodnenia je naopak až také široké, že zahŕňa *ľubovoľný typ epistemického zdôvodnenia*, t. j. zdôvodnenia, ktoré si nárokuje na to, že súvisí s pravdivosťou zdôvodňovaných presvedčení. Fakt, že to robí až na meta-úrovni, som pritom nepovažoval za podstatný. Teraz však vidím, že som mal *vopred* vysvetliť, akým spôsobom je v mojom chápaní zdôvodnenia zahrnuté internalistické aj externalistické epistemické zdôvodnenie. Mohol som tak predísť nedorozumeniu, že moje chápanie zdôvodnenia je „v jasnom nesúlade“ s internalistickými aj s externalistickými koncepciami (por. Nuhlíček 2013a, 147). Chápem, že (napríklad) kauzálna podmienka znie na prvý pohľad inak ako moje chápanie zdôvodňovania. Dúfam však, že sa mi podarilo objasniť, prečo sa problémy zdôvodňovania, ktoré som opísal v Taliga (2009), týkajú nielen tejto podmienky, ale aj *ľubovoľnej inej podmienky epistemického zdôvodnenia*. Navyše mám dojem, že moje a Nuhličkove názory nie sú až také odlišné, ako to vykresľuje Nuhlíček v (2013a). Napokon, sám Nuhlíček dospel v (2013b) k záveru, že *externalistické* teórie zdôvodnenia prínosnými nie sú. Ak si myslí, že sú prínosné *internalistické* teórie zdôvodnenie, stačí povedať v čom.

Neviem, či by Nuhlíček považoval za plodné aj zdôvodnenie, ktoré nesúvisí s pravdivosťou toho, čo je zdôvodňované. Jeho diskusia o internalistických teóriách v Nuhlíček (2013a, 147) budí tento dojem, no možno ide o dojem falošný. Môj názor som už uviedol: Keďže existujú vecné kritické argumenty proti možnosti získania definitívneho aj čiastočného zdôvodnenia, je neplodné vyžadovať zdôvodnenie, ktoré súvisí alebo chce súvisieť s pravdivosťou toho, čo je zdôvodňované. Avšak vyžadovanie zdôvodnenia, ktoré nesúvisí alebo nechce súvisieť s pravdivosťou toho, čo je zdôvodňované, je neplodné tobôž: Také zdôvodnenie neplní žiadny *epistemicky relevant-*

ný účel. Ak by však niekto predsa len trval na tom, že také zdôvodnenie plodné je, bolo by zaujímavé vidieť, ako by čelil otázke, či je relevantným zdrojom zdôvodňovania napríklad aj veštenie z krištáľovej gule (pozri aj Lammenranta 2008, časť 9). Pre túto chvíľu mám skôr dojem, že *neplodnou je sama požiadavka zdôvodnenia, nie moje chápanie zdôvodnenia*, ktoré som prevzal z Miller (1994, 3. kap.).

Odmietnutie zdôvodnenia vedie spolu s faktom, že existuje poznanie, k téze, že poznatky nemusia byť (a ani nie sú) zdôvodnené. Aj túto tézu považujem za plodnú, okrem iného preto, že vyjasňuje spor, či je skepticizmus paradoxný. Prejdem preto priamo k Nuhličkovej výčitke, resp. otázke, prečo zavádzajúco širim koncepciu „poznanie bez zdôvodnenia“ pod hlavičkou skepticizmu (Nuhlíček 2013a, 149). Zdá sa, že môj rozbor všeobecnej formy skeptických argumentov, inšpirovaný Hetheringtonom (1996), ktorý som uviedol v Taliga (2012, 422), bol príliš stručný. Keby som si dal viac záležať, azda by Nuhlíček netvrdil, že sa uspokojivo nevyjadrujem k nesúladu môjho chápania skepticizmu s tradičným chápaním skepticizmu (pozri Nuhlíček 2013a, 140-141), a že nevysvetľujem, prečo moje chápanie skepticizmu „bez okolokov spája[m] s výrazom ‚skepticismus‘, hoci ... [mu] očividne nezodpovedá ani skepticizmus, ktorý nájdeme v dejinách filozofie, ani skepticizmus, ktorý je v prevládajúcej miere predmetom súčasných epistemologických diskusií“ (Nuhlíček 2013a, 145). Zopakujem teda to, že tradičné a v súčasnosti prevládajúce chápanie skepticizmu považujem za zavádzajúce: Skeptik, ktorý argumentuje proti poznaniu *vymedzenému určitým spôsobom*,⁹ spochybňuje možnosť, že existuje poznanie *vymedzené týmto spôsobom*, nie možnosť, že existuje poznanie *vôbec*. Ak sa teda (tradične) skepticizmus interpretuje ako téza, podľa ktorej nič nevieme alebo ako argumentácia, „ktorej cieľom je spochybnenie poznania“ (Nuhlíček 2013a, pozn. 13), potom sa mi zdá, že sa ukracujeme o hlbšie pochopenie skepticizmu, pretože skeptik *nemá prečo tvrdiť*, že nič nevieme, a ani spochybňovať samu skutočnosť, že nejaké poznanie predsa len máme.

Napokon, sám Nuhlíček súhlasí s tým, že v skeptickej argumentácii je „predmetom spochybnenia ... *zdôvodnenie presvedčení*“ (Nuhlíček 2013a,

⁹ V Hetheringtonovej všeobecnej forme skeptických argumentov je toto vymedzenie obsiahnuté v úvodnej premise S1, ktorú následne skeptik spochybňuje. Pre ilustráciu si môžeme predstaviť, že v S1 sa poznanie definuje napríklad v termínoch externalistických kauzálnych teórií, ktoré podmienku zdôvodnenia naturalizujú, alebo v termínoch internalistických teórií, ktoré chápu podmienku zdôvodnenia v normatívnom zmysle.

pozn. 7). Vyplýva z toho spochybnenie poznania? Ako si všíma aj Nuhlíček, je tomu tak iba „za predpokladu tripartitnej definície poznania“ (Nuhlíček 2013a, pozn. 3). Musí však skeptik *akceptovať* tripartitnú definíciu? Nevie si predstaviť, prečo by to tak malo byť. To, že sa táto definícia predpokladá pri vymedzení poznania, ktoré skeptik kritizuje, ešte neznamená, že sa skeptik s ňou musí stotožniť. Naopak, už v Taliga (2010b) som tvrdil, že *pokiaľ chceme chápať rozdiel medzi skepticizmom a agnosticizmom*, nemali by sme skeptikovi podsúvať tézu, že nikto nič nevie a ani argumentačné stratégie spochybňujúce poznanie. Skeptické argumenty, podľa ktorých *nemôžeme mať zdôvodnenie v prospech pravdivosti určitého presvedčenia, že p, nehovoria predsa nič v neprospech pravdivosti tohto presvedčenia, ale „iba“ v neprospech tézy, že ono presvedčenie musí byť zdôvodnené, ak chce byť poznatkom*. Rozumiem teda, prečo skeptik popiera, že by presvedčenie, že *p*, mohlo byť poznatkom *v zmysle tripartitnej definície*. Nerozumiem však, prečo by sa skeptik mal s touto definíciou stotožniť, keďže jeho argumenty nehovoria nič proti pravdivosti presvedčenia, že *p*, ktoré tak (za predpokladu odmietnutia tripartitnej definície) stále môže byť poznatkom. Povedané v skratke: *To, na čom záleží, nie je zdôvodnenie presvedčenia, ale jeho pravdivosť*.

Prínos tohto chápania skepticizmu, ktoré som navrhol aj v Taliga (2010b), by sa mal teda hodnotiť vzhľadom na problémy, ktorým som sa v danej stati pokúšal čeliť. Ako som uviedol v Taliga (2010b, 695), išlo mi najmä o otázku, či sa musí skeptik nevyhnutne zaplietať do protirečení či paradoxov, alebo sa pretvarovať, ako sa to často v epistemologickej literatúre o ňom tvrdí. Súhlasím s Nuhlíčkom (2013a, pozn. 14), že *nie je nutné*, aby sa filozofi pokúšali zo skepticizmu urobiť logicky konzistentnú epistemologickú doktrínu. Ak však chcú odlíšiť skepticizmus od agnosticizmu, zdá sa mi, že to *nutné je*.¹⁰ V Taliga (2010b) mi teda pôvodne išlo o problém, či *preukazujeme dostatočné pochopenie skepticizmu, ak tvrdíme, že ide o vnútorne spornú epistemologickú doktrínu*. Nezaujímal ma otázka, či je skepticizmus logicky konzistentnou doktrínou z historického uhla pohľadu, ale otázka, či sú paradoxy pripisované skepticizmu *naozaj dôsledkom skeptických argumentov*. Som rád, že aj Nuhlíček a Szapuová odpovedajú na túto otázku záporne (por. Nuhlíček – Szapuová 2012, 156-157), ich odpovedi však prestávam rozumieť, ak skepticizmus *nepovažujú* za „logicky konzis-

¹⁰ Samozrejme, tento môj názor môže byť chybný a možno len nechápem, ako sa dá odlíšiť agnosticizmus od skepticizmu *a súčasne zachovať chápanie skepticizmu ako smeru spochybňujúceho poznanie*.

tentný ... smer“, ako to tvrdí Nuhlíček v (2013a, 145, pozn. 13). Existuje teda, podľa nich, nejaký skeptik, ktorý sa zaplieta do paradoxov opísaných napríklad v Démuth (2009)? V Taliga (2010b) som navrhol, že to nie je skeptik, kto sa zaplieta do paradoxov, ale agnostik, ktorý tvrdí, že nič nevieme. Dokonca ani paradox falibilizmu, ktorý Kvasz a Zelenák v práci Kvasz – Zelenák (2009) opisujú a pripisujú ho Popperovi, sa netýka skepticizmu, ktorý sa (podľa môjho čítania Poppera) v Popperovej filozofii nachádza: Skeptik môže s čistým svedomím akceptovať, že existujú tzv. nutné pravdy a súčasne tvrdiť, že nie sú zdôvodnené (pozri Taliga 2010b, 4. časť). To boli kľúčové problémy, ktorým som sa venoval v Taliga (2010b). Mrzí ma, že sa v priebehu našich diskusií čoraz viac od nich vzdávame.

