

Contents ▪ Obsah

<i>Editorial</i> [in English]	3
<i>Editoriál</i> [in Slovak]	4

ARTICLES ▪ STATE

S. Park: <i>Against the Besire Theory of Moral Judgment</i> [in English]	5
S. Yang: <i>Emotion, Experiential Memory and Selfhood</i> [in English]	18
M. Nuhliček: <i>Prečo externalistické kritériá zdôvodnenia principiálne zlyhávajú?</i> [<i>Why Do the Externalist Criteria of Justification Essentially Fail?</i> ; in Slovak]	37
M. Vraný: <i>Empirická výzva filosofickému pojetí mysli</i> [<i>An Empirical Challenge to the Philosophical Conception of Mind</i> ; in Czech]	50
P. Kořátko: <i>Fikce, skutečnost a radikální vyprávění</i> [<i>Fiction, Reality and Radical Narration</i> ; in Czech]	72

DISCUSSIONS ▪ DISKUSIE

S. Sousedík: <i>Na závěr úvah o intencionálních jsovcnech</i> [<i>To Conclude the Reflections on Intentional Entities</i> ; in Czech]	97
M. Mikušiak: <i>K logickej analýze lingvistického pojmu synonymie</i> [<i>On the Logical Analysis of the Linguistic Notion of Synonymy</i> ; in Slovak]	106

BOOK REVIEWS ▪ RECENZIE

J. Halas: M. Picha, <i>Kdyby chyby. Epistemologie myšlenkových experimentů</i> [in Slovak]	115
M. Kostelec: B. Geistová Čakovská, <i>Vztáhy synonymie. Logická sémantika verzus lingvistické teorie o synonymii</i> [in Slovak]	121
P. Zamarovský: S. Hawking – L. Mlodinow, <i>The Grand Design</i> [in English]	125

REPORTS ▪ SPRÁVY

F. Gahér: <i>50 rokov Katedry logiky a metodológie vied Filozofickej fakulty Univerzity Komenského v Bratislave [50th Anniversary of the Department of Logic and Methodology of Sciences, Faculty of Philosophy, Comenius University in Bratislava; in Slovak]</i>	129
M. Vacek: <i>25 Years In Contradiction [in English]</i>	131
V. Vohánka: <i>Projekt Analytická teologie [The Analytical Theology Project; in Czech]</i>	133

Dear Readers,

Organon F commences its 20th year of existence. Now, it seems incredible that, in spite of various unfavorable circumstances, it has not only survived, but also has even grown in content and form. We publish more papers in English and still more authors from around the world publish their papers in the journal. Also, and even more importantly, we are very happy that now our journal attracts more authors from Slovakia and that it publishes a lot of papers by young researchers. It is evidence that, despite the prevalence of the non-analytic philosophy at the Slovak universities, analytic philosophy has become established in the area.

Doubtless, Organon F is taken as a respected philosophical journal. The addition of the “international” to the “philosophical journal” stresses not only the fact that the journal is issued by institutions of two countries and that the papers appearing in it are written in various languages but also the fact that it is indexed in the most prestigious world databases and that the advisory board consists of, over and above the most prominent Slovak and Czech philosophers, famous philosophers from around the world. Strengthening the international scope of the journal is our primary ambition.

The Journal distribution, starting with this anniversary volume, undertakes the principal changes. Since our ambition to open up Organon F for a much wider audience and, hence, to proceed in the propagation of analytic philosophy, we come to change the edition form and the distribution policy.

Beginning with No. 1/2013, Organon F also appears in an online form and every paper is freely accessible on our website (<http://www.klemens.sav.sk/fiusav/organon>). The access to papers is unlimited and free, and there are no authorization codes. However, the print issues of the journal will still be available in selected libraries.

Allowing for these changes, which will hopefully lead to a better availability of the papers appearing in the journal, we decided to cancel the journal subscription fees for individuals and institutions. We suppose that having the papers online makes the subscription fees superfluous. However, it still holds that selected libraries and philosophical institutions will receive the printed version for free. Also, those individual readers interested in the printed version are highly invited to contact our office directly at filoorgf@savba.sk. These changes do apply to the regular as well as to the supplementary issues of the journal.

We believe that the new form of the journal distribution will be compelling not only to our readers, but also to our authors. The papers appearing in Organon F will be easily and illimitably available from all around the world. For that reason the papers, especially those written in English, can reach much more readers than before.

We believe that you will appreciate the recent changes concerning availability and distribution of our journal and that you will continually hold favor for it.

The Editors

Vážené čitateľky,
vážení čitatelia,

Organon F vstupuje do 20. ročníka svojej existencie. Je to priam neuveriteľné, že takýto časopis dokázal v našich podmienkach nielen prežiť, ale sa obsahovo aj formálne úspešne rozvíjať. Publikujeme čoraz viac článkov v angličtine a dávame väčší priestor zahraničným autorom z celého sveta. Osobitne nás však teší fakt, že v poslednom čase prejavujú o časopis zvýšený záujem naši domáci autori, ale aj to, že sa v ňom začína objavovať viac kvalitných príspevkov od mladých bádateľov. Svedčí to o tom, že napriek dominantnému postaveniu neanalytickej filozofie na našich vysokých školách sa analytická filozofia v našom prostredí udomácnila.

Organon F patrí medzi rešpektované filozofické periodiká. Pred časom sme k prívlasktu „filozofický časopis“ pridali „medzinárodný“, čím sme chceli podčiarknuť nielen skutočnosť, že sa v ňom objavujú príspevky v rôznych jazykoch a že sa na jeho vydávaní podieľajú vedecké inštitúcie z dvoch krajín, ale aj fakt, že je indexovaný v tých najprestížnejších svetových databázach a že v jeho redakčnej rade pôsobia okrem najvýznamnejších českých a slovenských analytických filozofov aj významní predstavitelia z celého sveta. Našou ambíciou je naďalej posilňovať medzinárodný rozmer *Organonu F*.

Jubilejný ročník sa nesie v znamení zásadných zmien v distribúcii časopisu. Keďže našim úmyslom je sprístupniť *Organon F* čo najširšiemu publiku a naďalej tak zvyšovať záujem o analytickú filozofiu, meníme od roku 2013 formu vydávania a distribučnú

politiku. Od čísla 1/2013 *Organon F* vychádza aj elektronicky a všetky príspevky budú voľne dostupné na našej webovej stránke (<http://www.klemens.sav.sk/fiusav/organon>). Prístup k článkom bude neobmedzený, zdarma a nebude potrebné žiadne prístupové heslo. Tlačená verzia časopisu nezaniká a bude dostupná vo vybraných knižniciach.

Vzhľadom na tieto zmeny, ktoré povedú k lepšej dostupnosti článkov vychádzajúcich v časopise *Organon F*, sme sa rozhodli zrušiť predplatné časopisu pre jednotlivcov aj inštitúcie. Predplatné stráca svoju funkciu, keď články z časopisu sú voľne prístupné na internete. Zároveň však platí, že vybraným knižniciam a filozofickým pracoviskám budeme tlačennú verziu poskytovať zdarma. Pre individuálnych záujemcov bude tlačenná verzia časopisu k dispozícii v redakcii. Ďalšie informácie o jej dostupnosti radi poskytneme priamo v redakcii (filoorgf@savba.sk). Tieto zmeny sa týkajú riadnych aj mimo-riadnych čísiel časopisu.

Sme presvedčení, že nový spôsob distribúcie časopisu bude zaujímavý nielen pre čitateľov, ale aj pre našich autorov. Články, ktoré vychádzajú v *Organone F*, budú pohodlne a neobmedzene prístupné z celého sveta, a preto predovšetkým anglicky píšuci autori majú možnosť osloviť neporovnateľne viac potenciálnych čitateľov než doteraz.

Veríme, že aktuálne zmeny týkajúce sa dostupnosti a distribúcie časopisu *Organon F* privítate a že aj v novej podobe si zachová Vašu doterajšiu priazeň, za ktorú ďakujeme.

Redakcia

Against the Besire Theory of Moral Judgment¹

SEUNGBAE PARK

Division of General Studies, Ulsan National Institute of Science and Technology
Ulsan 689-798, Republic of Korea
nature@unist.ac.kr

RECEIVED: 08-03-2012 • ACCEPTED: 30-05-2012

ABSTRACT: This paper critically examines two objections and raises a new objection against the besire theory of moral judgment. Firstly, Smith (1994) observes that a belief that *p* tends to expire whereas a desire that *p* tends to endure on the perception that not *p*. His observation does not refute the sophisticated version of the besire theory that to besire that *p* is to believe that *p* and to desire to act in accordance with the belief that *p*. Secondly, Zangwill (2008) claims that the strength of motivation may vary while the degree of belief remains constant. Besirists would reply that a besire admits of both degree and strength. Finally, I argue that the belief that *p* endures while the desire to act on the belief that *p* expires with the introduction of a new bodily condition, and hence that the belief and the desire are distinct mental states.

KEYWORDS: Belief – besire – desire – motivation – perception.

1. Introduction

What is the psychological structure of moral judgment? Some meta-ethicists advocate a besire theory of moral judgment which asserts that a moral judgment is a besire. A besire is alleged to be a unitary mental state. It is similar to a belief in that it represents a state of affairs, and to

¹ This paper improved significantly thanks to the insightful and meticulous comments from Boram Lee at University of Connecticut and anonymous referees for *Organon F*.

a desire in that it motivates an action. But it differs from a belief in that it is motivationally efficacious, and from a desire in that it purports to depict a fact in the world. Thus, the defining characteristic of a besire is that it performs both the function of representing a purported fact and the function of motivating an action:

Besires are not supposed to be gerrymandered mereological sums of two states – a belief plus a desire – each of which could occur without the other. Besires are conceived of as unitary states that have both the representational characteristics of beliefs and the motivational characteristics of desires (Smith 1994, 119). (Zangwill 2008, 51)

Note that the besire theory is refuted, if we can imagine a situation where the belief and the desire come apart. So far as I can tell, there can be a simple version and a sophisticated version of the besire theory, depending on what the content of the desire is. I will explicate them below and then argue that none of them is tenable.

The debate over the ontological status of besires sheds light on the psychological nature of moral judgments. It also has ramifications on the debate over the ontological status of moral facts, truths, and knowledge. If moral besires are true, there should be moral facts in the world that render them true.² The besires would be true in virtue of moral states of affairs. Furthermore, if the true besires are backed up by good evidence, they could count as moral knowledge. Moral knowledge alone would essentially motivate agents to act morally. If the main thesis of this paper is true, however, besires do not exist, and hence they cannot serve as a means to establish the existence of moral facts, truths, and knowledge.

My discussion below proceeds under the assumption that a belief is motivationally inert. Bromwich (2010) argues, however, that a belief that *p*, whether moral or non-moral, is motivationally efficacious in the sense that it disposes an agent to answer in the affirmative to the question: is it true that *p*? Many eminent philosophers are on her side. Nevertheless, I criticize her arguments in Park (forthcoming) and defend the Humean view that a belief is devoid of dispositional content. My main idea therein is that we answer affirmatively to the question not because a belief disposes us to do so but because we have a standing and default disposition to do so, and this dis-

² I agree with Goldman (1999) that the correspondence theory of truth is the best theory of truth, and that it is committed to worldly truth-makers.

position is external to the belief that p . Rather than expound the idea here and go off on a tangent, I refer the interested reader to the aforesaid article.

2. Smith's strategy

The simple version of the besire theory claims that to besire that p is to believe that p and to desire that p . Note that the belief and the desire have the same propositional content, p . On this version, to besire that we ought to save a drowning child, for example, is to believe that we ought to save a drowning child and to desire that we ought to save a drowning child. What does it mean for an agent to desire that we ought to save a drowning child? Perhaps, the agent abhors the absence of the moral fact in the world that we ought to save a drowning child, so he wishes to make it the case. Making it the case that we ought to save a drowning child would mean that the truth-value of the moral belief in question changes from falsehood to truth. It sounds, however, implausible that we can create such a fact. It is not even clear how we can bring about the existence of such a fact. Therefore, no sensible besirist would advance the simple version.

Smith's critique of the simple version can be illustrated with the example of water in the refrigerator. Suppose that you believe that there is water in the refrigerator. You feel thirsty, so you desire for the state of affairs in which there is water in the refrigerator. You open the refrigerator only to find that there is no water in it. In such a situation, you would tend to stop believing that there is water in the refrigerator, but you would tend to continue to desire for the state of affairs in which there is water in the refrigerator. After all, your thirst has not yet been quenched. Thus, a belief that p tends to expire whereas a desire that p tends to persist on a perception that not p :

... a belief that p tends to go out of existence in the presence of a perception with the content that not p , whereas a desire that p tends to endure, disposing the subject to bring it about that p . (Smith 1994, 115)

It would be incoherent for a mental state to tend to expire and persist at the same time. Therefore, Smith concludes, a besire that p does not exist. A shortcoming of Smith's critique is that it may undermine the simple version, but not the sophisticated version of the besire theory.