4. Asymetria medzi zdôvodňovaním a kritikou

V krátkosti sa ešte dotknem rozdielov medzi mojimi a Nuhlíčkovými názormi, ktoré na prvý pohľad nevyzerajú až tak dôležito. Pokúsim sa však ukázať, že ide o rozdiely kľúčové. Ako som zdôrazňoval už v Taliga (2009, 362; resp. 2012, 419), deduktívne argumenty sa dajú využiť buď na obhajobu alebo na kritiku určitého tvrdenia p , avšak ani jedna z týchto aktivít nemá nič spoločné so zdôvodňovaním p . Ak teda ponúkam argumenty určené na obhajobu p (napríklad na obhajobu svojich tvrdení) alebo argumenty, ktoré slúžia na kritiku p (napríklad na kritiku tézy, podľa ktorej je zdôvodnenie potrebné na poznanie), neznamená to ešte, že uvádzam dobré dôvody za (obhajoba) alebo proti (kritika) p . Preto si ani nemyslím, že je namieste podozrenie „z protirečenia medzi tým, čo deklaruje[m], a tým, čo skutočne robí[m]“ (Nuhlíček 2013a, 140).

Pokiaľ som si vedomý, moja pozícia nie je vnútorne sporná, no možno sa mýlim. Jej dôsledkom je však chápanie argumentov, ktoré je nezlučiteľné s ďalším Nuhlíčkovým názorom, a to, že „argument vo filozofii môže vystupovať v dvojakej úlohe: 1. Ako dôkaz tvrdenia v zmysle logického vyplývania z premís; 2. ako zdôvodnenie tvrdenia v zmysle pripojenia epistemicky podporných tvrdení“ (Nuhlíček 2013a, pozn. 19). Keďže takéto chápanie argumentov odmietam, nikdy mi nenapadlo explicitne upozorniť na rozdiel medzi 1 a 2. Z vyššie uvedeného je asi zrejmé, prečo odmietam 2. Prečo však odmietam 1? Jednoducho preto, lebo sformulovať argument, ktorého záver *vyplýva* z premís, znamená *odvodit'* záver z premís, neznamená to však záver *dokázať*. Na to, aby išlo o dôkaz záveru, by museli byť (ako podotýka

napr. aj Cmorej 2001, 83) premisy *pravdivé*, čo znamená, že na to, aby sme mohli *vedieť*, že ide o dôkaz záveru, by sme museli *vedieť*, že premisy sú pravdivé. Okamžite tak narážame na problém nekonečného regresu alebo na chybu *petitio principii*, ktorým som sa venoval v Taliga (2009). Mojou úlohou teda nebolo nič iné ako „pripomenúť nám, že je rozdiel medzi tým, keď chceme byť o niečom presvedčení, a tým, keď máme dobré dôvody, aby sme o tom boli presvedčení“ (Laudan 1998, 1132). Keďže sa však zdá, že ide o upozornenie, ktoré budí z nejasných príčin pohoršenie, pripomením aj to, že neprítomnosť dobrých dôvodov v prospech pravdivosti tvrdenia *p* nehovorí *vôbec nič* proti pravdivosti *p*.

Nuhlíček sa pýta: „Ak Taliga *nepozná* podmienky poznania, tak čo mu umožňuje tak horlivo obhajovať názor, že podmienka zdôvodnenia medzi ne nepatrí?“ (Nuhlíček 2013a, 149) Je pravda, že v Taliga (2012, 423) som tvrdil, že *neviem*, ktoré podmienky *sú* nutné a dostatočné pre poznanie. Z toho však *nevyplýva*, že by som nemohol poznať tie podmienky, ktoré nutnými či dostatočnými pre poznanie *nie sú*. V Taliga (2009) som sa pokúsil predstrieť racionálne uvažovanie využívajúce argumenty proti zdôvodňovaniu, ktoré som stručne zhrnul aj v tomto článku a ktorých zásobu zaujímavým spôsobom rozširuje aj Nuhlíček v (2013b). Toto uvažovanie ma doviedlo k záveru, že podmienka zdôvodnenia nie je nutnou a ani dostatočnou podmienkou poznania. Opakujem, že tento záver nie je ničím zdôvodnený, čo však neznamená, že je nepravdivý. Na jeho vyvrátenie je potrebné predstrieť protiargument, ktorý som však v Nuhlíček (2013a) nepostrehol. Nuhlíček síce nasleduje tradíciu analytickej epistemológie a rozlišuje medzi internalistickým a externalistickým zdôvodnením, pokúsil som sa však ukázať, že toto rozlíšenie je skôr v súlade ako v rozpore s názorom, že zdôvodňovanie je neplodné. Podmienku zdôvodnenia nezachráni ani názor, že existuje zdôvodnenie, ktoré „nemá nič spoločné s pravdivosťou presvedčení“ (Nuhlíček 2013a, 147). Ak je našim cieľom *poznanie*, potom je predkladanie a identifikácia dôvodov, ktoré nesúvisia s pravdivosťou zdôvodňovaného presvedčenia, azda aj možné, lenže *zbytočné*. Predkladanie a identifikácia dôvodov, ktoré s pravdivosťou zdôvodňovaného presvedčenia súvisia, je však *nemožné*. Práve to tvrdí, aspoň podľa môjho názoru, skepticizmus. Zdá sa mi, že tento záver je v súlade so závermi Nuhlíčkových statí (2013a) a (2013b), no možno sa mýlim a Nuhlíček z nejakého dôvodu chce obhajovať podmienku zdôvodnenia ako nutnú alebo dostatočnú podmienku poznania.

Literatúra

- CMOREJ, P. (2001): Úvod do problematiky metodológie vied (IV). Dôkazy a argumenty. *Organon F* 8, č. 1, 79-90.
- DÉMUTH, A. (2009): *Poznanie, vedenie alebo interpretácia?* Pusté Úľany: Schola Philosophica.
- HEMPEL, C. G. (1965): Studies in the Logic of Confirmation. In: Hempel, C. G.: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 3-51.
- HETHERINGTON, S. C. (1996): *Knowledge Puzzles. An Introduction to Epistemology*. Boulder – Oxford: Westview Press.
- KVASZ, L. – ZELENÁK, E. (2009): A Problem for Popper's Fallibilism. In: Cohen, R. S. – Parusniková, Z. (eds.): *Rethinking Popper (Boston Studies in the Philosophy of Science, vol. 272)*. Boston: Springer, 71-81.
- LAMMENRANTA, M. (2008): The Pyrrhonian Problematic. In: Greco, J. (ed.): *The Oxford Handbook of Skepticism*. Oxford: Oxford University Press, 9-33.
- LAUDAN, L. (1998): A Confutation of Convergent Realism. In: Curd, M. – Cover, J. A. (eds.): *Philosophy of Science. The Central Issues*. New York: W. W. Norton & Co, 1114-1135.
- LEHRER, K. (1999): *Teória poznania*. Bratislava: Infopress.
- MILLER, D. W. (1994): *Critical Rationalism. A Restatement & Defence*. Chicago – La Salle: Open Court Publishing Company.
- NUHLÍČEK, M. (2013a): Čo je zdôvodnenie? (Odpoveď M. Taligovi). *Filozofia* 68, č. 2, 139-150.
- NUHLÍČEK, M. (2013b): Prečo externalistické kritériá zdôvodnenia principiálne zlyhávajú? *Organon F* 20, č. 1, 37-49.
- NUHLÍČEK, M. – SZAPUOVÁ, M. (2012): Poznámky ku skepticizmu alebo čo spochybňuje, a čo nespochybňuje skeptik. *Filozofia* 67, č. 2, 147-159.
- TALIGA, M. (2009): Nekonečný príbeh zdôvodňovania. *Filosofický časopis* 57, č. 3, 353-373.
- TALIGA, M. (2010a): Causality, Truth, and Reality. *Organon F* 17, č. 4, 488-507.
- TALIGA, M. (2010b): Paradox skepticizmu? *Filozofia* 65, č. 7, 695-705.
- TALIGA, M. (2012): Na čo sú dobré argumenty? (Odpoveď A. Démuthovi, M. Nuhlíčkoví a M. Szapuovej). *Filozofia* 67, č. 5, 417-425.

Elipsa a anafora

MARIE DUŽÍ

VSB-Technická universita Ostrava. Katedra informatiky FEI
17. listopadu 15. 708 33 Ostrava. Česká republika
marie.duzi@vsb.cz

Tento článek je kriticky polemickou reakcí na článek s názvem „Elipsa“, autora Miloše Kosterce, který vychází ve stejném čísle časopisu *Organon F*. Ačkoli je míněn jako kritická polemika, neznamená to, že bych považovala zmíněný článek o elipse za špatný či bezcenný. Právě naopak, Kostercův příspěvek k teorii sémantiky vět s pragmaticky neúplným významem považuji za velice cenný a inspirativní. Přesto cítím potřebu se k některým aspektům Kostercova přístupu vyjádřit, neboť jsou poněkud diskutabilní.

1. Shrnutí Kostercova přístupu

Autor se v článku zabývá sémantickým obsahem tzv. eliptických vět, tj. takových, kde je určitá část věty vypuštěna, neboť je zřejmá z kontextu. Autor nejprve vymezí, jakým typem elipsy se bude zabývat, neboť dle Nealeho lze rozeznat tři typy takového eliptického vypuštění, a to *konverzační* elipsa, *eliptická výpověď* a *větná* elipsa. Jelikož mu jde o logicko-sémantickou analýzu, nezabývá se pochopitelně pragmatickým aspektem konverzační elipsy. Na vhodných příkladech pak ukazuje rozdíl mezi eliptickou výpovědí a eliptickou větou. Zatímco eliptická výpověď může být gramaticky úplná věta, avšak elipticky byla vypuštěna určitá informace, která je zjevná z kontextu (např. že se nejedná o všechny lahve na světě, nýbrž pouze o ty na stole), eliptická věta je gramaticky neúplná, tedy neúplná na úrovni povrchové formy (např. „Jožo s Anetou“, kde je vypuštěno sloveso „seděl“).

Dále pak autor říká:

Na rozdiel od konverzačnej elipsy a eliptickej výpovede sa teda eliptické vety vyznačujú neúplnosťou na úrovni povrchovej formy. V tomto článku sa venujem eliptickým vetám a ich významu. (Kosterec 2013b, 357)

Poté, co takto vymezil rozsah svého zkoumání, přechází ke krátkému shrnutí tří nejznámějších přístupů k analýze eliptických vět, a to *syntaktické* teorie elipsy, a dvou sémantických teorií, tj. teorie *vlastnosti* a teorie *minimálního indexikalismu*. Co se týká jeho kritiky syntaktické teorie elipsy, nelze než souhlasit s autorovým kritickým odmítnutím tohoto přístupu. Jistě, to co bylo z věty vypuštěno (antecedent), není pouhý semivýraz, ale plnohodnotný výraz s určitým *významem*, který je nutno analyzovat a na základě tohoto významu provést žádoucí doplnění eliptické věty. Tedy vztah mezi antecedentem a vynechanou částí eliptické věty není dán syntaktickým, nýbrž sémantickým kritériem.

Přechází proto ve svém hodnocení ke dvěma sémantickým teoriím elipsy, a to teorii vlastnosti a teorii minimálního indexikalismu. Zde snad jenom malá připomínka k autorovým vyjádřením. Na závěr této kapitoly 2 říká:

Čo keby sme tvrdili, že antecedent a vynechané miesto v eliptickej vete nie sú syntakticky totožné, ale vzťahujú sa na tú istú vlastnosť? Dôsledkom by bolo to, že vzťah medzi antecedentom a vynechanou časťou eliptickej vety nebude daný syntaktickým, ale sémantickým kritériom. (Kosterec 2013b, 360)

Jistě je mu známo, že stejný význam antecedentu a vynechané části věty nemusí být dán pouze vztahem ke stejné *vlastnosti*. Navíc, v dalším textu používá termín „vlastnost“ poněkud volně, a to nejen pro vlastnosti, ale i pro binární vztahy, apod.

Kosterec zcela správně poukazuje na to, že syntaktická teorie elipsy nerozliší dva možné významy věty

„Martin videl svoju matku a aj Edo“

První možné doplnění bude na základě vlastnosti označené výrazem „vidět svou matku“ a druhé na základě vlastnosti „vidět Martinovu matku“.

Co se týká rozboru teorie vlastnosti, zde je toto rozlišení snadno proveditelné. Nesouhlasím pouze se způsobem, jak to Kosterec poněkud neob-

ratně vysvětluje na základě lambda kalkulu na začátku kapitoly 3.1.: „Rozdiel medzi oboma vlastnosťami spočíva v počte výskytov viazaných premenných v λ -výraze.“ (Kosterec 2013b, 361) Vlastnosť jistě nemá žádné výskyty proměnných, a to ani vázané ani volné. Pouze příslušné λ -termy, které danou vlastností označují, obsahují výskyty proměnných.

Horší však je to, že jeho kritika teorie vlastnosti se zdá být zcela neopodstatněná. Kosterec říká zhruba toto: Teorie vlastnosti přijímá podmínku totožnosti významu antecedentu a významu vynechané části eliptické věty. Totožnost významu je dána totožností vlastnosti, a ta je dána identitou λ -termů. (Pozn.: K určení totožnosti vlastností by stačila pouhá *ekvivalence* λ -termů, jejich identita pak zajišťuje totožnost významu.) Lambda termy jsou identické, pokud mají stejnou strukturu, mohou se lišit pouze pojmenováním λ -vázaných proměnných. Není sice jasné, zda je to takto formulováno v teorii vlastnosti, nebo zda je to Kostercův výklad, avšak v zásadě je toto kritérium jasné. Ovšem příklad, na kterém se snaží Kosterec ukázat, že vysvětlení elipsy podané v rámci teorie vlastnosti na základě totožnosti významu antecedentu a vynechané části eliptické věty, není dostatečně všeobecné, se mi jeví jako zcela nevhodný a nesprávný. Kosterec tvrdí, že věta

„Martin videl svoju matku a Edo verí, že aj Jana“

má kromě dvou významů, které lze v teorii vlastnosti zachytit (tj. „Martin videl svoju matku a Edo verí, že aj Jana videla svoju matku“ vs. „Martin videl Martinovu matku a Edo verí, že aj Jana videla Martinovu matku“) ještě třetí význam, který již na základě této sémantické totožnosti v teorii vlastnosti zachytit nelze, a to tento:

„*Martin videl svoju matku a Edo verí, že Jana videla Edovu matku.*“

Opravdu nerozumím tomu, jak by někdo mohl výše uvedenou větu takto chápat. Vždyť o Edově matce se v tomto kontextu vůbec nemluví! Anaforický odkaz „aj“ se vztahuje k vlastnosti, o které se mluvilo, a v úvahu zde přicházejí pouze ony dvě výše zmíněné vlastnosti, a to „vidět svou matku“ či „vidět Martinovu matku“.

Dále v kapitole 3.2 vysvětluje Kosterec teorii minimálního indexikalismu, ke které se pak zjevně přiklání i v návrhu svého vlastního řešení problému sémantiky eliptických vět. Opět jen stručně shrnu principy, ze kterých autor vychází. Teorie minimálního indexikalismu (TMI) sice rovněž připouští, že význam vynechané části eliptické věty se zakládá na významu

antecedentu, avšak eliptická věta obsahuje tzv. *implicitní proměnné*, které nejsou v eliptické větě explicitně vyjádřeny. Cituji:

Eliptická veta podľa TMI obsahuje implicitné syntaktické jednotky. To však má závažný dôsledok pre eliptickú vetu – nie je totiž kompletnou vetou, ale iba tzv. „vetným radikálom“. (Kosterec 2013b, 364)

Kosterec zde uvádí příklad takovéto implicitní proměnné, a to ve větě „Peter sedel s Máriou a Jožo s Anetou“. Druhá věta „Jožo s Anetou“ je jistě gramaticky neúplná, je to větný radikál, obsahuje implicitní proměnnou, která má označovat vztah, v jakém jsou Jožo s Anetou, tedy tuto proměnnou musíme doplnit: „Jožo R^2 s Anetou“. S tím nelze než souhlasit. Ovšem závěry, které z tohoto jednoduchého faktu Kosterec vyvozuje, se mi zdají být velice problematické. Jednak opět zcela nevhodně ztotožňuje vlastnosti s jejich zachycením pomocí λ -termů. Opět cituji:

Tento vzťah sémantických obsahov antecedentu a implicitnej premennej reprezentujúcej vynechanú časť eliptickej vety sa nápadne podobá kritériu sémantickej identity sémantických obsahov v TV. Je tu však zásadný rozdiel. V TV boli výskyty premenných neoddeliteľnou súčasťou λ -zápisov, ktoré zodpovedali sémantickým obsahom antecedentu a vynechanej časti eliptickej vety. TMI nepostuluje túto spätosť a uvažuje binárne relácie a individuové premenné ako sémanticky nezávislé objekty. (Kosterec 2013b, 366)

Myslím, že jsou zde směřovány různé věci dohromady. Výskyty proměnných jsou jistě neoddelitelnou součástí λ -termů, o tom snad nikdo nepochybuje. Jistě však teorie vlastnosti nepovažuje binární relace a výskyty proměnných v λ -termech za sémanticky závislé objekty! Jde prostě jen o to, zda ona eliptická věta je opravdu eliptická v tom smyslu, že se jedná o gramaticky neúplný větný radikál s *implicitními* proměnnými, nebo se jedná o větu s *explicitním anaforickým odkazem*. Ve druhém případě je povrchová forma této věty úplná, nejedná se zde o žádné *implicitní* proměnné, které by bylo nutno doplňovat.

Toto směřování dvou různých věcí, a to problému větného radikálu (či opravdové elipsy, chcete-li) obsahujícího *implicitní*, tedy nevyjádřené odkazy, s problémem vět s *explicitním* anaforickým odkazem, tedy gramaticky úplné věty, se pak táhne dále jako červená nit celým článkem a je *hlavním předmětem mé kritiky*. Jasně je to vidět na následujícím příkladě, na základě jehož analýzy ukazuje Kosterec „přednosti“ TMI. Vrací se opět k příkladu

věty „Martin videl svoju matku a aj Edo“. Říká, že „vlastnost vidět je binární“ (sic!), a tedy v této větě chybí dvě implicitní proměnné, a to R^2 a i , proto je nutno ji doplnit takto:

Martin videl svoju matku a aj Edo R^2 i .

Nyní máme tak říkajíc volné ruce k tomu, abychom doplňovali za tyto proměnné jak vztahy, o kterých se mluvilo, tak individua, o kterých se mluvilo.

Za prvé, nerozumím tomu, proč je nutno při tomto řešení zcela ignorovat *explicitní* proměnnou, která se ve větě nachází, a to „aj“. Tato proměnná (anaforický odkaz) evidentně odkazuje na *vlastnost* (tedy *unární vztah*), o kterém se mluvilo. V úvahu, jak již jsem řekla, přichází pouze dvě vlastnosti, a to „vidět svou matku“ či „vidět Martinovu matku“. Za druhé, a to je ještě horší, ona volnost, se kterou zde Kosterec přidal dvě implicitní proměnné a přitom zcela ignoroval *explicitní* proměnnou, která tam je, mu umožňuje dojít i k tomu údajně třetímu možnému významu věty „Martin videl svoju matku a Edo verí, že aj Jana“, při kterém větě vedlejší rozumíme tak, že Jana viděla Edovu matku. Jak jsem již uvedla, takovéto řešení nelze považovat za správné.

Vždyť věta „aj Edo“ či česky „Edo také“ může být použita v jiném kontextu. Např. takto: „Martin je chytrý a Edo také“. V tomto případě tedy budeme dle Kosterce opět ignorovat *explicitní* anaforický odkaz „také“ a doplníme implicitní proměnné – jaké a proč? Jaké je dle Kosterce kritérium, kdy je nutno tyto implicitní proměnné doplňovat? Dle mého názoru je to kritérium jasné. Je to nutné, jedná-li se o gramaticky neúplnou větu s neúplnou povrchovou formou, kde ona *explicitní* proměnná chybí, jako v případě „Jožo s Anetou“. Což ale není případ vět s *anaforickým* odkazem, nýbrž čistě eliptických vět.

Nyní ještě rozeberu závěr této kapitoly 3.2. Kosterec zde uvádí (cituji, zdůraznění kurzívou je mé):

Na otázky zo začiatku tejto kapitoly odpovedá TMI nasledovne:

Q: Ako rozumieme, aké časti boli vypustené z eliptickej vety?

A: Vieme to vďaka kontextu jej použitia, ktorého vplyv sa prejaví ohodnotením príslušných implicitných premenných.

Q: Aký je vzťah významov antecedentu a vynechanej časti eliptickej vety?

A: Hodnoty implicitných premenných sú zároveň sémantickým obsahom niektoorej predchádzajúcej časti textu alebo komunikácie.