The sophisticated version asserts that to besire that p is to believe that p and to desire to act in accordance with the belief that p . Let me provide

two examples. To besire that we ought to save a drowning child is to believe that we ought to save a drowning child, and to desire to act in accordance with the belief, e.g., to desire to save a drowning child. To besire that there is water in the refrigerator is to believe that there is water in the refrigerator, and to desire to act on the belief, e.g., to desire to open the refrigerator. Unlike the simple version, the sophisticated version takes into account what Little, Tenenbaum, and Bromwich observe about besires when they write as follows:

... it is a believing-attitude directed toward one proposition, and it is a desiring attitude directed toward another. (Little 1997, 64)

Besires are supposed to be complex mental states that have the direction of fit of belief towards one content (say p) and the direction of fit of desire towards another content (say q). (Tenenbaum 2006, 237)

... a belief disposes a subject to act as if the content of her belief is the case; a belief does not dispose a subject to bring about the content of her belief. (Bromwich 2010, 346, footnote)

Thus, on the sophisticated version, although the cognitive content of one's besire is that p , the motivational content of the besire is not to bring about the state of affairs that p , but to bring about the state of affairs that one acts in accordance with the belief that p . The besire that we ought to save a drowning child represents the moral fact that we ought to save a drowning child, and it motivates an agent to save a drowning child. The besire does not motivate the agent to bring about the moral fact that we ought to save a drowning child. Also, the besire that there is water in the refrigerator represents the fact that there is water in the refrigerator, and it motivates an agent to open the refrigerator. It does not motivate the agent to bring about the fact that there is water in the refrigerator. Therefore, the sophisticated version is not committed to the implausible view that we can generate a moral fact.

In order to refute the sophisticated version, Smith would have to present a perception that tends to expel the belief that p , but not the desire to act in accordance with the belief that p . Does such a perception exist? What perception would tend to oust the belief that we ought to save a drowning child, but not the desire to save a drowning child? As far as I can tell, no such perception exists. After all, you can maintain your belief that we ought to save a drowning child, whether or not you perceive a per-

son saving the child. No matter what you observe, you can stick to the moral belief. Suppose, however, that you gave up the moral belief after repeatedly seeing people ignore drowning children. A problem is that you would, then, tend to lose the desire to act on the belief, and tend to form a new desire, viz., the desire to act on the new belief that we ought not to save a drowning child. Consequently, no perception can drive a wedge between the belief that *p* and the desire to act on the belief that *p*.

Also, what perception would tend to dispel the belief that there is water in the refrigerator, but not the desire to act in accordance with the belief? The perception that there is no water in the refrigerator may tend to drive out the belief that there is water in the refrigerator, but it is not clear that it tends to leave intact the desire to act on the belief. Suppose, for example, that your wife opens the refrigerator, and that you see no water in the refrigerator. You would, then, tend to lose the desire to act on the belief and tend to form a new desire, the desire to act on the new belief that there is no water in the refrigerator. Thus, on the perception that not *p*, the desire to act on the belief that *p* tends to go out of existence along with the belief that *p*. Therefore, Smith's observation that on the perception that not *p*, the belief that *p* tends to expire whereas the desire that *p* tends to endure does not undermine the sophisticated version of the besire theory, which proposes that to besire that *p* is to believe that *p* and to desire to act on the belief that *p*.

3. Zangwill's strategy

Zangwill (2008) observes that the strength of motivation may vary while the degree of belief remains constant. Consider two agents who believe that it is wrong to take a bribe. They are both rational agents in that neither of them suffers from mental debilities, such as depression and listlessness. The degrees of their moral beliefs are the same in that they both believe with equal level of confidence that it is immoral to take a bribe. Suppose, however, that the first agent cares more about the demands of morality than the second agent, so that the first agent refuses the bribe whereas the second agent takes it. From the point of view of the sophisticated version of the besire theory, it is mysterious why the first agent rejects the bribe and the second agent accepts it. After all, they both believe with equal degree of confidence that it is wrong to take a bribe, so they must be equally motivated not to take a bribe and they should both turn down the bribe.

Besirists might say that the first agent declines the bribe while the second agent receives it because the first agent is rational while the second agent is irrational. This answer, however, is not adequate because both agents are, by hypothesis, rational. In response, the besirists might now say that the first agent rejects the bribe because he believes with a high degree of confidence that it is immoral to take the bribe, and that the second agent accepts the bribe because he believes with a low degree of confidence that it is immoral to take the bribe. This move, however, is not available to the besirists either because, by hypothesis, both agents believe with the equal degree of confidence that it is immoral to take a bribe. Besirists cannot, Zangwill concludes, account for the variation in the degree of motivation and the resulting difference in the behaviors of the two agents.

In contrast, Zangwill, a motivational externalist, has an adequate explanation: the first agent rejects the bribe, while the second agent accepts it because they have desires of different strengths not to take the bribe:

... there is a difference in motivation between people with moral beliefs of the same degree that *is* best explained by a difference between them in their desires. (Zangwill 2008, 56)

Thus, the variation in the motivations stems not from any variation in the degrees of besires but from a variation in the strengths of desires. What is crucial on Zangwill's account is that the desires are extraneous to the beliefs. Thus, the variation in the desires can be explained by motivational externalism, but not by the besire theory.

In my view, however, besirists would retort that a besire admits of both degree of confidence and strength of motivation. The aforementioned two agents hold the besire with equal degree of confidence, but they besire with different strengths of motivation that we ought not to take a bribe. The first agent besires strongly that it is wrong to take a bribe, whereas the second agent besires weakly that it is wrong to take a bribe. Their besires are of different strengths, and hence they lead to different levels of motivation and resultant behaviors. Thus, the variation in the motivations and resultant behaviors of the two agents originates not from their besires' degrees of confidence (which are assumed to be the same), but from their different strengths of motivation. In short, Zangwill overlooked the possibility that the variation in the motivations and behaviors of the two agents can be explained by the besires' having different strengths of motivation.

4. My strategy

Let me sketch my strategy for rebutting the sophisticated version of the besire theory. Imagine that while patrolling the beach, a lifeguard sees a child drowning in the sea. He believes that we ought to save drowning children, and he desires to save the drowning child. As a result, he jumps into the water. Unfortunately, a venomous sea snake bites him on his leg. The excruciating pain eliminates his desire to save the child. Not at all motivated to save the child, he simply flounders in the water. Just then, a fisherman comes along on a boat and asks, "Do you believe we ought to save the drowning child?" The lifeguard says, "Yes." The fisherman rows his boat to the child and pulls him out of the water. He comes back to the lifeguard only to learn that he has died. This tragic example illustrates how a new bodily condition puts an end to the desire to act in accordance with the belief that *p*, but does not put an end to the belief that *p*. After the poison got into his body, the lifeguard lost the desire to save the drowning child, but he continued to believe that we ought to save the drowning child. Thus, a belief that *p* can occur without a desire to act on the belief that *p*, contrary to what the sophisticated version claims.

Let me take another example. Imagine that your body is in need of water, so you feel thirsty. You believe that there is water in the refrigerator. You have the desire to act on the belief that there is water in the refrigerator, and hence you are motivated to open the refrigerator. In such a situation, besirists would say that you besire that there is water in the refrigerator. Suppose, however, that while walking to the refrigerator, you spot a bottle of water on a table, and you drink from it. As a result, your body no longer lacks water. Due to the new bodily condition, you no longer desire to open the refrigerator, but you continue to believe that there is water in the refrigerator. Thus, your belief endured while your desire to act on the belief expired with the introduction of the new bodily condition. Therefore, the belief and the desire cannot form a unitary mental state, viz., the besire that there is water in the refrigerator.

Besirists would retort that after drinking the water, you still have the desire to act on the belief that there is water in the refrigerator, in the sense that if someone asks you whether or not there is water in the refrigerator, you would respond in the affirmative. You had that desire even before drinking water. It appears that this desire to act on the belief persisted along with the belief, and hence the besire that there is water in the refri-

generator continued to exist in your mind. So does the lifeguard's desire to act in accordance with the belief that we ought to save drowning children. He did not lose the desire, given that he answered in the affirmative to the fisherman's question. Therefore, it still stands that the belief that p cannot occur without the desire to act on the belief that p .

My response to the preceding observation is to point out the differences between the desire to act on the belief before you drink water and the desire to act on the belief after you drink water. Let's call them 'D₁' and 'D₂' respectively. It is granted that D₁ and D₂ share the dispositional content to answer affirmatively to the question of whether or not there is water in the refrigerator. Recall, however, that D₁ motivated you to open the refrigerator whereas D₂ did not, and hence they have different motivational contents. They have different phenomenological contents too. D₁ has an unpleasant phenomenological content, viz., the thirsty sensation, which you wanted to get rid of, whereas D₂ does not have the unpleasant phenomenological content. Since D₁ and D₂ have different motivational contents and different phenomenological contents, they are different desires, although the umbrella expression 'the desire to act on the belief' is used to refer to either of them.

The same is true of the example of the lifeguard above. Let's use 'D₁' and 'D₂' respectively to refer to the desire the lifeguard had before he was bitten by the snake and the desire he had after he was bitten by the snake. It is granted that D₁ and D₂ share the dispositional content to answer affirmatively to the question of whether or not we ought to save a drowning child. But there is an important difference between them. D₁ motivated the lifeguard to save the child whereas D₂ did not. Recall that he lost the desire to save the child after being bitten by the snake. Since they have different motivational contents, they are different desires, although the umbrella phrase 'the desire to act on the belief' is used to refer to either of them.

Besirists would insist that D₁ and D₂ are the same mental state. Consider that a person can persist as the same person over time, even if he loses his mental features and physical features. Analogously, a mental state can persist over time, even if it loses its contents. D₁ lost the motivational content and the qualitative content, but it is the same mental state as D₂. After all, D₁ and D₂ are conjoined by the dispositional content to answer affirmatively to the question: is it true that p ? Besides, when you say yes to the question, you are acting in accordance with your belief that p . It is not the

case that you are no longer acting on the belief. Therefore, besirists maintain, either desire merits the appellation 'the desire to act on the belief'.

As I mentioned earlier, I argued in Park (forthcoming) that a belief that *p* does not embed the disposition to answer affirmatively to the question: is it true that *p*? However, let me assume here for the sake of argument that the disposition is internal to the belief, and that they jointly form the besire that *p*. What I want to point out here is that the existence of such a besire does not have any interesting ethical implication. After all, the scope of its motivational force is limited to answering a question honestly. It does not motivate us to act morally. The besire, for example, that we ought to save a drowning child does not motivate us to save a drowning child. It only motivates us to say yes to the question: is it true that we ought to save a drowning child? It is not clear that such besires are worth fighting for from the besirists' perspective. Such besires, even if true and justified, would not be the kind of moral knowledge that motivates us to act morally.

Let me distinguish between what I call moral besires and conversational besires. A moral besire motivates an agent to act morally. For example, the lifeguard above had a moral besire when he jumped into the water to save the drowning child. An agent whose mind is filled with moral besires tends to perform moral acts, and hence he is a virtuous person. In contrast, a conversational besire only motivates an agent to answer a question as he believes. In this sense, the lifeguard above only had a conversational besire when he responded to the fisherman affirmatively. An agent whose mind is full of conversational besires might be a cooperative conversational partner, but he might not be a virtuous person. He might be the kind of person who habitually commits immoral acts, such as murder, theft, and adultery. Honesty might be his sole admirable moral quality.

In any event, based on the two examples discussed earlier in this section, let me make a general claim about the relationship among a bodily condition, a belief that *p*, and a desire to act on the belief that *p*. Upon the introduction of a new bodily condition in certain circumstances, the belief that *p* endures whereas the desire to act on the belief that *p* expires. Put differently, a new bodily condition may not annihilate the belief that *p*, whereas it may exterminate the desire to act in accordance with the belief that *p*. Thus, the belief that *p* can exist without the desire to act on the belief that *p*. It is impossible that a mental state exists and does not exist at the same time. Therefore, the belief that *p* and the desire to act on the be-

lief that p cannot form a single mental state, i.e., a desire that p . The putative unitary mental state does not exist.

My strategy differs from Smith's strategy in the following respects. Smith invites you to imagine situations in which the perception that not p is introduced. He claims that the perception that not p tends to oust the belief that p , but not the desire that p . In contrast, I invite you to imagine situations in which a new bodily condition arises. I claim that there are some situations where the new bodily condition extinguishes the desire, but not the belief. Thus, in Smith's strategy what is newly introduced is the perception that not p , whereas in my strategy it is a new bodily condition. Also, in Smith's strategy the target that is extinguished by the newly introduced condition is the belief that p , whereas in my strategy it is the desire to act on the belief. In a nutshell, the two strategies produce variations in different effects by introducing variations in different causes. Finally, my strategy is successful whereas Smith's strategy is not in confuting the sophisticated version.

There is a further reason for thinking that a belief that p and a desire to act on the belief that p are distinct mental states. The strength of a desire to act on a belief that p varies together with the potency of its cause, not with the potency of the cause of the belief that p . For example, the strength of the desire to act on the belief that there is water in the refrigerator varies in accordance with the amount of water in your body. If your body is in severe need of water, the desire will be strong enough to impel you to open the refrigerator. If the lack of water is moderate, the desire may not be strong enough to induce you to open the refrigerator. In contrast, the degree of a belief covaries with the power of its cause, not with the potency of the cause of a desire. For example, the degree of the belief that there is water in the refrigerator is correlated with the vivacity of the perception that there is water in the refrigerator. If the perception is vivid, the degree of the belief will be high. If the perception is dull, the degree of the belief will be low.