Povedzme si ešte, prečo TMI neprijíma postulát totožnosti sémantických obsahov eliptickej a jej prislúchajúcej neeliptickej vety. *Eliptická*

veta podľa TMI obsahuje implicitné premenné a predstavuje vetný radikál. Neeliptická veta žiadne implicitné premenné neobsahuje. Hoci v konkrétnom použití vetného radikálu je jeho vyjadrený význam rovnaký ako význam prislúchajúcej neeliptickej vety, neplatí to vo všeobecnosti. Ten istý vetný radikál totiž môžeme použiť v inom kontexte, v ktorom bude vyjadrovať iný význam. To súhlasí s tvrdením TMI, že eliptická veta mimo kontext jej použitia predstavuje otvorenú propozičnú konštrukciu, ktorá samostatne nie je úplným návodom na priradenie pravdivostnej hodnoty nejakej propozícii. (Kosterec 2013b, 367-368)

Jelikož se Kosterec s těmito zásadami dále ve svém návrhu analýzy v podstatě ztotožňuje, musíme si je trochu rozebrat. Tak za prvé, pokud v uvedeném odstavci „otázky – odpovědi“ nahradíme všude termín „implicitní proměnná“ termínem „explicitní proměnná, tj. anaforický odkaz“, vidíme, že totéž platí pro věty s anaforickým odkazem. Avšak, za druhé, tyto věty dle dále zdůrazněné části Kostercova textu nelze považovat za eliptické. Čili pro věty s anaforickým odkazem neplatí to, že zde došlo k eliptickému vypuštění nějaké části z pohodlnosti natolik, že bez doplnění chybějící části není věta gramaticky úplná, je to pouhý větný radikál. (Což nevylučuje, že by se nemohl vyskytnout v jedné a téže větě obojí jev, jak eliptické vypuštění, bez kterého není věta gramaticky úplná, tak explicitní anaforický odkaz.) Není tedy důvod, aby se *význam věty s anaforickým odkazem zcela měnil v závislosti na kontextu*, jako je tomu dle Kosterce v případě vět eliptických. Jistě, význam věty s anaforickým odkazem je rovněž *otevřená propoziční konstrukce*, tedy je to pragmaticky neúplný význam, neboli neúplný návod na to, jak v kterémkoli stavu světa a v kterémkoli čase vyhodnotit pravdivostní podmínky. Ale tato otevřená konstrukce *není závislá na kontextu*.

Tedy významem věty „Edo také“ je v *libovolném* kontextu otevřená konstrukce s volnou proměnnou *také*, a to konstrukce, která *v*-konstruuje propozici v závislosti na ohodnocení proměnné *také*: λωλt [*také*_{wt}⁰Edo]. Je-li věta vyslovena v situaci, kdy mluvíme o tom, že Jožo je chytrý, ohodnotíme proměnnou *také* pragmaticky a přisoudíme jí jako hodnotu vlastnost být chytrý. Vyskytne-li se v lingvistickém kontextu „Jožo je chytrý a Edo také“, pak jako výsledek rezoluce anaforického odkazu „také“ dostaneme ohodnocení proměnné *také* vlastností být chytrý. Není zde vůbec žádný důvod pro to, abychom zcela měnili pragmaticky neúplný význam (tj. otevřenou konstrukci) větou vyjádřený a přidávali do ní dodatečné implicitní proměnné pro vztahy, vlastnosti či individua.

V případě anaforického odkazu tedy není žádný důvod k radikálnímu kontextualismu tak, jak jej navrhuje Kosterec. Otevřenou otázkou je to, zda je takovýto důvod v případě vět eliptických s neúplnou povrchovou formou. Nelze snad onu implicitní proměnnou doplnit jaksi jednou provždy, nezávisle na kontextu, a to pouze na základě povrchové formy či gramatiky? Jelikož Kosterec neuvádí žádná kritéria pro to, jak a jaké implicitní proměnné je opravdu nutno doplňovat, jeho článek na tuto otázku nedává v podstatě žádnou odpověď. O to však hůře, neboť je to další důvod k pochybnostem o správnosti jeho radikálně kontextualistického řešení.

2. Kostercova analýza eliptických vět v rámci TIL

Problémy, na které jsem upozornila výše, tj. způsob, jakým chápe Kosterec principy teorie *minimálního indexikalismu*, se promítají rovněž do analýzy v rámci TIL, kterou autor představuje v kapitole 3.3. Jedná se zejména o to, že přidává implicitní proměnné i tam, kde to není nutné, kde je již explicitní proměnná, tedy anaforický odkaz. Přitom tuto explicitní proměnnou ignoruje a nijak nespecifikuje kritéria, dle kterých bychom se měli řídit při rozhodování o tom, zda a jaké implicitní proměnné do věty vložit. Opět cituji:

Naša analýza elipsy je dvojstupňová. Najprv nahradíme eliptickú vetu anaforickou vetou. Tú ďalej nahradíme vetou obsahujúcou prístavok, ktorá čo najvernejšie zachytáva význam vety obsahujúcej anaforický výraz. Sémantickú analýzu aplikujeme na vetu s prístavkom. Naša analýza viet obsahujúcich anaforický výraz totiž predpokladá formuláciu prístavkových viet, ktorých sémantický obsah potom skúmame. Z tohto pohľadu je rozdiel medzi eliptickými vetami a vetami obsahujúcimi anaforický výraz iba v tom, že pri analýze eliptických viet najprv nájdeme im prislúchajúce anaforické vety. Elipsa je teda podľa našej analýzy v podstate komplikovanejšia forma anafory. (Kosterec 2013b, 368)

S tímto postojem by se dalo v principu naprosto souhlasit. Problém však tkví v tom, že Kosterec v podstatě považuje téměř všechny věty za eliptické, tedy i ty, kde je již explicitní anaforický odkaz obsažen, jejichž povrchová forma je úplná. Tento anaforický odkaz pak často ignoruje a spoléhá při analýze pouze na ony implicitní proměnné, které byly dodány v podstatě *ad hoc*.

Druhý problém, který v Kostercově analýze nacházím, je ten, že nerespektuje skutečnost, že věta se zájmenem jako „On je rozumný“ má prostě

pragmaticky neúplný význam, a to stejný, ve všech kontextech. Věnujme se nejprve tomuto druhému problému. Kosterec říká (zdůraznění kurzívou je moje):

Príkladom anaforického výrazu je výraz „On“ v nasledujúcom kontexte:

(10) Pavel je logik. On je rozumný.

Rozumieme, že výraz „On“ sa vzťahuje na Pavla. Zároveň je to iný spôsob odkazu na Pavla ako výraz „Pavel“. *Nesmieme si však zamieňať druhú vetu v kontexte (10) za vetu (11).*

(11) On je rozumný.

Zásadné rozdiely medzi týmito vetami spočívajú v ich sémantickom obsahu. Tento rozdiel ukážeme, keď druhú vetu v kontexte (10) obsahujúcu anaforický výraz nahradíme vetou

(12) On, teda Pavel, je rozumný.

Zároveň predpokladáme, že veta (12) neobsahuje vetu (11) ako samostatný podvýraz. Prístavok „teda Pavel“, ktorý sa dopĺňa k zámenu „On“, nahrádza vo vete (12) (resp., presnejšie, vo vete, ktorú dostaneme z (10), keď druhú vetu nahradíme vetou (12)) anaforickú väzbu, ktorá v (10) existuje medzi „On“ a „Pavel“. Ak budeme dôslední, musíme pripustiť, že kým druhá veta v (10) obsahovala anaforický výraz, tak veta (12) už anaforický výraz neobsahuje. Anafora sa totiž nahradila prístavkom. (Kosterec 2013b, 368-369)

Zde se náš přístup liší. Tento rozdíl sice není tak závažný, jako již zmíněné přidávání implicitních proměnných i tam, kde to není nutné, kterému se budu věnovat v zápětí, přesto i zde je dle mého názoru narušen princip kompozicionality a princip doslovné analýzy (Parmenidův princip), tak jak jim rozumíme a dodržujeme je v TIL.

Tak především, nemohu souhlasit s tím, že druhá věta v kontextu (10) není věta „On je rozumný“. To je popírání nosu mezi očima, nebo opravdu radikální kontextualismus. Prostě a jasně, druhá věta v souvětí „Pavel je logik. On je rozumný.“ je věta „On je rozumný“. Jejím sémantickým obsahem je *otevřená* konstrukce $\lambda\omega\lambda t$ [⁰Rozumný_{wt} On] s volnou proměnnou *On*. Nevidím žádný důvod, proč by tomu tak nebylo. Ovšem na druhé straně je pravda, že celé souvětí (10) má úplný význam, tedy jeho analýzou musí být ve výsledku *uzavřená* konstrukce. Toho dosáhneme vhodnou substitucí.

Dále pak nerozumím tomu, proč by neměla věta (12) obsahovat jako samostatný podvýraz větu „On je rozumný“. Vždyť tam ta věta je! A koneč-

ně, proč by neměla věta (12) obsahovat anaforický odkaz? Vždyť ten zde stále je, a to zájmeno „on“. Kosterec udělal pouze to, že tuto volnou proměnnou ihned ohodnocuje pomocí substituce entitou Pavel, kdežto já provádím substituci vně této věty.

Možná jde pouze o neobratné formulace. Přejdeme tedy k vlastní analýze souvětí (10). Kosterec navrhuje tuto konstrukci (opravuji přitom zjevný překlep):

$$(12T) [\lambda\omega\lambda t [{}^0\text{Rozumný}_{wt} {}^2[{}^0\text{SUB} [{}^0\text{Tr } {}^0\text{Pavel}] {}^0\text{On } {}^0\text{On}]]]$$

Já navrhuji konstrukci

$$(\text{Duží}) {}^2[{}^0\text{SUB} [{}^0\text{Tr } {}^0\text{Pavel}] {}^0\text{On } {}^0[\lambda\omega\lambda t [{}^0\text{Rozumný}_{wt} \text{On}]]]$$

Zjevně jsou obě analýzy ekvivalentní. Výhodou mé analýzy, jak jsem již zmínila, je to, že je striktně antikontextualistická. Věta „On je rozumný“ má stále stejný, pragmaticky neúplný význam, a tím je otevřená konstrukce $[\lambda\omega\lambda t [{}^0\text{Rozumný}_{wt} \text{On}]]$. Pouze ohodnocení proměnné *On* se provádí na základě kontextu, což může být jak kontext pragmatický, čili situace promluvy, tak kontext lingvistický, tj. universum diskursu.

Kostercova analýza tento princip antikontextualismu (pravda, v tomto případě téměř „nevinně“) narušuje a Kosterec ji obhájí tím, že v *tomto kontextu* čteme větu „On je rozumný“ jako „On, teda Pavel, je rozumný“. Nevidím sebemenšího důvodu, proč bychom jí zde neměli rozumět tak, jak jí rozumím já, a to „On je rozumný. Kdo on? Aha, přeci Pavel“.

V tomto případě je však opravdu diskutabilní, který způsob analýzy volit, zda (12T) nebo (Duží). Argumenty pro a proti se zdají být poměrně slabé, a jakoby to v tomto případě byla pouze věc vkusu. Přejdeme proto k závažnějšímu problému, který provází Kostercovu analýzu, a to je ono již několikrát zmíněné přidávání implicitních proměnných i tam, kde se vyskytuje explicitní proměnná (anaforický odkaz), a kde to tedy není vůbec potřeba.

Markantně je to vidět na příkladu analýzy věty

„Martin videl svoju matku a Edo verí, že aj Jana“

Při analýze poslední klauzule Kosterec v podstatě narušuje Parmenidův princip, na který se sám odvolává. Ignoruje *explicitní* proměnnou (anaforický odkaz „aj“) a přidává zde *implicitní* proměnné pro binární vztah *R* a dvě proměnné pro individua *S* a *x*. Jenom tak může dojít k závěru, že tuto větu pak lze číst rovněž tak, jak navrhuje, tedy

„Martin videl svoju matku a Edo verí, že aj Jana videla Edovu matku“

Toto čtení je evidentně nesprávné, protože Edova matka nebyla vůbec větou zmíněna, je tedy narušen princip doslovné analýzy a Parmenidův princip. Zastavme se však ještě na chvíli nad Kostercovými analýzami údajného trojího čtení věty „Martin videl svoju matku a Edo verí, že aj Jana“. Kosterec zde ignoruje explicitní proměnnou „aj“ a přidává tři implicitní proměnné R , S a x , aniž by jakkoli zdůvodnil, proč takto narušuje smysl věty „aj Jana“. Kosterec říká „Premenná R zodpovedá výrazu ‚v tom istom vzťahu ako‘. Premenná S zodpovedá výrazu ‚niekým‘ a premenná x zodpovedá výrazu ‚niekoho‘.“ (Kosterec 2013b, 372)

Toto vysvětlení je velice nejasné a problematické. Věta „aj Jana“ nezmiňuje žádný „ten istý vzťah“, a neobsahuje výrazy „niekým“ a „niekoho“. Navíc, Kosterec neuvádí typy vložených implicitních proměnných, což velice znesnadňuje pochopení jeho analýz. Vždyť typy jsou neoddělitelnou součástí konstrukcí TIL. Zkusme je tedy doplnit: $R/*_n \rightarrow_v (\delta\mu)_{\tau\omega}$, S , $x/*_n \rightarrow_v \iota$. Ještě jedna připomínka, a to závažná. Kosterec analyzuje výraz „verí“ jako *hyperintensionální vztah* ke konstrukci, což rovněž nevysvětluje. Přitom je evidentní, že kdyby toto vysvětlil, bylo by zřejmé, že výsledná analýza může být jednodušší, neboť pak není nutné Dvojí provedení předcházeno Trivializací. Výsledkem aplikace funkce *Sub* je přímo ta konstrukce, ke které je Edo vztažen. Kosterec pak navrhuje *trojí* možné čtení této věty a analyzuje je postupně takto (Kosterec 2013b, 374-375):

(8a) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Martinovou.

(8Ta) $[\lambda\omega\lambda t \ [^0 a \ [^0 Videl_{wt} \ ^0 Martin \ [^0 Matka_{wt} \ ^0 Martin]] \ [^0 Verí_{wt} \ ^0 Edo \ ^0 [^2 [^0 SUB \ [^0 Tr \ ^0 Videl_{wt}]] \ ^0 R \ ^0 R] \ ^0 Jana \ [^2 [^0 SUB \ [^0 Tr \ ^0 Matka_{wt}]] \ ^0 S \ ^0 S] \ ^2 [^0 SUB \ [^0 Tr \ ^0 Martin] \ ^0 x \ ^0 x]] \] \] \]]$

(8b) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Janinou.

(8Tb) $[\lambda\omega\lambda t \ [^0 a \ [^0 Videl_{wt} \ ^0 Martin \ [^0 Matka_{wt} \ ^0 Martin]] \ [^0 Verí_{wt} \ ^0 Edo \ ^0 [^2 [^0 SUB \ [^0 Tr \ ^0 Videl_{wt}]] \ ^0 R \ ^0 R] \ ^0 Jana \ [^2 [^0 SUB \ [^0 Tr \ ^0 Matka_{wt}]] \ ^0 S \ ^0 S] \ ^2 [^0 SUB \ [^0 Tr \ ^0 Jana] \ ^0 x \ ^0 x]] \] \] \]]$

(8c) Martin videl svoju matku a Edo verí, že Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Edovou.

$$(8Tc) [\lambda\omega\lambda t [{}^0 a [{}^0 Videl_{wt} {}^0 Martin [{}^0 Matka_{wt} {}^0 Martin]] [{}^0 Veri_{wt} {}^0 Edo \\ {}^0 [{}^2 [{}^0 SUB [{}^0 Tr {}^0 Videl_{wt}] {}^0 R {}^0 R] {}^0 Jana [{}^2 [{}^0 SUB [{}^0 Tr {}^0 Matka_{wt}] {}^0 S \\ {}^0 S] {}^2 [{}^0 SUB [{}^0 Tr {}^0 Edo] {}^0 x {}^0 x]]]]]]$$

Všimněme si, že je zde ještě jeden problém kromě těch, které jsem již zmínila. Ve všech případech analyzuje Kosterec větu „Martin videl svoju matku“ konstrukcí

$$\lambda\omega\lambda t [{}^0 Videl_{wt} {}^0 Martin [{}^0 Matka_{wt} {}^0 Martin]]$$

To je evidentně nesprávné, neboť je opět ignorován odkaz „svoju“. V češtině a slovenštině máme zvrtná zájmena tedy na rozdíl od angličtiny je zde zcela zřejmé, že správnější čtení je ono „sloppy“, tedy to, které aplikuje vlastnost „vidět svou matku“. Vlastnosti „vidět svou matku“ a „vidět matku Martina“ jsou evidentně různé, a ignorování tohoto rozdílu zkresluje význam první věty podstatným způsobem, jak jsme ukázali v článku Duží – Jespersen (2013). Pro ilustraci, porovnejme Kostercovy analýzy s analýzou dle mé metody. Především tedy typ vztahu $Veri^*$ (ostatní typy zřejmé): $Veri^*/(oi^*_n)_{\tau\omega}$: hyperintensionální, tj. explicitní postoj.

- a) „**sloppy**“ čtení (s vlastností $P^{Svou}/(ou)_{\tau\omega}$: „vidět svou matku“).
 $P^{Svou} = \lambda\omega\lambda t \lambda x [{}^0 Videl_{wt} x {}^2 [{}^0 Sub [{}^0 Tr x] {}^0 him {}^0 [{}^0 Matka_{of_{wt}} him]]]$
 „Martin videl svoju matku“: $\lambda\omega\lambda t [P^{Svou}_{wt} {}^0 Martin]$
 „Edo verí, že aj Jana“: $\lambda\omega\lambda t [{}^0 Veri^*_{wt} {}^0 Edo {}^0 [\lambda\omega\lambda t [aj_{wt} {}^0 Jana]]]$
 „Martin videl svoju matku a Edo verí, že aj Jana (videla svoju matku)“:
 $\lambda\omega\lambda t [[P^{Svou}_{wt} {}^0 Martin] \wedge [{}^0 Veri^*_{wt} {}^0 Edo [{}^0 Sub {}^0 P^{Svou} {}^0 aj {}^0 [aj_{wt} {}^0 Jana]]]]]$
- b) „**strict**“ čtení (s vlastností $P^{Mart}/(ou)_{\tau\omega}$: „vidět Martinovu matku“).
 $P^{Mart} = \lambda\omega\lambda t \lambda x [{}^0 Videl_{wt} x [{}^0 Matka_{of_{wt}} {}^0 Martin]]]$
 „Martin videl Martinovu matku“: $\lambda\omega\lambda t [P^{Mart}_{wt} {}^0 Martin]$
 „Edo verí, že aj Jana“: $\lambda\omega\lambda t [{}^0 Veri^*_{wt} {}^0 Edo {}^0 [\lambda\omega\lambda t [aj_{wt} {}^0 Jana]]]$
 „Martin videl Martinovu matku a Edo verí, že aj Jana (videla Martinovu matku)“:
 $\lambda\omega\lambda t [[P^{Mart}_{wt} {}^0 Martin] \wedge [{}^0 Veri^*_{wt} {}^0 Edo [{}^0 Sub {}^0 P^{Mart} {}^0 aj {}^0 [aj_{wt} {}^0 Jana]]]]]$

Myslím, že tato analýza je mnohem přehlednější, což však není argument, zajisté. Ovšem tato analýza nekládá *ad hoc* větné přístavky za *ad hoc* přidané implicitní proměnné, neignoruje explicitní proměnnou *aj* (význam anaforic-

kého odkazu), tedy nenarušuje ani žádoucí princip kompozicionality ani s tím úzce souvisící Pramenidův princip.

Navíc, a to bych chtěla zdůraznit, zvolíme-li analýzu postoje *Veri* jako *explicitní hyperpropoziční* postoj, tj. vztah k *významu* vložené věty tedy ne k propozici, kterou věta označuje, pak musíme být obzvláště opatrní. Taková analýza musí *zcela respektovat perspektivu* toho (v našem případě Edy), komu je takovýto explicitní postoj připisován „zvenku“, z perspektivy mluvčího. Jak může Kosterec vědět, že Eda věří, tedy má vztah k *významu* jím tak radikálně pozměněné věty, kde místo „aj Jana“ stojí dle Kosterce „Jana je v tom istom vzťahu, teda vzťahu vidieť, s niekým, teda matkou niekoho, teda Martinovou / Janinou / Edovou“? To je opravdu *nesprávná analýza explicitního hyperpropozičního postoje*.