Moreover, the degree of a belief that p may be inversely proportional to the strength of a desire to act on the belief that p . For example, you may be confident that there is water in the refrigerator, but you may not desire at all to open the refrigerator. This happens when you have a vivacious perception of water in the refrigerator, but there is enough water in your body. In contrast, you may be unsure that there is water in the refrigerator, but you may strongly desire to open the refrigerator. This happens when you

have an obscure perception of water in the refrigerator due to an eye disease, but there is a severe shortage of water in your body. In short, different factors exert different influences on the belief that *p* and the desire to act on the belief that *p*. This phenomenon reinforces the view that the belief that *p* and the desire to act on the belief that *p* are separate mental states.

Why does a belief that *p* endure whereas a desire to act on the belief that *p* expires, given some appropriate change of bodily conditions? The answer lies in the fact that the belief and the desire are under the influence of different physical factors. The amount of water in the body determines whether a subject feels the desire to drink water or not. It is not surprising that a subject loses the desire after the intake of water rises above the needed threshold. In contrast, the transmission of signals along the optic nerves determines whether a subject believes that there is water in the refrigerator or not. It is natural that a subject maintains his perceptual belief even after the amount of water in the body increases. Since the belief and the desire are under the control of different factors, the elimination of the bodily condition that caused the desire puts an end to the desire but not to the belief.

In general, different physical factors are responsible for different beliefs and for different desires. A visual belief is generated by the physical processes in our eyes, while an auditory belief is generated by the physical processes in our ears. A desire for water is created by the lack of water in our body, while a desire for food is generated by the release of a hormone called ghrelin. It is up to scientific inquiry, not to conceptual analysis, to reveal the one-to-one correlations between beliefs and physical factors, and between desires and physical factors. What is important for my purpose here is that beliefs and desires are causally related to different physical factors, so they are separate mental states, and besires do not exist.

A recent finding in neuroscience speaks in favor of the view that in general a belief and a desire are distinct mental states. Some neuroscientists have discovered that a region of a rat's brain called the rostral AGM activates when the rat chooses an action among multiple alternatives which lead to different rewards:

Our results indicate the involvement of the rostral AGM not only in action selection but also in valuation, which is consistent with the finding that AGM activity is modulated by expected reward. (Jung Hoon Sul et al. 2011, 6)

This neurological finding indicates a possible difference between neural substrates underlying our decisions about what to do and those underlying our deliberations about what to believe. It may be discovered in the future that a certain region in our brain activates when we are motivated to save a drowning child, but not when we merely believe that we ought to save a drowning child. Also, the region may turn out to be inactive in the brains of dejected agents suffering from volitional debilities, such as *accidie* and *akrasia*, even when they claim to believe that we ought to save a drowning child. Such neural discovery would amount to empirical evidence disconfirming the desire theory of moral judgment.

Tenenbaum, however, would not be swayed by the aforementioned neural discovery. Suppose that agents suffering from *accidie* and *akrasia* claim to believe that we ought to save a drowning child, and yet are not motivated to save a drowning child. On Tenenbaum's account, it is wrong to attribute moral beliefs to them because they have not fully grasped the content of the moral beliefs:

So the motivational cognitivist is committed to seeing those motivational failures as in themselves failures to fully grasp the content of one's moral beliefs, or somehow failing to have the same kind of moral beliefs as the moral agent. (Tenenbaum 2006, 257)

For Tenenbaum, appropriate motivation is constitutive of a moral belief. Lethargic agents do not have moral beliefs, although they claim that they do, because they are not motivated to act appropriately. Their failure to be motivated to act morally proves, by Tenenbaum's stipulated definition of 'moral belief', that they do not have moral beliefs.

In my view, however, a puzzle arises. What is it, if not a moral belief, that the listless agents have when they sincerely claim to believe that we ought to save a drowning child? Obviously, they take it to be a belief, although they are aware that they are not motivated to act morally. Tenenbaum, on the other hand, refuses to call it 'belief' on the grounds that it does not motivate them to save a drowning child. The dejected agents would insist that what they have in their minds is a moral belief, pointing out that they have better epistemic access to their own mental states than Tenenbaum does, and that Tenenbaum's definition of 'moral belief' is no more correct than their definition of 'moral belief'. Of course, these arguments are not decisive, but Tenenbaum has the burden to account for their mental states.

5. Conclusion

The sophisticated version of the besire theory asserts that to besire that *p* is to believe that *p* and to desire to act in accordance with the belief that *p*. I have argued that in some cases, a new bodily condition that does not obliterate a belief that *p* terminates a desire to act on the belief that *p*. It is impossible for a mental state both to endure and to expire at the same time. Therefore, the besire that *p*, as posited by the sophisticated version, does not exist. It is an open question whether moral knowledge exists or not. What this paper is intended to establish is the claim that if moral knowledge exists at all, it is purely a cognitive state,³ and hence it does not motivate an agent to act morally. What motivates an agent to act morally is not a moral belief but a mental state extraneous to it.

References

- BROMWICH, D. (2010): Clearing Conceptual Space for Cognitivist Motivational Internalism. *Philosophical Studies* 148, No. 3, 343-367.
- GOLDMAN, A. (1999): *Knowledge in a Social World*. Oxford: Oxford University Press.
- JUNG, H. S. – SUHYUN, J. – DAEYEOL, L. – MIN, W. J. (2011): Role of Rodent Secondary Motor Cortex in Value-Based Action Selection. *Nature Neuroscience* 14, 1-7. DOI: 10.1038/nn.2881
- LITTLE, M. (1997): Virtue as Knowledge: Objections from the Philosophy of Mind. *Noûs* 31, No. 1, 59-79.
- PARK, S. (2012): Against Moral Truths. *Cultura. International Journal of Philosophy of Culture and Axiology* 9, No. 1, 179-194.
- PARK, S. (forthcoming): Against the Motivational Efficacy of Beliefs. In: Imafidon, E. (ed.): *Ontologized Ethics: A Critical Reader in (African) Meta-Ethics*.
- SMITH, M. (1994): *The Moral Problem*. Oxford: Blackwell.
- TENENBAUM, S. (2006): Direction of Fit and Motivational Cognitivism. In: Shafer-Landau, R. (ed.): *Oxford Studies in Metaethics*. Oxford: Oxford University Press.
- ZANGWILL, N. (2008): Besires and the Motivation Debate. *Theoria* 74, 50-59.

³ By the way, I have constructed what I call the dilemma of moral perception in Park (2012) against the view that a moral judgment has cognitive content. In this paper, however, I set aside the debate between cognitivists and noncognitivists.

Emotion, Experiential Memory and Selfhood

SUNNY YANG

Department of Philosophy, Seoul National University
Seoul 151-742, South Korea
yangseonee@snu.ac.kr

RECEIVED: 03-03-2012 • ACCEPTED: 28-09-2012

ABSTRACT: Recently, emotion has attracted much attention in many areas of philosophy. In the philosophy of mind, some argue that emotions are individuated and identified with reference to feelings, beliefs, desires, or perceptions. Furthermore, they are often claimed to be changeable, unstable, and ambivalent. However, despite their instability, emotions are sometimes long-standing. They have, in addition, perspective. These characteristics of the emotions, I argue, help us in solving one of philosophy's most enduring problems, that is, the problem of personal identity. In order to illustrate this claim I elaborate on the conception of 'experiential memory' suggested by Wollheim. To understand memory as experiential, I argue, we need to understand the affective element attached to some memories. I argue that memory affects not only my past thought but also my past emotions, and those emotions deriving from the past stay on to affect my whole being and my future. Hence, I argue that experiential memory is not just confined to the recalling of events or experiences that the subject has experienced, but concerns the narrative structure of a person's life as a whole.

KEYWORDS: Emotion – experiential memory – Lockean psychological continuity theory – personal identity – Wollheim.

1. Introduction

What are the conditions under which a girl *A* at time *t*, and 10 years later, a woman *B*, are the very same person? To answer this question, I focus on a mental connectedness theory of personal identity. In much recent

work, accounts of the criteria of personal identity have divided into two camps: bodily continuity and psychological continuity theses respectively. The former has sought the criterion in the domain of the physical, identifying personal identity with sameness of body; the latter has sought it in the domain of the mental, equating personal identity with continuity of memory. The latter view has been represented in recent literature by psychological continuity theories of personal identity that take Locke's original insight and develop it into a more detailed identity criterion. These memory-based theories of personal identity take memory as providing a retrospective form from present experiences to past experiences that explains the unity of a person's life.

In what follows, I reject the psychological continuity theory described above. Instead, I defend a different version of a psychological continuity theory according to which memory has as much to do with a person's future as with a person's past, due to the role that emotions play. In order to do this, the view which follows will elaborate 'experiential memory' suggested by Wollheim.¹

2. Experiential memory

The central role of experiential memory in explaining personal identity has been neglected by many philosophers in the mental connectedness tradition. In this section, I first develop and defend a psychological criterion for personal identity in and over time based on Wollheim's notion of experiential memory. In order to understand memory as experiential, we need to understand the affective element attached to some memories. Secondly, I draw on Wollheim's work to provide an account of personal unity that relies on memory with an emphasis on its dispositional and emotional aspect, making memory as important to a person's future as to her past.

Why is emotion the key player in the "affective" component of experiential memory? This is because experiential memory has a qualitative aspect and emotional aspect has qualitative thing. What then does the 'qua-

¹ Experiential memory is also termed personal memory which is in contrast with merely factual memory of one's past. The distinction between personal and factual memory is developed further in Andrew Hamilton (forthcoming); see also Goldie (2012).

litative aspect' of experiential memory mean? If we take emotions as feelings or perceptions, there is a proprioceptive residue of sensation and a point of view.² In order to understand this, let us consider the difference between a person remembering herself watching a horror movie and remembering *that* she watched a certain horror movie. We can say that the former is an experiential memory that reminds her of what it was *like* to watch such a horror movie. In experientially remembering watching a horror movie, a person thinks and *feels* as she experienced when she watched it. When I experientially remember watching a horror movie all those years ago, I tend to remember also the thoughts I had, the feelings I experienced, while watching the movie, and when I do so, those thoughts and feelings tend to take hold of me. I once again sense my muscles tensing, my pulse quickening, and adrenaline flowing, and those sensations deriving from the past persist to affect my whole being and future. The view described here does not mean that whenever I relive past events I should feel exactly as I did then. In defence of a psychological continuity theory which I have described above, in what follows, I shall develop four claims, made by Wollheim:

- a) an event remembered is one that I have lived through;
- b) an event is remembered *as* it was experienced and not just as it occurred;
- c) when we remember an event experientially an event is remembered from a point of view and this point of view is represented within memory (see Wollheim 1984, 104);
- d) an affective tendency is attached to such memories (see Wollheim 1980, 309-310).

Experiential memories take the form of 'I remember doing such-and-such' or 'I remember your, or X's, doing such-and-such', rather than 'I remember that I X-ed' (cf. Wollheim 1979, 195; see also Hamilton 1995). Now in order to see the distinction between 'I remember X-ing' and 'I remember that I X-ed', let us take an example: What Wollheim is intending to describe is the fact that when I remember something in my past

² Many philosophers of emotion have recently taken emotions to be perceptions. Some people suggest that we should take emotion to be perception, or proprioception. For a discussion of taking emotion or feeling as perception, or proprioception, see Solomon (2004).

i) I draw my attention to *what it was like to be* experiencing something from my point of view. ii) The affect that a memory preserves reconnects me with a broader affective dimension, that of a specific time in my life. I argue that memory affects not only my past thought but also my past emotions and those emotions deriving from the past persist on to affect my whole being and my future. Hence for example, when I remember my failing some examination, the despair, disappointment, or regret that I experience involves the sum of my experience from that point on, considered as a unified whole. My past experiences affect my present condition – my sense of disappointment – and this feeling expresses my awareness of the past. Then my concern for my future – a desire for a successful life – causes painful feelings and emotions, for example, my heart torn with anguish and regret, in my present condition. This is possible through memory's recreation of the past and creation of the future. Hence, for example, the shame or regret that I experience when I remember my past wrongdoing is part of my identity up to now as well as at that time. In this way, I argue, we bring the person under the influence of his past, such that the past is reconstructed and then influences the future. Hence, a life – the diachronic expansion of a person – is unified. Furthermore, we can say that this kind of unity of consciousness over time requires unity of consciousness at one time.³ A view that I defend on unity of consciousness is the self-conscious knowledge of bodily identity, or proprioceptive knowledge concerning my body.⁴

The problem of personal identity, so far I have discussed, is the problem of finding ontological criteria for the identity of the different temporal 'stages' of a person. In what follows, I shall show that emphasizing the emotional, experiential character of experiences and recollections contribute to solve the ontological problem of personal identity. I shall try to solve the transitivity problem that a memory theory faces by showing; 1) that the difference emotion makes is in the character of psychological continuity; 2) that the emotional aspect of memory allows for certain dispositions connected with it to live on when the experienced event that caused it is forgotten.

³ For the discussion of unity of consciousness, see Hamilton (forthcoming).

⁴ A further discussion on proprioceptive knowledge about my body, see Hamilton (forthcoming).

3. Identity and change

3.1. Transitivity implies changes

Now some people say that personal identity is constituted by sameness of the psychological subject. But it is not obvious what is meant by this. Locke, for example, claims that personal identity consists, not in sameness of substance, but in ‘sameness of consciousness’ and holds that continuation of the consciousness itself, not that of body or soul in which that consciousness resides, is necessary for identity. Butler claims that consciousness is ‘successive’ and, therefore, “cannot be the same in any two moments, nor consequently the personality constituted by it” (Butler 1975, 102). Reid argues that “consciousness, and every kind of thought are transient and momentary, and have no continued existence” (Reid 1975, 116). Consciousness is, according to these theorists, a momentary phenomenon and does not remain numerically the same over time. If this is the case, it is not clear what it means to say that the same consciousness has continued to exist.

In order to answer the question of what it means to say that the same consciousness has continued to exist, Locke claims that if you cannot currently remember the past thoughts and action of some person, then you simply are not the same person who had and did them.⁵ Here what Locke makes necessary for identity with a ‘past self’ is, as Shoemaker notes, “not that one remember the actions and experiences for that past self but that one have ‘memory continuity’ with that past self” (Shoemaker 1984, 81). However, one might raise an objection to Locke’s memory criterion of personal identity which focuses on the idea of *transitivity*. Reid tries to show this with his ‘brave officer’ example (cf. Reid 1975, 114-115). At a certain time a boy steals apples. Years later the same person, now a young officer, performs an act of bravery in battle, remembering still his boyhood’s theft of the apples. Many years later our man is an elderly general, who remembers the act of bravery in battle, but no longer remembers incident of stealing apples. Reid charges that on Locke’s theory the general is the same person as the young officer, who is the same person as the boy (because

⁵ For Locke, ‘person’ is a forensic term, which is related to our practices of attributing responsibility and distributing rewards and punishments. According to him, a person should not be held responsible and punished for actions of his which he cannot remember doing.

identity is transitive), but the general is *not* the same person as the small boy because he has no memory of the boyhood incident.⁶

Locke himself acknowledges that ‘the same Consciousness’ is not the same ‘individual Action’ but ‘a present representation of a past Action’ (Locke 1979, 337). Here we should allow, as Shoemaker points out, that “one’s current person-stage contains a memory of something even if one has temporarily forgotten that thing, as long as one has the potentiality of remembering it. ... In such a case the stage will retain a ‘memory trace’ that is the basis of that potentiality”, he adds (see Shoemaker 1984, 81).

3.2. Memory is a disposition

Now one might wonder how the view described in the previous section might be put to use in accounting for a person’s emotional change in relation to their identity. I can recognize that my emotions or feelings have changed when I look back on past experience. For example, when I was 5 years old, I was scared of a monster which appeared in a horror movie, but now I am not. Change can also be seen in, for example, recovery from shock or grief. How then can Wollheim explain our emotional change – for example, from having been afraid to no longer being so – by appealing to experiential memory?

If we apply the above claim b) – an event is remembered *as* it was experienced and not just as it occurred – to explain one’s identity we are left with the following: what it is to be a psychologically identical being is to be a creature who can be, from a first-person, phenomenological perspective, in the same state. Concerning this claim, Lowe argues that if a person “remembers his own thoughts, experiences, feelings and actions in a special way – that is, so-called a ‘first person’ way and if the person remembers them, as it were, ‘from the inside’, as episodes undergone *by him*”, this is different from the fact that the person “remember events in the life of another person only in an ‘external’ or ‘third-person’ way, as episodes undergone *by somebody else*” (Lowe 1995, 109).⁷

⁶ Locke might reply to Reid as follows: the general is not the same *person* as the boy, while pointing out that the general is nonetheless the same *human being* or *man* as the boy. This is pointed by E. J. Lowe; see Lowe (1995, 113).

⁷ I shall argue that this generates the question of synchronic unity. The synchronic unity question is: in virtue of what do different experiences or mental states occurring at the same time count as belonging to one and the same person?

Wollheim says that “the affective tone that accrues to the memory – and the impact of which is just what the beliefs, desires, and feelings – will differ qualitatively from that which would have accrued to it had the other entertained the memory” (Wollheim 1979, 230). And this is so because “when the rememberer’s affective store from which the affective tone comes was reinforced by an experiential memory, the impact of the memory was mediated by his current beliefs, desires and feelings, which become increasingly peculiar to him” (Wollheim 1979, 230). However, this seems to make trouble with Wollheim rather than helping him. For, as Lowe notes, “if one *can* have first-person memories of episodes in the lives of other people, then, clearly, such memory does *not*, after all, provide a satisfactory criterion of personal identity across time” (Lowe 1995, 110).

Faced with this difficulty, Wollheim would argue that it is not occurrent psychological states alone that are part of one’s psychological make-up, but underlying traits and dispositions as well – this concerns my claim a), namely, *a remembered event is one that I have lived through*. If we take a memory to be disposition, it is difficult to say a memory disposition which I have and a memory disposition which someone else has would manifest identical experiential memory states, even though both rely on the same initiating event.

This is because a memory disposition, established in me by my living through the event, would manifest itself differently from others’. If this is so it is impossible, as Wollheim points out, that “mental dispositions could shift from one person to another and remain intact” (Wollheim 1984, 113). Since “dispositions form... a web or network: they are ancillary to one another, and there seems no method for determining what in the way of other dispositions, what in the way of beliefs, emotions, desires, fears, and other memories, would have to be transferred along with it if the original memory is to run across persons” (Wollheim 1984, 113).⁸ Furthermore, if we accept my claim c), namely, when we remember an event experientially an event is remembered from a point of view and this point of view is represented within the memory, then we can understand that memory has as much as to do with a person’s future as with a person’s past. It is a point of view by which one’s disposition would, on occasions, manifest itself in appropriate mental states.

⁸ For these reasons, I find both the idea of memory replication and memory transfer implausible.

Having elaborated on my four claims, namely, a) memory is a disposition; b) memory has a first-person perspective; c) memory has a point of view; d) an affective tendency is attached to some memory, in what follows I shall argue that these four characteristics of experiential memory play a vital role in explaining the problem of both synchronic and diachronic unity.

3.3. Self-knowledge of emotional change: from the past to the present

With the understanding of psychological sameness described in section 3.1, we can explain the effect of a person's emotional change on his or her identity. I shall defend the Lockean memory criterion of personal identity, showing that change is constitutive of, rather than a threat to, personal identity.⁹ According to Wollheim, emotional change is possible because when my affective store, from which the affective tone comes, was reinforced by an experiential memory, the impact of the memory was mediated by my current *beliefs*, *desires* and *feelings*. Hence, I should not be scared of the monster in the horror movie due to my current *belief* that it is fictional character, even though I was scared of it in my childhood. Furthermore, I am not experiencing grief over the death of my father, due to my present *recognition* that the perceptual image of my father is weakened, the *belief* that I can no longer see him, and due to my *desire* for well-being in his absence. Hence, we can say that I am a psychologically identical being, from a first-person phenomenological perspective, even though my emotions or experiences have changed.

Although Wollheim claims that change is constitutive of, rather than a threat to, a person's identity, he does not explain how change is constitutive of personal identity. Change must occur in order that a person is not for an indefinite period struck down, for example, by grief or disappointment. On this question, Wollheim claims that "it is in the nature of persons that they change", since time is not merely a static but a transient thing and a development of the new out of the past and the present. Hence "as they live, they necessarily change, and therefore it would be a merit in a criterion of personal identity". Thus Wollheim argues that the mental connectedness criterion of personal identity must permit change. "It must

⁹ Some people read Locke's transitivity of consciousness as what William James (1890, Ch. 10) later calls the 'stream of consciousness.' I share this view. Cf. Lowe (1995, 114).

exhibit how change is compatible with identity” (Wollheim 1980, 313). Yet he does not develop this point. Hence, let us consider how such an account might be developed.

It might be said that emotional change is a kind of recovery. But once a person notices herself to have recovered, for example, from deep grief or shock, she could say that she is not the same as the person who once grieved or shocked. That is, she has changed. But the recognition that her feelings have been changed with passage of time is not itself a feeling. Then what is it? In order to appreciate this, let us consider ‘two senses of experience’ suggested by William James. William James defines the concept of memory as “the knowledge of an event, or fact with the additional consciousness that we have thought or experienced it before” (James 1890/1950, 648). He presents two senses of experience as follows:

A succession of feelings, in and of itself, is not a feeling of succession. And since, to our successive feelings, a feeling of their succession is added, that must be treated as an additional fact requiring its own elucidation. (James 1890/1950, 628-629)

Here James seems to suggest that the feeling of succession cannot be much like sensations or emotions. He suggests that it is one thing to be presented by a succession of feelings; it is another to appreciate it as a succession. On this view, memory plays an important role in taking notice of this succession, since without the ‘taking notice’, mere occurrence of successive perceptions would not provide us with a self-conception. But ‘taking notice’ is not itself a feeling. One might say that ‘taking notice’ of emotional change is *disaffection* with oneself. When someone, for example, finds herself not only changed but also finds that she has become an alien to herself many times over, she realizes that she is in a disaffected state. Martha Nussbaum sees this disaffected state as being in the state of “a detached, unemotional, or intellectual deliberation” (Nussbaum 2001, 24). She views this disaffected state as involving various cognitive features, such as a ‘personal point of view’, ‘cognitive denseness’, and ‘cognitive freshness’ (see Nussbaum 2001, 63-65; 79-85).¹⁰ These features do not imply noncognitive components for they impart the cognitive content of the judgments.

¹⁰ In this paper I do not discuss these features in more detail. Ben’zev (2004, 453-455) provides a critical assessment of these features in his paper ‘Emotions are not mere Judgment’.

Although I agree with Nussbaum that the recognition of emotional change depends on cognitive features, I do not accept her view that cognitive elements are both necessary and sufficient for such recognition, since in my view we can sometimes recognize our emotional change in terms of self-knowledge of bodily feelings or affects. Hence, in explaining the fact that grief fades as time passes by, Nussbaum ignores the aspects of *felt* urgency and the affective side of the emotional experience. She seems to admit this point when she says that “it is the emotion itself, and not some further reaction to it, that has urgency and heat” (Nussbaum 2001, 78). But she views this urgency as a cognitive feature for it “comes not from the unthinking force, but from my thought that my well-being is threatened by that force” (Nussbaum 2001, 78). For example, if there is urgency in being hit by a gust of wind, it is not after all a noncognitive urgency – the urgency, if it is there, comes not from the unthinking force, but from my thought that my well-being is threaten by that force (cf. Nussbaum 2001, 78). If this were right, as Peter Goldie notes, Nussbaum’s cognitivism fails to recognise that “how profoundly and systematically our emotional feelings can mislead us – how the emotions can distort perception and reason” (Goldie 2004, 91).

In order to appreciate the point, in what follows I demonstrate how affect or bodily feeling also plays an important role in self-knowledge of emotional changes. I will do this by taking an example from *Remembrance of Things Past* by Proust (1981).¹¹ When Marcel hears the news of Albertine’s death, he suffers great anguish. Up until this point he had believed that he did not love her. However, upon hearing the distressing news, he realizes with certainty that he did indeed love her. According to Proust, Marcel has deceived himself about his love for Albertine. If we hold, like the cognitivist, that recognition of our condition can be best explained by a detached unemotional and non-affective cognition, then we do not have the resources to explain Marcel’s shock and suffering. This intellectual judgment that he did not love Albertine is shattered by his anguish. Of Marcel’s self-deception, Proust writes:

I had been mistaken in thinking that I could see clearly into my own heart. But this knowledge, which the shrewdest perceptions of the mind would not have given me, had now been brought to me, hard,

¹¹ Nussbaum also discusses this example. See Nussbaum (1990).

glittering, strange, like a crystallised salt, by the abrupt reaction of pain.
(Proust 1981, 426)

Marcel's feeling of anguish is said to *feel*, 'like a thunderbolt', and forces his assent to his state of love (Proust 1981, 431). His state is revealed by his feelings, not his intellect. Proust tells us that our feelings are systematically buried by the workings of habit, the primary manner in which self-deception operates. The truth of Marcel's condition is brought out by his feelings. His painful anguish demonstrates his prior state of self-deception. If this is indeed a typical or common phenomenon, then it shows that a full account of our condition cannot be given simply by appeal to a cognitive-evaluative state. Confronted with this objection, Nussbaum might say that the relevant feelings are not bodily but can themselves be unpacked 'cognitively'. She claims that the occurrence of non-cognitive features in emotional experiences is contingent to the definition of emotions. Thus, Nussbaum would say that Marcel's feeling – anguish – wasn't just a physical sensation of pain that articulated his emotion, but the pain of a particular loss. She writes:

We also have a type of pain that probably is necessary for grief: namely, the pain *that an important element of one's life is gone*. But of course that is not a noncognitive element, and we have already included it in our cognitive/evaluative account, which has stressed, with Chrysippus, that such losses are bad and that it's right to be upset about them. (Nussbaum 2001, 64)

If this were true, she might say that Marcel's feelings, in our example, is psychological pain, which would yield to a purely 'cognitive' analysis. Yet if this were right, Nussbaum must explain what 'affect' is and whether it can be fairly contrasted with cognition. However, she remains unclear on this matter. Hence, I argue that the attempt by Nussbaum to reduce the role of cognition and affect in our recognition of emotional change to a single cognitive element is implausible, since both cognition and affect play an important role in our mental lives.

3.4. Self-knowledge of emotional change: from the past-present to the future

Now I shall discuss how experiential memory can play a role for illuminating a person's future. Memory is broadly understood as ability. Memory

as a kind of mental phenomenon has been divided into two categories by a number of philosophers: memory, on the one hand, is a mental state, on the other hand, a mental disposition.¹² What is the difference between a mental disposition and a mental state? A mental disposition differs from a mental state in that mental states are events, in that they can be ascribed a discrete temporal location, whereas mental dispositions are manifested intermittently over a certain period. Mental states are “episodic or transient phenomena”, whereas mental dispositions are persisting and they “manifest themselves intermittently” (Wollheim 1984, 34-35). A disposition explains events that occur in a person’s life and underlies mental states.