Kosterec nejen že ignoruje *explicitní* proměnnou „aj“, ale i další explicitní odkaz, a to „svoju“, který se vyskytuje v první klauzuli, kde „natvrdo“ analyzuje ono „svoju“ jako *Martinovu*. Zavádí sice implicitní proměnnou *S*, o které tvrdí, že stojí jako význam „někoho“, ale ignoruje fakt, že tedy by musel tutéž proměnnou použít i v první klauzuli „Martin videl svoju matku“ (naše analýza zde ponechává pro větší názornost proměnnou *him*, protože v češtině a slovenštině je evidentně správnější „sloppy“ čtení). Ale to je zřejmě již důsledek toho, co jsem několikrát v této krátké polemice zdůraznila, že Kosterec přidává implicitní proměnné i tam, kde toho není zapotřebí, a navíc neposkytuje žádné kritérium pro to, kdy a jaké proměnné a jim odpovídající „větné přístavky“ přidávat. V důsledku toho se jeho analýza jeví jako kontextualistická, *ad hoc* a v některých případech je bohužel zcela *nesprávná*, neadekvátní.

Všimněme si ještě Kostercovy analýzy věty „Watson podozrieval všetkých, ktorých Sherlock Holmes“. Zde je jeho postup použit zcela *správně*. Jistě, klauzule „ktorých Sherlock Holmes“ je větný radikál s neúplnou povrchovou formou. Evidentně zde chybí vztah, který má spojit daná *x* (kterých) se Sherlockem Holmesem, proto zcela správně doplníme implicitní proměnnou, jejímž oborem proměnnosti jsou binární vztahy, tedy proměnnou $R/*_n \rightarrow_v (ou)_{\tau\omega}$. (Jak jsem již uvedla, takovéto důsledné uvádění typu vkládaných implicitních proměnných však v Kostercově analýze bohužel chybí.) Na základě daného kontextu pak druhou klauzuli upravíme substitucí, tedy za proměnnou *R* substituujeme konstrukci binárního vztahu *Podzrívat*. Diskutabilní je pak již pouze to, zda provést substituci uvnitř předzpracovávané otevřené konstrukce, nebo vně.

3. Závěr

V této polemice jsem přednesla kritické poznámky ke Kostercově článku *Elipsa*, který vychází v tomto čísle časopisu *Organon F*. Moje námitky lze shrnout takto.

Za prvé, Kosterec si zřejmě neuvědomuje, že věta, která obsahuje explicitní indexické zájmeno, má svůj význam, i když pragmaticky neúplný. Je to *otevřená* konstrukce, ve které ono indexické zájmeno vyjadřuje volnou proměnnou. Valuace této proměnné se pak provádí na základě kontextu, ať již lingvistického či situací promluvy. Kosterec postupuje tak, že v lingvistickém diskursu tuto volnou proměnnou upravuje substitucí *uvnitř* oné otevřené konstrukce, kdežto moje analýza provádí substituci *vně* této konstrukce tak, aby význam věty (ona otevřená konstrukce) zůstal zachován nezávisle na kontextu. V tomto případě je jistě diskutabilní, který postup je správnější, a je to opravdu spíše „kosmetická“ záležitost, neboť oba postupy jsou zcela ekvivalentní. Nakonec, i když provedeme substituci uvnitř konstrukce, lze vždy snadno původní otevřenou konstrukci rekonstruovat.

Za závažnější však považuji to, že Kostercova analýza není obecně správná v případě vět s explicitním anaforickým odkazem. Kosterec tyto věty analyzuje stejným způsobem jako čistě eliptické věty, jejichž povrchová forma není úplná, jsou to tedy gramaticky nesprávné „větné radikály“. V případě gramaticky neúplných větných radikálů je v první fázi analýzy *nutno* doplnit implicitní proměnné tak, aby se gramatická forma zúplnila. Teprve poté můžeme ve druhé fázi analýzy za tyto implicitní proměnné substituovat, podobně jako to provádíme v případě vět s explicitním anaforickým odkazem, jejichž povrchová forma je úplná. Kosterec však tyto věty s explicitním anaforickým odkazem analyzuje tak, že explicitní anaforický odkaz (proměnnou) často ignoruje a dosazuje implicitní proměnné, jejichž volba se zdá být zcela *ad hoc*. Důsledkem pak může být neadekvátní analýza, která nezachycuje správně smysl věty.

Kostercův postup je správný a vhodný v případě eliptických větných radikálů, a to považuji za jeho *cenný příspěvek* k aplikacím TIL, neboť větnými radikály, tj. čistě eliptickými větami s neúplnou povrchovou formou, se u nás dosud v rámci TIL nikdo nezabýval. Je jen škoda, že Kosterec neuvádí žádné vodítko pro to, kdy a které implicitní proměnné je nutno doplnit. Na otázku *kdy* bych odpověděla tak, že *vždy* a *pouze* tehdy, kdy se jedná o větný radikál s neúplnou povrchovou formou. Na otázku, které proměnné je nutno doplňovat, je již těžší odpověď. Zřejmě bychom se měli řídit typovou

analýzou. Zmiňuje-li např. eliptický větný radikál pouze dvě individua, pak zřejmě chybí aplikace binárního vztahu, která by je spojovala. Implicitní proměnná, kterou je nutno doplnit, je pak zřejmě proměnná $R \rightarrow_v (ou)_{\tau\omega}$, což se zdá být nejčastější případ. Jistě to však není jediná možnost a bylo by velice cenné ověřit Kostercův postup na více příkladech eliptických větných radikálů tak, aby bylo možno formulovat kritérium pro doplňování implicitních proměnných příslušných typů.

Poslední poznámka. Nedávno vyšel další Kostercův příspěvek na podobné téma s názvem „*Anaforický retáзец*“ (Kosterec 2013a). Zde uvádí autor další námitku proti způsobu analýzy anaforických odkazů tak, jak jsem jej navrhla, a to je případ řetězení anaforických odkazů. Kosterec tvrdí, že moje metoda a zejména její *implementace* pomocí „programátorských“ proměnných *referentů k diskursu* nedokáže tyto řetězce správně zachytit. Zde se jedná o zřejmé nepochopení metody. Anaforické řetězení lze samozřejmě v této metodě správně analyzovat. Funkce *Sub* operuje na konstrukcích, lze ji tedy snadno iterovat. Vhodnou iterací aplikace funkce *Sub*, tj. postupnými Kompozicemi, pak snadno ony řetězce analyzujeme. Rovněž implementační metoda dynamického diskursu toto dokáže správně zohlednit. Jelikož však problém anaforického řetězení se již týká jiného článku, tedy ne toho, ke kterému se vztahuje tato polemická diskuse, ponechám rozbor anaforických řetězců napříště.¹

Literatura

- DUŽÍ, M. – JESPERSEN, B. (2013): Procedural Isomorphism, Analytic Information, and β -Conversion by Value. *Logic Journal of the IGPL* 21, No. 2, 291-308. doi: 10.1093/jigpal/jzs044.
- KOSTEREC, M. (2013a): Anaforický retáзец. *Organon F* 20, mimoriadne číslo 2, 140-159.
- KOSTEREC, M. (2013b): Elipsa. *Organon F* 20, č. 3, 355-376.

¹ Tento výzkum je podporován interní agenturou VSB-TU Ostrava, projekt č. SP2013/207 „Využití umělé inteligence při získávání znalostí o procesech, jejich modelování a dolování“.

Peter Kosso: *A Summary of Scientific Method*

Dordrecht – Heidelberg – London – New York: Springer 2011, 56 strán

Napriek tomu, že žijeme vo vedecko-technickom veku, neustále sa stretávame s nepochopením vedeckej práce a s neschopnosťou jasne a prístupne formulovať odpovede na otázky zásadnej dôležitosti, aké si kladie aj autor recenzovanej knihy už v predhovore: „Čo je veda? Čo je vedecká metóda? V čom je veda vedecká?“ (s. vi). Vzhľadom na autorov dar vyjadrovať náročné myšlienky jednoduchým jazykom s použitím nenáročných ilustrácií z každodenného života sa mu podľa nás podarilo na tieto otázky odpovedať spôsobom, ktorý čitateľ ocení.

Práca *A Summary of Scientific Method (Prehľad vedeckej metódy)* je krásnym opisom toho, čo robí vedu vedeckou. Samozrejme, je na čitateľovi, aby posúdil, do akej miery korešpondujú jeho názory so skutočným výkonom vedcov a s tým, čo robia, keď robia vedu. Kniha je však napísaná jasným a čistým štýlom, ktorý je prístupný a informatívny pre vedcov a študentov vedy, ako aj pre filozofov. Táto rozsahom skromná práca je dôkazom možnosti vyjadriť mnoho na malom priestore. Má 6 zrozumiteľných kapitol, z ktorých každá objasňuje jeden z aspektov vedeckej metódy. Každá kapitola sa začína abstraktom a kľúčovými slovami, čo uľahčuje prácu s textom. Na záver autor ešte raz zhrňa výsledky a uvádza zásadné práce ako odporúčanú literatúru ku každej kapitole.

V prvej kapitole (*Science and Common Sense*) nás Peter Kosso oboznamuje s myšlienkou spektra medzi istotou a dohadom, ktorá podľa neho zastáva významnú rolu vo vedeckej metodike. Obvykle veda pracuje s entitami alebo procesmi, ktoré nie sú priamo pozorovateľné, dôkaz je teda nepriamy a teoretické výsledky nie sú isté na 100%. To však neznamená, že teórie sú číre dohady. Dôležité je práve spektrum medzi istotou a dohadom. Úlohou vedeckej metódy je lokalizovať konkrétnu teóriu v tomto spektre. Po zhromaždení dôkazov môže teória zmeniť svoje postavenie v spektre a buď sa priblížiť k istote, alebo naopak k dohadu. Pre vedeckú metódu je teda dôležitý stupeň dôkazu a efektívne použitie evidencie, ako uvidíme neskôr. Ak teda existujú stupne či úrovne dôkazu a spektrum medzi dohadom a istotou, potom musia byť dôležité detaily a nuansy vedeckej metódy, ktoré majú vplyv na stupeň a lokalizáciu teórie v spektre. Pochopenie týchto stupňov či úrovní dôkazu je podľa autora najdôležitejší a najnáročnejší aspekt porozumenia vedeckej metóde.