In order for my current mental state and the earlier experience to be connected some kind of causal connection is needed. But what kind of connection? Not just any causal connection; the current mental state and the past event have to be connected in the right way. This idea can be illuminated in terms of what Wollheim calls a ‘dependency relation’. What then is the ‘dependency relation’? If we see memory as a disposition as Wollheim does, it entails that it has ‘causal dependence’, surely on earlier events. Wollheim calls this the dependency relation. Thus there is a three term relation to be found in the orbit of experiential memory. First, “the original event causes the disposition to be”. Second, “the disposition causes the mental state to occur” (Wollheim 1984, 98).¹³ Third, the mediating disposition has the role of keeping an event “causally alive” (Wollheim 1984, 98). Here we can say that “causation occurs not just once but twice in the orbit of experiential memory, or that the memory onwardly transmits the causal influence of the remembered event” (Wollheim 1980, 308). Given that memory is a disposition, it exerts a causal influence over the manifestations of a person’s later life. Hence, we can explain a person’s future life in terms of memory as disposition. Having established that memories, as dispositions, have a backward and a forward looking dimension, Wollheim argues that a memory criterion of personal identity is not merely an

¹² In developing this idea I draw heavily on aspects of Wollheim’s philosophical psychology. See Wollheim (1984); and also Wollheim (1980).

¹³ It could be argued that for a disposition to occur upon some event there had to be some prior disposition that disposed me to develop a disposition upon facing a certain type of events. So after all, it is dispositions all the way down. I would like to thank for this point to an anonymous reviewer of *Organon F*. But I leave the metaphysical issue of what dispositions are for a later occasion.

‘indicative’ criterion: “it is creative in that, when it holds, it brings about or helps to bring about that for which it is a criterion” (Wollheim 1984, 109).

On this view, a person’s memories are not merely instances of the spatio-temporal connections of psychological states, in the way that the traditional mental connectedness proponents, including Locke and Hume, propose. According to Locke and Hume, memory is a kind of store of the occurrent psychological state. On this traditional view, memory has a retrospective form only. However, the view which I defend says that it is not the occurrent psychological states alone that are part of one’s psychological make-up, but underlying traits and dispositions as well. On this view, memory and the forward-looking dimension in the temporal unity of a person’s life are interwoven.¹⁴ The most distinctive feature of this forward-looking dimension is the emotional tone and richness of each person’s life, since as Byers notes, for Proust, “memory is the emotionally textured expanse or register of a person’s life, as without it the past would be empty, rather than emotionally nuanced, textured, and on occasion, devastating”, as in *Traumatic Stress and Disorder* (see Byers 2005, 175). This idea suggests that in a person’s life memory is deeply implicated in the forward-looking orientation or impetus of a person’s life. Now in order to understand this idea, the implication that memory has a point of view needs to be spelled out. This concerns my claim c), namely that when we remember an event experientially, it is remembered from a point of view and this point of view is represented within the memory.

The discussion which follows will show how memory has a perspective or a point of view in our emotional change. This idea will show us how experiential memory helps to create one’s future. The reason why the emotional change is possible is that memory embedded in feelings also has a perspective, or a forward-looking dimension. The idea that memory em-

¹⁴ I am indebted for this point to Philippa Byers. See Byers (2005), especially, Chapter 5, Section 5.5. Byers discusses the relation between memory and time drawing on Proust’s *In Search of Lost Time*. On Byers’s analysis of Proust, the time that memory recalls is time that memory recreates. According to Byers’s Proust, although the stuff of memory is the past, what is produced from the past, as the result of the labour of remembering, is new. Memory calls forth something new, rather than something repeated (cf. Byers 2005, 170). This view implies that memory not only recreates time but also has a future. In this respect, Byers criticizes the traditional memory-based theories of personal identity, such as Locke’s, where memory is characterized as presenting a retrospective form of unity.

bedded in feelings has a perspective can be shown by using an analogy with *perceptions*. We can compare a memory which has a perspective with the Muller-Lyer lines: the Muller-Lyer lines are known to be equal while they continue to appear to be of different lengths. Many people have used this example to illustrate how, at times, our perceptions seem to behave with a mind of their own despite our best efforts in trying to control them. This is because, the content of memory embedded in feelings, resembles the content of sense-perception in that both kinds of representational content may not be revised in the light of belief and better knowledge.

We can also apply this kind of analogy to the emotions as follows: just as our perception that the two lines differ in length persists in the faces of our belief that the lines are the same length, so some emotions may persist even though we have relevant and countervailing knowledge. Thus, we can say that a person can have a sense-perception or can feel emotion that *p*, and at the same time consciously believe that not-*p*. However, this is not the case in the case of contradictory beliefs. The point that I want to make here is that in the cases of emotion or experiential memory, we are able to be in a contradictory or ambivalent state while in the case of belief, this is impossible. This is because, we can say, both memory and emotion have a perspective.

If emotion and memory do not have a perspective, then we cannot understand how someone who has been in the grip of jealousy is eventually indifferent concerning the object of jealousy, for example. In order to appreciate this point, let us imagine the following example: Suppose that Kate wants to gain revenge on her ex-husband who had an affair with her friend, Lucy and left her. Kate is also jealous of Lucy, and wants to gain revenge on her as well. Kate knows that her ex-husband depended on her financially and also Lucy is impoverished. Kate desires to see him financially ruined so that he will regret leaving her. But one day Kate falls in love with a man whom she dreams of marrying. Then she realizes that she has become a person who is no longer obsessed with her ex-husband, and no longer jealous of Lucy.

Having established that emotion and experiential memory have a perspective, we can say that Kate's jealousy is comparable to the following cases: the snake is not dangerous and at the same time someone fears it; a person believes that the Muller-Lyre lines are equal while they continue to appear to be of different lengths. As I have said before, this implies that Kate's jealousy, for example, like sense-perception can present in the face of better

knowledge. She may know that Lucy is superior to her and more attractive than her. However, the emotion does not tend to get out of control, due to the passivity of feeling. In this respect, I argue that the emotion which does not vanish in the face of better knowledge is 'recalcitrant emotion' in the sense that to some extent our emotions and emotional responses are passive, and cannot be controlled.¹⁵ This is because the subject's attitude toward those involved in jealousy is very complex and the desire in which jealousy is involved depends on all kinds of things: the circumstances, character-traits and moods that the person has.

Now if emotion and memory does not also have a perspective, then we cannot understand how Kate who has been in the grip of jealousy is eventually indifferent concerning the object of jealousy. But from Kate's self-concern in her future, this emotional change is understandable. Her self-concern and self-love, which possess perspective, have forward-looking dimensions. If we adopt my claim a), *a remembered event is something that I have lived through*, then we can say that the person has been and will be persisting, since a person is only able to experience something on the basis of his or her past action or experiences and concern for his or her future life. The reason why she is psychologically continuous over time is because she is aware of the fact that her past actions or experiences, and her concern for her future life, influence her present condition. Hence we can say that the recovery from the obsession or jealousy, for example, due to her awareness of a future shows that she is a successor of a series of past embodied subjects, who are similar but not numerically identical to each other. If this is right, we can say that change does not imply the loss of identity, rather change reveals how one and the same person is obsessed and miserable and then becomes no longer obsessed and miserable. This view implies that the unity of a person's life includes changes, for our condition and experiences change.

In explaining change there are elements of both discovery and creation. Self-knowledge of emotional change involves discovering something new in oneself. Self-knowledge of something new is awareness of both the fact that the past is differentiated from the present, and the fact that the future is the new creation of the past and present. In this respect, we can say that experiential memories of someone recreate a dimension of her/his past life

¹⁵ The familiar example for emotional recalcitrance is phobia: the phobic who is afraid of something despite believing that it poses little or no danger.

and create her/his future life. She tends to remember the thoughts and the feelings that she had. Furthermore, her remembering drives her life towards the future. This shows us that experiential memory is not just concerned with the recollection of events or experiences that the subject has undergone, but concerns the narrative structure of one's life as a whole.

4. Conclusion

In this paper I have defended a psychological continuity theory of personal identity which claims that memory has as much to do with a person's future as with a person's past with reference to emotions. Specifically, on the view which I have defended, memory does not simply invoke past events, but creates dispositions for future actions, thoughts and emotions. In this respect, we can say that memory is both past and future-regarding.

In order to illustrate this idea, I have elaborated on the conception of 'experiential memory' suggested by Wollheim. The characteristics of experiential memory are: first, it is a memory of an event lived through; second, it involves remembering an event *as* I experienced it, intrinsic to which are affective elements, as opposed to the mere recall of facts; third, it is a memory of an event from a person's own point of view. Hence, when we experientially remember a past event with feeling and perspective, our remembering not only *recreates* a dimension of our past life but also drives our future life in a certain direction. If my argument is right, we can say that experiential memory is not just confined to the recalling of events or experiences that the subject has experienced, but concerns the narrative structure of a person's life as a whole.

One might say that the view which I have discussed implies a narrative conception of the self. What then is the narrative conception of the self? The proponent of a narrative thesis claims that autobiographical memory is an essentially constructive and reconstructive phenomenon rather than a merely reproductive one. Yet the problem with autobiographical memory is that it involves the possibility of revision when a person reconstructs his or her past life.¹⁶ However, if revision is necessary to being a narrative, a person's life which is reconstructed by revision is not a genuinely identical

¹⁶ Strawson (2004) presents the problem of the possibility of revision in his argument against the narrative thesis.

one. If this is true, I do not want to commit myself to the narrative theory. Whatever narrativity means by the unity of a life, it seems to me that it is not the same as personal identity but seems to presuppose it, since in order for us to stitch together the event of our life into a whole self we should already know which events those are. So some notion of personal identity is presupposed. Hence, in order to do justice to the narrativity thesis, narrative theorists need to respond to some facts which seem to be problematic for their account: for example, people can remember and relive only a small part of their lives, those things which happened to us when we were very young being forever inaccessible to most of us. I have attempted to meet this kind of objection by taking memory to be a disposition. However, if we adopt the narrative thesis, then we cannot avoid the following objection: there is an important difference between what one *actually* did at a particular time and a *story* that is made up to explain those events that they do remember. I leave this kind of difficulty open in this paper. Instead I want to suggest that as a philosopher, I'd suggest we go for an historical narrative for a past self while trying to be creative about a future self, leaving fiction for the novelists, since I am not convinced that I have been constructed. Autobiographies are not sufficiently reliable for philosophical account of identity.¹⁷

References

- AINSLIE, D. (2001): Hume's Reflections on the Identity and Simplicity of Mind. *Philosophy and Phenomenological Research* 62, 557-578.
- BEN-ZE'EV, A. (2004): Emotions are not Mere Judgments. *Philosophy and Phenomenological Research*, 450-457.
- BYERS, P. (2005): *Personal Identity: Temporality and Continuity*. PhD thesis, University of Sydney.
- BIRO, J. (1976): Self-Identity and Memory. *The Review of Metaphysics* 26.

¹⁷ This paper was presented in the conference of *Emotion, Others, and the Self* at Abo Akademi University, Turku, in August 2005. I am very grateful to the audience at the conference, especially, Lili Alanen, Kathy Behrendt, and Michael Lacewing for their useful comments. I also thank Andy Hamilton and Roger Squires, who suggested, in personal correspondence, some important improvements. Finally, I would like to thank Ronald De Sousa and Peter Goldie, who gave me an inspiration on Philosophy of Emotion.

- BUTLER, J. (1975): Of the Personal Identity. In Perry, J. (ed.): *Personal Identity*. California University Press.
- CAMPBELL, J. (1994): *Past, Space, and Self*. Cambridge, MA: MIT Press.
- DEONNA, J. A. (2006): Emotion, Perception and Perspective. *Dialectica* 60, 29-46.
- EVANS, G. (1982): *The Varieties of Reference*. Oxford: Clarendon Press.
- GOLDIE, P. (2004): Emotion, Feeling and Knowledge of the World. In: Solomon, R. (ed.): *Thinking about Feeling*. Oxford – New York: Oxford University Press.
- GOLDIE, P. (2012): *The Mess Inside: Narrative, Emotion and the Mind*. Oxford: Oxford University Press.
- HAMILTON, A. (1995): A New Look at Personal Identity. *Philosophical Quarterly* 45, 332-349.
- HAMILTON, A. (forthcoming): Memory and Body: A Study of Self-Consciousness.
- HUME, D. (1978): *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge, 2nd ed., Revised by P. H. Nidditch. Oxford: Clarendon Press.
- JAMES, W. (1890/1950): *The Principles of Psychology*. New York.
- LOCKE, J. (1975). *Essay Concerning Human Understanding*. Edited by Peter H. Nidditch. Oxford.
- LOWE, E. J. (1991): Substance and Selfhood. *Philosophy* 66, 81-99.
- LOWE, E. J. (1995): *Locke: On Human Understanding*. New York: Routledge.
- LOWE, E. J. (2000): Personal Identity and Self-Knowledge. In: *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.
- NOONAN, H. W. (1989): *Personal Identity*. Routledge.
- NUSSBAUM, M. C. (1990): *Love's Knowledge: Essays on Philosophy and Literature*. New York: Oxford University Press.
- NUSSBAUM, M. C. (2001): *Upheavals of Thought*. Cambridge: Cambridge University Press.
- PARFIT, D. (1987): *Reason and Persons*. Oxford: Oxford University Press.
- PENELHUM, T. (2000): *Themes in Hume: The Self, The Will, Religion*. Oxford: Clarendon Press.
- PERRY, J. (1975): *Personal Identity*. California University Press.
- PROUST, M. (1981): *Remembrance of Things Past*, Vol III. Transl. by Scott Moncrieff, C. K. – Mayor, A. – Kilmartin, T. (rev). London: Vintage.
- REID, T. (1975): Of Mr. Locke's Account of Our Personal Identity. In: Perry, J. (ed.): *Personal Identity*. California University Press.
- SCHECHTMAN, M. (1994): The Same and the Same: Two Views of Psychological Continuity. *American Philosophical Quarterly* 31, 199-212.
- SCHECHTMAN, M. (1997): *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- SHOEMAKER, S. (1963): *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.
- SHOEMAKER, S. (1984): Personal Identity: A Materialist Account. In: Shoemaker, S. – Swinburne, R. (eds.): *Personal Identity*. Oxford: Basil Blackwell.
- STRAWSON, G. (2004): Against Narrativity. *Ratio* 17, 428-452.

- WIGGINS, D. (1967): *Identity and Spatio-Temporal Continuity*. Oxford: Blackwell.
- WILLIAMS, B. (1973): *Problem of the Self*. Cambridge: Cambridge University Press.
- WOLLHEIM, R. (1979): Memory, Experiential Memory, and Personal Identity. In: McDonald, G. F. (ed.): *Perception and Identity: Essays Presented to A. J. Ayer with His Replies to Them*. London: Macmillan.
- WOLLHEIM, R. (1980): On Persons and Their Lives. In: Rorty, A. (ed.): *Explaining Emotion*. University of California Press.
- WOLLHEIM, R. (1984): *The Thread of Life*. Cambridge: Cambridge University Press.

Prečo externalistické kritériá zdôvodnenia princiipiálne zlyhávajú?

MARTIN NUHLÍČEK

Katedra filozofie a dejín filozofie. Filozofická fakulta Univerzity Komenského v Bratislave.
Gondova 2. 814 99 Bratislava. Slovenská republika
martin.nuhlicek@yahoo.com

ZASLANÝ: 25-10-2011 • AKCEPTOVANÝ: 08-07-2012

ABSTRACT: The article concerns two basic approaches to the problem of epistemic belief-justification: internalism and externalism. It aims to show that internalism as well as externalism, when confronted with the problem of philosophical skepticism, face various kinds of problems, which lead to implausibility of their respective accounts of justification. The author focuses especially on the externalist approach which was invented as a direct response to the threat of skepticism. The central part of the article contains a brief analysis of main attributes of externalism, and subsequently its criticism which aims to show that the criteria of justification proposed by externalists do not accomplish the basic function of distinguishing between justified and unjustified beliefs. The author argues that the discussed deficiencies of externalism result from its elementary rationale, which implies that they are incurable, and therefore the externalist criteria of justification inevitably fail.

KEYWORDS: Epistemology – externalism – internalism – justification – skepticism.

Známa tripartitná definícia vedenia hovorí, že vedenie je presvedčenie, ktoré je pravdivé a zároveň zdôvodnené (pozri napr. Ayer 1956, 28-34). Podmienky presvedčenia a pravdivosti sú triviálne – nie je možné vedieť to, o čom nie sme presvedčení, alebo to, čo nie je pravda – preto tieto dve podmienky v definícii vedenia obyčajne nie sú predmetom vážnejších epistemologických diskusií. Inak je to pri podmienke zdôvodnenia. Filozofi sa

zhodujú, že táto podmienka má slúžiť na vylúčenie náhodne pravdivých presvedčení (presvedčení, ku ktorým sme dospeli nesprávnym spôsobom, ale napriek tomu náhodou vyjadrujú pravdu). No pri otázke, v čom zdôvodnenie spočíva a aké sú jeho kritériá, sa názory filozofov markantne rozchádzajú.

Predmetom môjho príspevku je kritické analyzovanie dvoch základných prístupov k podmienke zdôvodnenia presvedčení, ktorými sú internalistický a externalistický prístup. Cieľom môjho príspevku je ukázať, že v konfrontácii s problémom skepticizmu internalizmus aj externalizmus upadajú do principiálnych, t. j. dodatočnými teoretickými úpravami neodstrániteľných problémov, ktorých dôsledkom je neplauzibilitasť jedného i druhého prístupu pri opisovaní zdôvodnenia. Špecificky sa pri tom zameriam na kritiku externalistického prístupu, ktorého proponenti ho prezentujú ako originálne a výnimočne účinné riešenie epistemologických otázok.

Najskôr pár poznámok o zdôvodnení. V prirodzenom jazyku sa predpokladá, že ak je niečo dobre zdôvodnené, tak existujú nejaké dostatočne presvedčivé dôvody v prospech danej veci. Zdôvodnenie¹ teda súvisí s dôvodmi: Zdôvodnené presvedčenie je také presvedčenie, ktoré je podložené nejakými vhodnými dôvodmi. Ako dôvod môže slúžiť čokoľvek, čo „podopiera“ naše presvedčenie; čokoľvek, čo sme schopní a ochotní chápať ako evidenciu² v prospech obhajovanej veci. Ako dôvod môže slúžiť iné presvedčenie, logický argument, jednoduché zmyslové vnemy, informácie od iných ľudí atď. Na základe povedaného môžeme formulovať definíciu, že dôvod je čokoľvek, na základe čoho subjekt³ zastáva presvedčenie, že to a to sa má tak a tak, a na čo sa subjekt môže odvolať ako na evidenciu v prospech svojho presvedčenia.

Bežné nefilozofické intuície, ale aj tradičná (z karteziánu vychádzajúca) epistemológia obsahujú predpoklad, že zdôvodnenie presvedčení je niečo

¹ Terminologická poznámka: *Zdôvodnenie* je dôvod alebo súbor dôvodov, ktoré hovoria v prospech daného presvedčenia. *Zdôvodňovanie* je proces, ktorým presvedčenie nadobúda zdôvodnenie. Napokon, *zdôvodnené* presvedčenie je presvedčenie, ktoré už nadobudlo zdôvodnenie.

² Pre jednoduchosť budem v ďalšom predpokladať, že neexistuje striktný epistemický rozdiel medzi „dôvodom“ a „evidenciou“. Budem predpokladať, že „byť dôvodom v prospech presvedčenia“ a „byť evidenciou v prospech presvedčenia“ sú (zhruba) totožné epistemické vlastnosti.

³ V súlade s epistemologickou tradíciou budem výrazom „subjekt“ označovať nositeľa presvedčení resp. vedenia.

subjektívne, relativizované vzhľadom na konkrétny subjekt. Každý subjekt má svoje vlastné dôvody, ktoré podporujú jeho presvedčenia a pre ktoré ich zastáva. Nie je vylúčené, že dôvody jedného subjektu si môže osvojiť druhý subjekt – nakoniec, v tom je podstata presvedčania a presadzovania názorov – ale vždy sú to konkrétne dôvody konkrétneho subjektu, ktorý ich predkladá, používa, odvoláva sa na ne a pod.

Takéto uvažovanie nás privádza k inému, v tradičnej epistemológii bežne obsiahnutému predpokladu, podľa ktorého dôvody, z ktorých pozostáva zdôvodnenie, sú niečím, k čomu má subjekt *intelektuálny prístup* z hľadiska svojej kognitívnej perspektívy.⁴ Podľa tohto chápania platí, že evidencia, ktorá nie je pre daný subjekt prístupná, nemôže zdôvodňovať presvedčenia daného subjektu. Vďaka tomu, že tento prístup k zdôvodneniu chápe dôvody ako niečo interné vzhľadom na subjekt, presnejšie ako niečo, čo je vždy (aspoň potenciálne) obsahom mysle subjektu, obyčajne sa označuje ako *internalizmus*.⁵ Internalistické teórie zdôvodnenia presvedčení sú teórie, ktoré obsahujú požiadavku oboznámenosti subjektu s podmienkami zdôvodnenia; jednoduchšie povedané, sú to teórie, ktoré pre zdôvodnenie (a nepriamo aj pre vedenie) vyžadujú, aby subjektu boli známe dôvody v prospech jeho presvedčení, ako aj spôsob, akým tieto dôvody podporujú dané presvedčenia.⁶

Existujú kritické hlasy, spochybňujúce plauzibilitu internalistického prístupu k zdôvodneniu na základe jeho (údajnej) neschopnosti poradiť si s *problémom skepticizmu*. Problém skepticizmu spočíva v existencii špecifických filozofických argumentov, ktoré smerujú k tomu, že ani najlepšie subjektívne zdôvodnenie presvedčení objektívne nezaručuje akýkoľvek relevantný súvis presvedčení s pravdou. Skeptická argumentácia sa opiera o možnosť tzv. skeptických scenárov, t. j. hypotetických situácií, v ktorých sú subjekty zásobované nepravdivými vnemami a presvedčeniami pochádza-

⁴ Kognitívna perspektíva subjektu je súbor všetkých epistemických informácií (t. j. presvedčení, vnemov, argumentov atď.), ktorými subjekt aktuálne disponuje, alebo ktorými je schopný potenciálne disponovať – t. j. získať ich napr. pomocou odvodenia, introspekcie, rozpamätania sa a pod. z toho, čím disponuje aktuálne.

⁵ Označenie *internalizmus* pre tradičný spôsob uvažovania o zdôvodnení sa zaužívalo spätne až v konfrontácii s novším, netradičným prístupom k zdôvodneniu, nazývaným externalizmus.

⁶ Nie je nevyhnutné od každého subjektu vyžadovať podrobnú znalosť epistemických vzťahov a procesov; postačujúci môže byť aj určitý laický stupeň oboznámenosti s podmienkami zdôvodnenia – napr. v podobe schopnosti odpovedať na otázku: Odkiaľ, resp. ako vieš to, čo tvrdíš, že vieš?

júcimi z mylného zdroja. Asi najznámejšími skeptickými scenármi sú Descartov scenár opisujúci zlého démona s nadprirodzenou mocou nad našimi mysľami (Descartes 1997, 21-26) alebo Putnamov scenár opisujúci mocného vedca ovládajúceho naše mozgy pomocou rozvinutej technológie (Putnam 1981, 5-8). Uvedené scenáre vykresľujú spôsoby, akými môžu vznikáť naše vnemy a presvedčenia v rozpore s bežne predpokladanými spôsobmi poznávania. Dôležitou okolnosťou je to, že hoci vytváranie vnemov a presvedčení v rámci skeptických scenárov je umelé, tieto vnemy a presvedčenia sú na nerozoznanie od pravých vnemov a presvedčení. Skeptické scenáre sú programovo „dizajnované“ tak, aby sa z hľadiska subjektu nedalo rozoznať, či ide o pravý alebo mylný podnet.⁷ Skepticizmus ako filozofický postoj však netvrdí, že skeptický scenár je aktuálne pravdivý – na to by nemal dobré dôvody – ale presadzuje slabšie tvrdenie, že nedokážeme vylúčiť, že skeptický scenár je pravdivý. V dôsledku toho pri ktoromkoľvek našom presvedčení nemáme epistemickú istotu, že *nepochádza* zo skeptického zdroja, kvôli čomu nemôžeme legitímne tvrdiť, že je pravdivé, a teda, že tvorí vedenie. Skepticizmus vedie k záveru, že máme oveľa menej vedenia, ako bežne predpokladáme – ak vôbec nejaké vedenie máme.⁸

Skepticizmus predstavuje pre internalizmus vážnu hrozbu vzhľadom na to, že v rámci internalizmu je zdôvodnenie presvedčení viazané na konkrétny subjekt a jeho kognitívnu perspektívu. Subjekt zo svojej osobnej kognitívnej perspektívy principiálne nedokáže vylúčiť možnosť skeptického klamu pri ktoromkoľvek zo svojich presvedčení – pripomínam, že skeptické scenáre sú „dizajnované“ práve tak, aby to subjekt nemohol dokázať. Preto ak chápeme zdôvodnenie internalisticky, tak kvôli nevylúčiteľnosti skeptických scenárov platí, že ani najlepšie možné subjektívne dôvody dostatočne nezaručujú správnosť našich presvedčení. Presvedčenia síce môžu byť subjektívne zdôvodnené, ale to logicky a dokonca ani pravdepodobnostne nezaručuje, že sú objektívne pravdivé, a teda, že tvoria vedenie. Internalistické chápanie zdôvodnenia je teda kritizované najmä za to, že nedokáže relevantne prispieť k identifikovaniu pravého, nenáhodného vedenia.

⁷ Pri pokuse overiť pravosť podnetov nepomôže ani žiadne dodatočné skúmanie alebo hľadanie evidencie, pretože akýkoľvek dodatočný test môže byť tiež „dielom“ zlého démona, resp. mocného vedca.

⁸ Skeptici tradične spochybňujú vedenie o vonkajšom svete, minulosti, budúcnosti a iných myšliach. Naopak, rozumové alebo introspektívne vedenie spochybnili len výnimočne.