Druhá kapitola (*Empirical Foundations*) načrtáva, v akom vzájomnom vzťahu sa voči sebe nachádzajú pozorovanie a teória. Na začiatku kapitoly sa definu-

jú pojmy teórie, hypotézy a zákona, ktoré sú nevyhnutné na pochopenie vedeckej metódy. Autor zastáva názor, že teória opisuje aspekty prírody, ktoré nie sú dostupné našej zmyslovej skúsenosti, nemožno ich teda pozorovať priamo. Ide pritom o aspekty, ktoré sa dajú použiť na vysvetlenie pozorovaného. Súčasná veda pracuje s entitami, ako sú subatomárne častice, zakrivený časopriestor či elementárne struny, a práve tie sú základom teórií, ktoré dnes veda rozvíja napriek nemožnosti ich priameho pozorovania. Z tohoto dôvodu sú podľa autora teoretické, nie však nereálne. Teória sa ukazuje byť pravdivá, ak presne opisuje nepozorovateľné, ktoré skutočne existuje, a zhoduje sa teda s faktom, ktorý je skutočným stavom vecí v prírode. „Fakty sú; teórie opisujú“ (s. 8).

Za najvšeobecnejšie teórie autor považuje vedecké zákony. „Inými slovami, zákony sú teórie zvláštného druhu, ktoré identifikujú celé kategórie vecí a opisujú ich vzťahy v najvšeobecnejších pojmoch“ (s. 8).

Hypotézy podľa autora predstavujú málo testované teórie, ktorých spojenie s faktom nie je známe alebo je nejasné. Existuje však dôvod veriť, že nové dôkazy pomôžu vytvoriť toto premostenie a posunú hypotézu v spektre bližšie k istote. Kľúčová otázka znie, ako sa to deje. Samé induktívne zovšeobecňovanie pozorovaní neimplikuje tvrdenia o nepozorovanom či nepozorovateľnom, a tak vedecká metóda sa nemôže redukovať len na indukciu, ale musí byť čímsi viac, keďže veda prekračuje pozorovateľné. Z čistého pozorovania by sme nikdy nemohli dospieť k zakrivenému časopriestoru či elementárnym strunám. Ďalším dôvodom nedostatočnosti indukcie je predpoklad čistého, ničím nekontaminovaného toku informácií z vonkajšieho sveta k nám. Akékoľvek vedecké pozorovania či experimenty sú ovplyvnené vedeckými teóriami. Einstein to jasne vyjadril už v roku 1926 v rozhovore s Heisenbergom: „Teória rozhoduje o tom, čo sa dá pozorovať“ (Heisenberg 1997, 78).

Autor uvádza štyri dôvody, na ktorých demonštruje závislosť vedeckého pozorovania od teórie a teda nemožnosť jeho neutrality: 1. Nikdy nemožno pozorovať všetko. Potrebná je selekcia, ktorá nemôže byť náhodná. O tejto selekcii rozhoduje vždy teória, ktorá určuje, čo je a čo nie je relevantné. 2. Po uskutočnení pozorovania sa treba venovať výsledkom a zamerať úsilie na to podstatné, o čom opäť rozhoduje určitá znalosť, sprostredkovaná teóriou a vlastnou skúsenosťou vedca, umožňujúca mu niektoré aspekty ignorovať. 3. Podmienky, za ktorých sa pozorovanie uskutočňuje, musia byť kontrolované. Každý prístroj si vyžaduje určité teoretické porozumenie, ktoré umožňuje jeho správne nastavenie. 4. Namerané výsledky si vyžadujú interpretáciu, ktorá vždy predpokladá teoretický kontext. Bez neho by sme neboli schopní správne porozumieť bublinám v bublinovej komore a nemohli by sme ich interpretovať ako dráhu alfa častice, mali by sme pred sebou len oddelené bublinky. Vedecká metóda sa ukazuje ako dynamický proces, v ktorom informácie prúdia od sveta k nám a od

nás späť k svetu, od teórií k pozorovaniam a vice versa. Kladieme prírode otázky a príroda odpovedá, pričom jej odpovede nás vedú stále k novým otázkam.

Abstrakt tretej kapitoly (*Empirical Testing*) začína vetou: „Empirické testovanie vedeckých hypotéz je vždy nepriame“ (s. 13). Túto, pri prvom prečítaní prekvapivú myšlienku autor dôsledne dokazuje pomocou trhliny, ktorú nachádza v Popperovom falzifikačnom modeli. Podľa tohoto modelu predstavuje falzifikácia predikcie vyvrátenie hypotézy. Logická štruktúra skutočného empirického testovania však naznačuje opak. Z nepotvrdenej predikcie nemožno nevyhnutne usudzovať na falzifikáciu hypotézy. Jednoduchý hypoteticko-deduktívny (H-D) model testovania treba rozšíriť o experimentálne podmienky (C – *Experimental Conditions*) a pomocné teórie (A – *Auxiliary Theories*), predstavujúce spôsoby testovania hypotézy a zakladajúce sa na základných znalostiach, ktoré vedci získavajú počas dlhoročného štúdia. V jednoduchom H-D modeli vytvárame z hypotézy dedukciou predikcie, ktoré sú predmetom testovania. Platí teda: ak *H* (hypotéza), tak *p* (predikcia). V modeli rošírenom o (A) a (C) zvažujeme aj teoretické detaily spôsobu vytvárania predikcií a praktických detailov uskutočňovania experimentu, ktoré predstavujú zásadné prvky každého vedeckého testovania. V takom prípade *p* nadobúda podobu: „*p* = ak *C*₁ a *C*₂ a *C*₃ a ..., tak *E*“ (s. 18), kde *E* predstavuje očakávaný výsledok experimentu a H-D premisa sa mení na: „ak *H* a *A*₁ a *A*₂ a *A*₃ a ..., tak *p*“ (s. 19). Definitívna podoba H-D premisy má potom formu:

Ak (H a *A*₁ a *A*₂ a *A*₃ a ...), tak (ak *C*₁ a *C*₂ a *C*₃ a ..., tak *E*)

Úsudok zachytávajúci testovanie predikcie, ktorá sa ukáže nepravdivá, má formu:

Ak (H a <i>A</i> ₁ a <i>A</i> ₂ a <i>A</i> ₃ a ...), tak (ak <i>C</i> ₁ a <i>C</i> ₂ a <i>C</i> ₃ a ..., tak <i>E</i>)	(H-D premisa)
nie <i>E</i>	(empirická premisa)

nie (H a *A*₁ a *A*₂ a *A*₃ a... *C*₁ a *C*₂ a *C*₃ a ...)

Logická štruktúra tejto úsudkovej schémy ukazuje, že napriek nepravdivosti predikcie nemôžeme považovať hypotézu za definitívne vyvrátenú, keďže chyba mohla nastať kdekoľvek v dlhom reťazci pomocných teórií a experimentálnych podmienok. Musíme brať do úvahy aj zásadný rozdiel medzi experimentálnymi podmienkami a pomocnými teóriami, keďže prvé z nich môžeme pozorovať, a tak podrobiť kontrole ich splnenia, kým skutočnosti opísané pomocnými teóriami sú nepozorovateľné a ich pravdivosť môžeme overiť jedine nepriamo. Preto hypotézy nemôžeme testovať priamo, čo potvrdzuje vetu citovanú v úvode. Empirické testovanie teda nepredstavuje dostatočný nástroj na určenie pravdivosti hypotézy. Rozhodne však nie je zbytočné, práve naopak, umožňuje nám prehodnotiť všetky aspekty, ktoré mohli zlyhať a usilovať sa o ich nápravu. Pe-

ter Kosso v závere kapitoly výstižne konštatuje: „Vedecká metóda je založená na evidencii a logike, ale detaily empirického testovania ukazujú, že ani evidencia, ani logika samy neriešia otázku, ktoré teórie sú pravdepodobne pravdivé a ktoré nepravdivé. O čo iné teda ide?“ (s. 20)

Kapitola s názvom *Sieť poznania* (*The Network of Knowledge*) vysvetľuje autorovu základnú myšlienku o koherencii empirického a teoretického vo vedecskom poznaní. Bežne je známe, že vedecká metóda sa musí zakladať na experimentálnej činnosti aj teoretickej práci. Obidve musia byť v súlade a netolerujú sa žiadne rozpory. „Najzákladnejšia požiadavka súdržnosti je logická konzistencia. Sieť vedeckých poznatkov nemôže tolerovať rozpor... Rozpor nemožno ignorovať“ (s. 23).

Okrem konzistencie vedeckých tvrdení je nutná aj ich kooperácia. Nestačí, že sú vo vzájomnej zhode, zásadnými sú aj vzťahy, ktoré medzi sebou majú. Pomocou teoretických výpovedí vysvetľujeme naše pozorovania aj iné teoretické tvrdenia, ako je uvedené v predchádzajúcej kapitole, kde pomocné teórie zohrávali významnú úlohu pri získavaní evidencie inej, testovanej teórie.

Konexie medzi vedeckými ideami majú rozličnú povahu a charakter a sú nevyhnutným predpokladom fungovania vedy ako komplexnej štruktúry. O tejto skutočnosti svedčí aj fakt, na ktorý autor upozorňuje, že vedecké publikácie majú spravidla viac než jedného autora. Pokiaľ množstvo odborných znalostí konverguje k spoločnému záveru, disponujeme dobrým dôvodom, aby sme takémuto záveru dôverovali. Vedecký postup je dynamický proces medzi teóriami navzájom či teóriou a pozorovaním. Ak sa hypotéza ustáli v sieti „vedeckého poznania, máme stále lepší dôvod veriť, že je pravdivá. Toto je vedecká metóda“ (s. 24).

V predposlednej kapitole (*Scientific Change*) autor venuje pozornosť zmenám vo vede. Vo všeobecnosti možno konštatovať, že Kosso súhlasí s Kuhnovou koncepciou, predstavenou predovšetkým v jeho slávnej publikácii *Štruktúra vedeckých revolúcií*. Na druhej strane vplyv paradigmy podľa neho Kuhn preceňuje. Zdôraznili sme už to, že veda nie je statický, ale dynamický, neustále sa vyvíjajúci systém, čoho si jej aktéri plne uvedomujú. Aby bolo možné udržať progres vedy, musí jej metóda obsahovať mechanizmus umožňujúci zmenu spôsobom, ktorý zvyšuje pravdepodobnosť pravdivosti jej výsledkov a nie naopak. V dejinách vedy sa odohralo už niekoľko významných kvalitatívnych zmien, ktoré predstavovali zmenu jednej paradigmy na inú. To, čo Kuhn nazýva paradigmou, je „sieť základných znalostí, opisných kategórií, odborného jazyka, experimentálnych zásad atď...“ (s. 27). Bez toho veda nemôže existovať. Predpoklad stabilného základu je nevyhnutný už len z dôvodu vzájomnej komunikácie vedeckých pracovníkov a ich spolupráce. Pokiaľ by sme nehovorili rovnakým jazykom, disponovali odlišnými teoretickými znalosťami a skúsenosťou, nemohli by sme vedu vôbec pestovať. Pri pohľade z opačnej perspektívy „nemôžeme ignorovať neurčitosť súčasných vedec-

kých presvedčení“ (s. 30). Musíme mať vždy na pamäti, že naše poznanie nie je definitívne a byť neustále otvorení jeho novej revízii spôsobovanej permanentným nárastom rozsahu nášho poznania.

V záverečnej kapitole (*Scientific Understanding*) sa pozornosť venuje vedeckému chápaniu. Poznanie samé ešte nie je zárukou porozumenia faktom, ktorými disponujeme. Ak chceme prírode rozumieť, musia všetky jej teoreticky opísané časti vytvárať akúsi mozaiku. Kosso vychádza z porovnania vedy a umenia. Krása umeleckého diela totiž nespočíva v dokonalom prevedení jeho jednotlivých častí, ale v ich vzájomnej súhre. Harmonické zladenie častí v celok je kritériom krásy. Inak tomu nie je ani vo vede: „...vedecké porozumenie súvisí s krásou. Obe sa dosahujú iba v starostlivom zväžení vzťahov a súladom medzi časťami“ (s. 33). Napriek tomu, že s porozumením sa stretávame častejšie v spoločenských vedách, využívajúcich ako svoju metódu na tento účel hermeneutiku, konceptuálne dochádza k obdobnému procesu aj v prírodných vedách, problematizujúcich odlišné predmety či aspekty sveta. Každý jednotlivý symbol, myšlienka či akt má význam len v širšom kontexte, do ktorého ho situujeme. Osve sú všetky bez významu, pretože bez vzťahu k inému ich nemožno vymedziť ani definovať. „Porozumenie závisí od koherencie“ (s. 34). Znakom porozumenia je korešpondencia teoretických tvrdení zapadajúcich do siete utkanej z množstva predchádzajúcich poznatkov.

Ak to zhrnieme, Peter Kosso vníma vedu ako ucelenú, no flexibilnú sieť poznatkov, spoločenský fenomén a záležitosť rozsiahlej skupiny špeciálne teoreticky, ale aj prakticky vybavených jedincov, z ktorých každý prispieva svojím kúskom práce k celkovej výstavbe pýchy západnej civilizácie. Niektorí schopnosťou vytvárať hypotézy, iní ich testovaním a ďalší pochybovaním a následným uskutočňovaním potrebných revízií na spustenie opätovného procesu testovania. Takto stále dokola, aby sme boli zakaždým o krok bližšie k odpovediam na otázky, ktoré vyvstávajú s každou novou skúsenosťou. Veda neposkytuje definitívne poznanie, je procesom, ale dáva nám dobré dôvody, aby sme verili jej záverom. Myslíme si, že takýto postoj je vzhľadom na súčasný, ale aj minulý stav vedy akceptovateľný a vďaka štýlu, ako sa v práci artikuluje, predstavuje publikácia jeden z mála prístupných, no zároveň odborne fundovaných výkladov o metóde uplatňovanej vo vedeckej praxi pre nezainteresovaného čitateľa aj študenta ktorejkoľvek prírodnej vedy či filozofie (vedy). Publikáciu odporúčame najmä začínajúcim študentom doktorského štúdia so záujmom o teoretické aspekty vedy.

Ladislav Országh
orszagh.l@gmail.com

Literatúra

HEISENBERG, W. (1997): *Část a celek. Rozhovory o atomové fyzice*. Olomouc: Votobia.

PhiLang2013 – Third International Conference
on Philosophy of Language and Linguistics

In May, 9-11 2013, the University of Łódź hosted a conference on the relationship between philosophy of language and linguistics entitled *PhiLang2013 – Third International Conference on Philosophy of Language and Linguistics*. Organized by Piotr Stalmaszczyk, the conference put together philosophers and linguists from all around the world and so provided an excellent opportunity for them to discuss.

The conference commenced with the opening lecture given by Emma Borg (Reading). In her ‘Semantics Minimalism: What Does It Claim and Why?’, Borg went through the main semantic positions including minimalism, indexicalism, relativism, occasionalism or contextualism and, consequently, motivated the first one. A fantastic introduction to the conference.

The first plenary lecture was given by Gary Kemp (Glasgow). Kemp outlined and further elaborated the relationship between ‘Quine and Wittgenstein as regards Linguistics as a Science’ and their objections to ontological weirdness of philosophical theories. After that, participants split up into three (rather monothematic) sessions one of which, section B, started with Mark Pinder’s (Bristol) ‘Borg’s Minimalism, Modularity and the Problem of Paradox’. In it, Pinder challenged the claim that speakers cognize minimal semantic theory that yields truth conditions for their language’s sentences. Marián Zouhar (Bratislava) in his ‘On Underdetermination of Contextualism’ argued that arguments for contextualism could be used in non-contextualist’s theories too, leaving thus contextualism rather underdetermined. In ‘Some Remarks about Nonindexical Contextualism’, Tadeusz Ciecierski (Warsaw) outlined several arguments for nonindexical contextualism, namely evaluation arguments, operator arguments and epistemic arguments. Finally, Tim Pritchard’s (London) ‘Semantics and the Understanding of Artifact Words’ explored the difference between account of constitution of our basic understanding on one side, denotationally-focused account of word meaning on the other.

The second plenary lecture, called ‘Categorial Grammar and the Foundations of the Philosophy of Language’, was given by Mieszko Tałasiewicz (Warsaw). Tałasiewicz argued that it is a new subject-predicate distinction that could elucidate a lot of problems of philosophy of language. The section B

continued with 'Kripke, Putnam and the Description Theory' by Luis Fernández Moreno (Madrid). In it, Fernández Moreno discussed such a version of the description theory of reference as to substitute the resort to causal chains by a social alternative. 'Referential Thinking and Putnam's Brain in a Vat' by Roman Godlewski (Bydgoszcz) proposed to meet Putnam's argument by means of various referential considerations. Piotr Stalmaszczyk (Łódź) and Bartosz Stopel (Silesia) were concerned with literary theory. While '(Un)translability of Proper Names and (Im)Possible Worlds' of the former discussed an option to analyze literary texts using possible worlds apparatus, 'Philosophy of Language and Literary Theory: a Problematic Relation?' of the latter focused on several issues of meaning, interpretation and intention in literary studies.

The second day of the conference started with a co-authored plenary lecture given by Kasia M. Jaszczolt (Cambridge) and Chi-Hé Elder (Cambridge). Their 'Conditional Utterances and Conditional Thoughts: A Radical Contextualist Account' focused primarily on the relationship between antecedent and consequent of conditionals when it comes to direct as well as indirect conditionals. Siu-Fan Lee (Hong Kong) in her 'Who Wants To Be a Russellian about Names' as a first A-section speaker of the day, presented several arguments against Russell's project of names, while Gabriele M. Mras's (Vienna) 'Relations-Russell with Frege against Bradley' discussed Russell's stance about external and internal relations. Definite description played the crucial role also in Wojciech Rostworowski's (Warsaw) 'Free Logical Theory of Definite Descriptions' and Thomas J. Hughes's (Durham) 'The Origin of Referential Definite Descriptions'. The last speaker before lunch was Filip Kawczyński (Warsaw) with a paper called 'New Approaches towards Frege's Puzzle'. After lunch, Paul Livingston's plenary lecture (New Mexico) wondered into 'The Sense of Finitude and the Finitude of Sense' in which he quite interestingly scrutinized the relation between the finite and the infinite in language. After that lecture, Section A continued with 'What Can Be (precisely) Said about Perception' by Pawel Grabarczyk (Łódź) and Ali's Abasnezhad's (Lille) 'Precise Reference and Vague Referent'. Interestingly though, both papers went rather behind the philosophy of language and discussed its cognitional and metaphysical aspects, respectively. Alicja Chybińska (Warsaw) introduced (at least to some of us) 'The Concept of Truth and Falsehood in the Philosophy of Kazimierz Twardowski', Jiří Raclavský (Brno) presented 'Language in Synchronic/Diachronic Sense and some Puzzles in the Philosophy of Language' and, finally, Başak Aray told us something about 'Ogden and Analytic Philosophy'.

The last day of the conference commenced with the fifth plenary lecture called 'The Meaning of Formal Semantics'. In it, Chris Fox (Essex) compared truth conditional and truth theoretic perspectives of the analysis of proposi-

tional statements and explored the criteria as how to decide between them. An experimental approach to philosophy was presented by Barcz Michal's and Leon Ciechanowski's (both from Warsaw) paper entitled 'An Ambiguity of 'Intentionally': Has Ascombe's Thesis Been Experimentally Refuted?' followed by Tabea Reiner's (München) 'Semantics and Semantics'. The last plenary lecture gave John Collins (East Anglia) on 'What is a Variable such that a (Natural) Language May Realize One'. In it, Collins clarified the notion of variable as well as argued for several grades of variable involvement. Finally, Natalia Karzewska (Warsaw) discussed the problem as 'What Kind of Disagreement is there in Faultless Disagreement? Revision of Kölbel's Relativist Refutation against the Possibility Thereof' and Krzysztof Kosecki (Łódź) took a look at the status of categories in signed languages.

Of course, the above talks are just a part of all what was presented at the conference. But even that suffices for the claim that PhiLang belongs among the best conferences dedicated to the philosophy of language and linguistics in Europe.

Martin Vacek

martinvacekphilosophy@gmail.com

Pro-Fil je elektronický časopis, který publikuje příspěvky z oblasti filosofie. Čtenáři zde naleznou texty, stati, rozhovory, recenze, anotace, upozornění a zprávy. Časopis zveřejňuje odborné studie i příspěvky vhodné pro výuku filosofie.

Pro-Fil, Vol. 13 (2012), No. 2

Z obsahu

Jesús Padilla Gálvez: Euklidovský prostor a zorný prostor jako estetický problém

Iva Svačinová: Epistemologické výzvy platonismu a Gödelovo pojetí matematické intuice

Jan Votava: Parapsychologie a pojem replikace

Monika Bečvářová: „Nature's Bastards“ z Royal Society: Obhajoba přírodní filosofie v díle Margaret Cavendishové (1623-1673)

Miloš Kratochvíl: K hermeneutice vědeckých teorií cestou Héléne Metzgerové

František Škvrnda: Čo iné je Platónov Sókratés, než len Platón vpredu, Platón vzadu a v strede Chiméra?

Marek Hrubec: Idea otroctví v dějinách evropského a severoamerického myšlení

Lukáš Likavčan: Význam kontextu

<http://profil.muni.cz>