Od konca 60. rokov 20. storočia sa začali objavovať názory, ktorým na tradičnom internalistickom prístupe vážne prekážala slabá previazanosť zdôvodnenia presvedčení s ich pravdivosťou. Začali presadzovať úvahu, že ak by zdôvodnenie nebolo celkom osobnou, subjektívnou vecou, ale uviedlo by sa do užšieho súvisu s pravdivosťou presvedčenia, tak by sa eliminovala možnosť mať presvedčenia, ktoré síce formálne splňajú podmienky zdôvodnenia, ale sú objektívne nesprávne. Zdôvodnenie by tak dokázalo účinne plniť úlohu indikátora pri selektovaní správnych presvedčení. Priekopníkom a najvýznamnejším zástancom tohto typu uvažovania je Alvin Goldman (1979, 1986).

Naznačený prístup radikálne zmenil dovedty zaužívané chápanie zdôvodnenia. Zatiaľ čo tradičný internalizmus zdôrazňuje podmienku angažovanosti subjektu⁹ pri zdôvodňovaní presvedčení, spomínaný novší prístup hovorí, že presvedčenie je zdôvodnené, keď splňa určité objektívne podmienky. Hlavná zmena oproti internalizmu spočíva v tom, že podmienky zdôvodnenia nemusia byť nevyhnutne prístupné z hľadiska kognitívnej perspektívy subjektu. Faktory pravdivosti a objektívnej správnosti, o ktoré novší prístup opiera podmienku zdôvodnenia presvedčení, sú vzhľadom na subjekt externé.¹⁰ Ak zdôvodnenie závisí od externých, vzhľadom na subjekt nie interných faktorov, tak prístupnosť samého zdôvodnenia nie je z hľadiska subjektu samozrejma. Najmä kvôli týmto charakteristikám sa pre tento nový prístup k zdôvodneniu zaviedlo pomenovanie *externalizmus*.

Externalistické teórie zdôvodnenia presvedčení sa vyznačujú dvomi hlavnými črtami:

1. Podmienky zdôvodnenia sú stanovené tak, že ich splnenie zabezpečuje, že zdôvodnené presvedčenie má tendenciu byť pravdivé.
2. Od subjektu sa nevyžadujú žiadne špeciálne vedomosti o podmienkach zdôvodnenia alebo o zdôvodňovacích vzťahoch.

⁹ Angažovanosť subjektu pri zdôvodňovaní sa dá chápať dvojako: buď ako „technická“ angažovanosť, t. j. tak, že subjekt je nositeľom a držiteľom dôvodov, alebo ako „morálna“ angažovanosť, t. j. ako úsilie subjektu vyhýbať sa nezdôvodneným, resp. pochybným zdôvodneným presvedčeniam.

¹⁰ Uvedené faktory sú externé v tom zmysle, že sú nezávislé od presvedčení, názorov a postojov subjektu.

Externalisti predpokladajú, že na to, aby sme mohli mať zdôvodnené presvedčenia, nemusíme byť špecialistami na otázky kognitívnych mechanizmov. Nemusíme mať žiadne špeciálne presvedčenia o fungovaní kognitívnych procesov¹¹ – na zdôvodnenie presvedčenia stačí to, aby procesy, v ktorých má zdôvodňované presvedčenie pôvod, fakticky fungovali správne. To je v rámci externalizmu kľúčový bod: Na zdôvodnenie presvedčenia nie je rozhodujúca subjektívna schopnosť uvádzať podporné dôvody (ako pri internalizme), ale objektívna správnosť spôsobu, akým presvedčenie vzniklo a akým súvisí s tým, čoho sa týka.

Najznámejším príkladom externalistickej teórie zdôvodnenia je reliabilizmus. Reliabilizmus pracuje s pojmom spoľahlivosti u kognitívnych procesov: Presvedčenie chápe ako zdôvodnené, ak je produktom spoľahlivého kognitívneho procesu, čiže procesu, ktorý má tendenciu produkovať pravdivé presvedčenia (o reliabilizme bližšie pozri napr. Goldman 1979).¹² Atraktívnou črtou reliabilistického prístupu je to, že poskytuje jednoduchý nástroj na identifikovanie správnych presvedčení: Ak je kognitívny proces, v ktorom má pôvod skúmané presvedčenie subjektu, spoľahlivý, tak dané presvedčenie môžeme označiť ako zdôvodnené, t. j. pravdepodobne pravdivé.¹³ (Ďalšie podrobnosti konkrétnych externalistických teórií zdôvodnenia nie sú pre cieľ tohto príspevku dôležité.)

Prečo externalistické kritériá zdôvodnenia principiálne zlyhávajú? Pri odpovedi na centrálnu otázku z názvu tejto state chcem obhajovať názor, že uplatnenie externalistických kritérií a podmienok zdôvodnenia vedie k skeptickému záveru, že pri žiadnom presvedčení nie je možné určiť, či je zdôvodnené alebo nezdôvodnené. Keďže zdôvodnenie je jednou z podmienok vedenia v rámci tripartitnej definície vedenia, tak z toho vyplýva ďalší záver, podľa ktorého pri žiadnom presvedčení nie je možné určiť, či je ve-

¹¹ Kognitívne procesy sú všetky mentálne procesy, ktoré vedú k vzniku presvedčení; napr. zmyslová percepcia, rozpamätávanie sa, usudzovanie a ďalšie.

¹² Reliabilizmus sa medzi externalisticky zameranými epistemológmi rozšíril natoľko, že v súčasnosti pre viacerých splyva so samým externalizmom. (Ani ja sa v tomto príspevku tomuto javu celkom nevyhnem.)

¹³ Reliabilisticky zdôvodnené presvedčenie nie je automaticky pravdivým presvedčením. Spoľahlivý kognitívny proces zaručuje iba toľko, že presvedčenia, ktoré sú jeho výsledkom, majú *tendenciu* byť pravdivé. To nevyklučuje ojedinelé prípady nepravdivých presvedčení. Z hľadiska subjektu je to však dostatočný indikátor správnosti presvedčení (neomylné kritérium by sa zrejme hľadalo veľmi ťažko).

dením alebo nie je vedením. Ak teória zdôvodnenia vedie k stavu, že pri akoľvek presvedčení sa zdôvodnenie vôbec nedá identifikovať, môžeme to nazvať jej zlyhaním. Chcem ukázať, že toto zlyhanie je principiálne, to znamená, že sa mu nedá vyhnúť úpravou vybraných bodov teórie, ale vyplýva zo samej podstaty externalizmu.

Primárny zdroj problémov spočíva v už spomenutej externalistickej téze, podľa ktorej na to, aby mal subjekt zdôvodnené presvedčenie, nemusí mať žiadne špecifické vedomosti o fungovaní svojich kognitívnych procesov – na zdôvodnenie údajne stačí, aby dotknuté procesy fakticky fungovali správne. Táto problematickosť sa ukáže, ak si položíme otázku, akým spôsobom je v externalistických podmienkach možné určiť, či je ľubovoľné presvedčenie zdôvodnené alebo nezdôvodnené.¹⁴ V súlade s predpokladom, z ktorého vychádza externalizmus, subjekt nemusí byť schopný identifikovať odpoveď na túto otázku. Pri skúmaní zdôvodnenia presvedčení (a odvodené aj vedenia) je však dôležité mať nejaký spôsob, algoritmus, umožňujúci získať informáciu o zdôvodnení alebo nezdôvodnení presvedčení. V tejto súvislosti externalizmus ponúka možnosť, aby spoľahlivosť kognitívnych procesov subjektu, od ktorých závisí zdôvodnenie jeho presvedčení, preskúmali špeciálni vedci, ktorí môžu vyhodnotiť korelácie presvedčení subjektu s realitou a poskytnúť odpoveď. Externalizmus totiž predpokladá, že to, ktoré procesy sú fakticky spoľahlivé, je celkom kontingentné. Kompetencie v oblasti skúmania a rozhodovania o epistemických procesoch preto presúva z filozofov na špeciálnych vedcov – filozofi pracujúci „z kresla“ totiž neboli vyškolení v disciplínach, ktoré externalisti považujú za rozhodujúce na tieto účely, ako sú napr. neurofyziológia alebo psychológia (viac k tejto téme pozri Fumerton 1995, 171-173).

S touto možnosťou sa však spája vážna ťažkosť. Predstavme si, ako by naznačené skúmanie zdôvodnenia vyzeralo prakticky. Majme subjekt X, u ktorého máme záujem preskúmať správnosť niektorého z jeho presvedčení, napríklad preto, lebo od neho závisí dôležité rozhodnutie. Oslovíme preto špeciálneho vedca Y, aby preveril spoľahlivosť kognitívneho procesu u X, ktorý viedol k skúmanému presvedčeniu. Detaily vedcovej práce nechajme bokom; dajme tomu, že vedec Y vo výsledku vyhlási, že skúmané presvedčenie subjektu X je zdôvodnené. Tento postup však obsahuje problém: Vedec Y je totiž tiež samostatným poznávajúcim subjektom, ktorý používa

¹⁴ Externalistické podmienky budem demonštrovať na príklade reliabilizmu; predpokladám, že môj postup môže byť inštruktívny aj pre iné externalistické teórie.

svoje vlastné kognitívne procesy pri tvorbe svojich vlastných presvedčení. Ak má byť vedec Y dôveryhodný pri posudzovaní presvedčení subjektu X, tak musí byť jasné, že tvorba príslušných presvedčení u Y prebieha správne. Na takéto zistenie, či si Y utvára zdôvodnené presvedčenia, ako už vieme, externalizmus umožňuje osloviť ďalšieho špeciálneho vedca, Z, ktorý preskúma kognitívne procesy a presvedčenia u vedca Y. Tým sa však Z dostane do rovnakej pozície ako predtým Y, v ktorej bude potrebné mať informáciu, podľa ktorej príslušné presvedčenia vedca Z, týkajúce sa vedca Y, si Z utvoril fakticky správne. Do tejto pozície sa zároveň automaticky dostane každý ďalší overujúci výskumník na každej ďalšej úrovni. Vzniká tak nekonečný regres, ktorého dôsledkom je to, že nikto nemá dôveryhodné vedomosti o zdôvodnení kohokoľvek presvedčení – hoci nie je vylúčené, že skúmané presvedčenia môžu byť „fakticky“ zdôvodnené.

Externalizmom ponúkaný pokus skúmať zdôvodnenie presvedčení z hľadiska tretej osoby teda vedie k praktickej nerealizovateľnosti algoritmu, ktorým by bolo možné získať informáciu, či sú ľubovoľné presvedčenia ľubovoľného subjektu zdôvodnené alebo nezdôvodnené.¹⁵ Uvedený problém budem nazývať *problém tretej osoby*.

Niektorí zástancovia externalizmu naznačujú cestu, ako sa vyhnúť opísanému regresu, ktorá spočíva v možnosti overiť spoľahlivosť svojich vlastných kognitívnych procesov z hľadiska samého subjektu. Úvaha, ktorá otvára túto možnosť, je približne nasledovná: Ak je externalizmus pravdivý a ak je skúmaný kognitívny proces spoľahlivý, tak je možné s využitím tohto procesu pozorovať relevantné fakty a ich korelácie s utváranými presvedčeniami. Takýmto spôsobom je možné overiť spoľahlivosť kognitívnych procesov aj bez problematického zapojenia iných subjektov (bližšie o tejto úvaha pozri napr. Fumerton 1995, 173-177).

Tradične zmýšľajúci epistemológ však na tomto mieste spozornie alebo, ako hovorí R. Fumerton, priam začne zúriť:

Na zdôvodnenie spoľahlivosti percepcie nemožno použiť percepciu! Na zdôvodnenie spoľahlivosti pamäti nemožno použiť pamäť! Na zdôvod-

¹⁵ Pridruženým problémom je bludný epistemický kruh, ktorý by vznikol, keby špeciálny vedec overoval spoľahlivosť percepcie nejakého subjektu: Výsledky overenia by daný subjekt vnímal zrakom, sluchom a pod., teda tým, čo by ešte len malo byť overené. Sám kognitívny proces, ktorého spoľahlivosť by sa práve mala zistiť, by používal na získanie informácie, že je spoľahlivý.

nenie spoľahlivosti indukcie nemožno použiť indukciu! Takéto pokusy o odpoveď na skeptikovu výzvu obsahujú do očí bijúcu, skutočne trápnu kruhovosť. (Fumerton 1995, 177)

Nie je to nepodobné situácii, v ktorej by chcel astrológ dôkaz, že astrológia je spoľahlivá, vyčítať z hviezd, alebo situácii, kedy by človek sám sebe vydával úradné potvrdenia.

Externalisti sa proti výčitke kruhovosti bránia, že kritici nesprávne chápu spomenutú kruhovosť a vyvodzujú z nej nezamýšľané dôsledky. J. Greco pripúšťa, že ak by subjekt pri overovaní spoľahlivosti kognitívneho procesu vopred predpokladal ako premisu, že jeho proces je spoľahlivý, tak s určitosťou by išlo o neprijateľné kruhové usudzovanie. No obratom podotýka, že externalisti nič také nerobia: O kognitívnych procesoch nič vopred nepredpokladajú, len umožňujú, aby ich subjekt používal pri dosahovaní kognitívnych cieľov. Vychádzajú totiž z toho, že pri odpovedaní na akékoľvek otázky, vrátane otázok o poznaní, nemáme inú možnosť, ako využívať tie spôsoby myslenia a poznávania, ktoré nám boli dané do vienka (bližšie pozri Greco 2000, 184-187). Externalisti sa nazdávajú, že kruhovosť obsiahnutá v ich úvahách nie je zhubná, je v súlade so spôsobom, ako pracujeme s informáciami v bežnom živote a je podstatne lepším riešením než všeobecný skepticizmus.¹⁶

Napriek tomu, že externalistická argumentácia nie je celkom nezaujímavá, zdá sa, že predložený návrh, podľa ktorého subjekt môže sám overovať spoľahlivosť svojich kognitívnych procesov, je predsa len neefektívny. „Samooverovanie“ totiž nie je subjektu nápomocné pri rozlišovaní spoľahlivých a nespoľahlivých procesov. Predstavme si, že určitý kognitívny proces, ktorý je fakticky nespoľahlivý, by z hľadiska subjektu sám seba ohodnotil ako spoľahlivý – mohlo by sa to stať práve preto, že je fakticky nespoľahlivý. V externalistických líniách uvažovania by sme takto „overený“ proces mali uznať ako spoľahlivý, čo je však zjavne neprijateľné. Externalistický „samooverovací“ návrh teda nevedie k požadovanému cieľu, ktorým je rozlíšenie

¹⁶ Vymenované charakteristiky chápu externalisti ako prednosti svojej teórie vďaka špecifickej metodológii, s ktorou pristupujú k skúmaniu poznania. Hlavným zreteľom ich prístupu je vyhnutie sa skeptickým prvkom, s ktorými permanentne zápasí internalizmus. Dobrá teória má podľa externalistov spĺňať tri základné kritériá: 1. Má byť v súlade s našimi predteoretickými intuíciami ohľadom vedenia; 2. má byť imúnna voči skepticizmu (pretože ani naše intuície nie sú skeptické); 3. má byť psychologicky plauzibilná (bližšie pozri Greco 2000, 15-18).

zdôvodnených a nezdôvodnených presvedčení. Tento problém budem označovať ako *problém prvej osoby*.

Ďalším problémovým bodom externalizmu, na ktorý chcem upozorniť, je *kondicionálny tvar* niektorých kľúčových formulácií. Vráťme sa napríklad k externalistickej úvahe o možnosti „samooverovania“ spoľahlivosti kognitívnych procesov zo strany subjektu samého: *Ak* je externalizmus pravdivý a *ak* je skúmaný kognitívny proces spoľahlivý, tak je možné s využitím tohto procesu pozorovať relevantné fakty a ich korelácie s utváranými presvedčeniami. Zaiste, *ak* sú fakticky splnené podmienky, že externalizmus je pravdivý a že daný kognitívny proces je spoľahlivý, tak subjekt pomocou neho samého skutočne môže správne overiť spoľahlivosť tohto procesu. Očividným problémom však je to, že *ak* sa subjekt pýta na spoľahlivosť niektorého svojho kognitívneho procesu, tak prirodzene nedokáže vopred povedať, či je splnená podmienka, že tento proces je fakticky spoľahlivý. Napokon, práve nevedomosť subjektu v tejto veci je zrejším dôvodom, prečo sa na ňu pýta. Vzniká tu primitívny kruh: Na to, aby subjekt dokázal odpovedať na otázku, či je určitý jeho kognitívny proces spoľahlivý, musel by už vopred poznať odpoveď na túto otázku. Kondicionálna formulácia v tvare „*ak – tak*“ teda vonkoncom nedokáže pomôcť pri identifikovaní spoľahlivých procesov z hľadiska subjektu. Táto ťažkosť je súčasťou problému prvej osoby.

Pregnantné zhrnutie problémových bodov externalizmu ponúka B. Stroud. Upozorňuje, že plauzibilná teória zdôvodnenia, resp. vedenia musí byť všeobecná; čiastkové vysvetlenia, vzťahujúce sa len na vybrané subjekty alebo vybrané okolnosti, nie sú tým, k čomu sa v epistemológii pokúšame dospieť (Stroud 2000, 111-112). Hlavným problémom externalistických koncepcií je to, že tento cieľ principiálne nedokážu dosiahnuť. Opisujú síce zdôvodnenie a vedenie u iných ľudí, ale nevyžadujú, a preto ani neposkytujú dôvody (ako sme videli pri probléme tretej osoby a probléme prvej osoby) v prospech presvedčení z hľadiska prvej osoby. V konečnom dôsledku to vedie k nepríjemnému záveru, že sami teoretici externalizmu, ktorí sa na teórie pozerajú z hľadiska prvej osoby, nemajú vyhovujúce dôvody na akceptovanie svojich vlastných teórií. No tým externalizmus zlyháva vo svojom epistemologickom poslaní, pretože osvojením si externalistického prístupu „nebudeme bližšie k pochopeniu, ako je možné naše vlastné vedenie“ (Stroud 2000, 112).

Proponentovi externalizmu nič nebráni v tom, aby zrekonštruoval subjektívne dôvody, ktoré má v prospech svojej teórie. Za predpokladu, že ex-

ternalizmus je pravdivý a že príslušné kognitívne procesy, ktorých pomocou externalista rekonštruuje svoje dôvody, pracujú spoľahivo, to bude prijateľné vysvetlenie. Lenže toto externalistovo subjektívne „vysvetlenie“ nevylučuje možnosť, že sa fakticky mýli a jeho procesy pracujú nespoľahivo. Opäť teda vzniká primitívny kruh: Na to, aby subjekt mohol v externalistických líniách uvažovania dôveryhodne zdôvodniť, že externalizmus je pravdivý, musel by už vopred vedieť, že externalizmus je pravdivý.¹⁷ Stroud preto uzatvára, že *externalisti nerozumejú svojmu vlastnému vedeniu* (vrátane vedenia o externalizme), čo naznačuje, že externalizmus je teoretickou slepou uličkou. Stroud ústami fiktívneho externalistu hovorí:

Neviem, či rozumiem ľudskému vedeniu alebo nie. Ak je to, čomu verím, pravdivé a moje presvedčenia o tom vznikajú spôsobom, ktorý je podľa mojej teórie správny, tak viem, ako nastáva ľudské vedenie, a v tomto zmysle mu rozumiem. Ak však moje presvedčenia nie sú pravdivé alebo nie sú utvárané správnym spôsobom, tak vedeniu nerozumiem. Kladiem si otázku, ktoré z toho platí. Kladiem si otázku, či rozumiem ľudskému vedeniu alebo nie. (Stroud 2000, 119)

Kvôli problému, ktoré som pomenoval ako problém 3. osoby a problém 1. osoby platí, že v rámci externalizmu principiálne nie je možné dôveryhodne identifikovať zdôvodnenie akýchkoľvek presvedčení. Niektorí autori argumentujú, že ak je nemožné vhodným spôsobom preukázať, že niektoré presvedčenie je zdôvodnené, tak mu patrí status nezdôvodneného presvedčenia (pozri napr. McGrew – McGrew 2007, 73 a nasl.). Títo autori používajú na opísanie stavu, ku ktorému nás privádza externalizmus, výraz *epistemická anarchia*. Charakterizujú ju ako stav, v ktorom (na úrovni teórie) „v princípe neexistuje žiadny spôsob, ako jasne odlíšiť pravé epistemické princípy od absurdných epistemických princíпов, a teda žiadny spôsob, ako jasne odlíšiť presvedčenia, ktoré majú pozitívny epistemický status, od tých, ktoré ho nemajú“ (McGrew – McGrew 2007, 82). Opísanú principiálnu neschopnosť externalizmu z hľadiska subjektu diferencovať správne a nesprávne presvedčenia považujem za hlavný dôvod jeho teoretického zlyhania.

¹⁷ Stroud porovnáva kruhové usudzovanie externalistov so známou kruhovou úvahou, ktorú nájdeme u Descarta: Boh garantuje, že naše poznávacie schopnosti nás nezavádzajú, a teda keď ich správne použijeme, môžeme pomocou nich podať dôkaz Božej existencie (bližšie pozri Stroud 2000, 112-113).

Ak sa vrátíme k úvodu tohto článku, môžeme si pripomenúť, že podmienka zdôvodnenia presvedčení, ktorú obsahuje tripartitná definícia vedomia, slúži primárne na vylúčenie náhodne pravdivých presvedčení a v širšom zmysle má prispievať k selekcii správnych, podložených, k pravde smerujúcich presvedčení. Pokúsil som sa ukázať – za predpokladu, že neexistuje tretia cesta okrem internalizmu a externalizmu¹⁸ – že ani jeden z uvedených prístupov k zdôvodneniu nedokáže splniť úlohu, ktorá sa od zdôvodnenia vyžaduje. Internalizmus ostáva bezradný zoči-voči skeptickým hypotézam, pretože zo subjektívnej kognitívnej perspektívy nemáme nástroje, ako overiť pravosť zdôvodňujúcich faktorov. Externalizmus prichádza so sľubnou myšlienkou, že zdôvodnenie by malo priamočiarejšie zabezpečovať objektívnu správnosť presvedčení, ale návrhy na realizáciu tejto myšlienky sú z praktického hľadiska nerealizovateľné. Externalizmus v konečnom dôsledku trpí podobným nedostatkom ako internalizmus: Neposkytuje dostatočné prostriedky na to, aby sme dokázali s istotou identifikovať, či sú naše presvedčenia (v širokom zmysle) správne alebo nie. Ako som ukázal, ide o principiálnu, neodstrániteľnú chybu, ktorá vyplýva z vlastnej povahy externalizmu.

Záverom článku je konštatovanie, že ak sa pokúšame na problematiku zdôvodnenia presvedčení nahliadať prostredníctvom dichotómie internalizmus – externalizmus, tak výsledkom je to, že neexistuje žiadna dôveryhodná perspektíva, z ktorej by mohol ktokoľvek posudzovať kohokoľvek presvedčenia, vrátane vlastných presvedčení subjektu. Tým nie je vylúčené, že možno máme množstvo „fakticky“ správne utvorených presvedčení – ibaže kým sme „odstrihnutí“ od možnosti identifikovať, či ich skutočne máme a ktoré sú to, táto informácia nám nijako nepomôže pri epistemologickom úsilí porozumieť zdôvodneniu a vedeniu. Z pertraktovaných teoretických pokusov uchopiť a vhodne opísať problematiku zdôvodnenia nie je teda možné ani jeden hodnotiť ako úspešný – o čom napokon svedčí aj fakt, že žiadny z týchto pokusov sa doposiaľ nedočkal všeobecnej akceptácie.¹⁹

¹⁸ Existujú hybridné teórie, ktoré spájajú vybrané internalistické prvky s vybranými externalistickými prvkami. Som však presvedčený (bez toho, aby som mal priestor to bližšie vysvetliť), že arbitrárne spájanie častí jedného s časťami druhého neprekonáva, ale naopak kumuluje problémy oboch prístupov do jednej teórie.

¹⁹ Príspevok vznikol s podporou grantu UK G-11-087-00.

Literatúra

- AYER, A. (1956): *The Problem of Knowledge*. London: MacMillan.
- DESCARTES, R. (1997): *Meditácie o prvej filozofii*. Bratislava: Chronos.
- FUMERTON, R. (1995): *Metaepistemology and Skepticism*. Lanham: Rowman & Littlefield Publishers.
- GOLDMAN, A. (1979): What Is Justified Belief? In: Pappas, G. (ed.): *Justification and Knowledge*. Dordrecht: D. Reidel, 1-23.
- GOLDMAN, A. (1986): *Epistemology and Cognition*. Cambridge (MA): Harvard University Press.
- GRECO, J. (2000): *Putting Skeptics in Their Place*. Cambridge: Cambridge University Press.
- MCGREW, L. – MCGREW, T. (2007): *Internalism and Epistemology*. Abingdon: Routledge.
- PUTNAM, H. (1981): Brains in a Vat. In: Putnam, H.: *Reason, Truth, and History*. Cambridge: Cambridge University Press, 1-21.
- STROUD, B. (2000): Understanding Human Knowledge in General. In: Stroud, B.: *Understanding Human Knowledge*. Oxford: Oxford University Press.

Empirická výzva filosofickému pojetí mysli¹

MARTIN VRANÝ

Ústav filosofie a religionistiky. Filozofická fakulta Univerzity Karlovy.
Nám. J. Palacha 2. 116 38 Praha. Česká republika
martin.vrany@gmail.com

ZASLÁN: 07-07-2012 • AKCEPTOVÁN: 16-11-2012

ABSTRACT: The article reflects on influential views of the mind that come from cognitive science and seem to undermine the traditional philosophical view that the mind is simply unified and transparent to itself. Specifically, the modularity thesis is presented, along with its important modifications and criticisms, suggesting that the apparent unity can be ascribed only to higher cognition, if at all. Various theories of why the mind seems to be unified while being composed of autonomous modules are discussed. The overview results in the conclusion that our linguistic capacity plays a prominent role in the unity of the mind.

KEYWORDS: Cognitive architecture – consciousness – language – philosophy of mind – modularity – unity of mind.

Velký rozdíl mezi myslí a tělem tu pozoruji především v tom, že tělo je ze své přirozenosti vždy dělitelné, zatímco mysl je naprosto nedělitelná: neboť když ji, čili sebe, nakolik jsem pouze věc myslící, uvažuji, vůbec v sobě nemohu rozlišit části, nýbrž se chápu výhradně jako jednu a ucelenou věc – a byť je podle všeho celá mysl sjednocená s celým tělem, poznávám, že utnutím nohy či paže či jakékoli jiné tělesné části se nic z mysli neztrácí, a ani o schopnostech chtění, smyslového vnímání, chápání atd. nelze říci, že jsou jejími částmi, protože je to jedna a táž mysl, která chce, smyslově vnímá a chápe. (Descartes 2001, 61)

¹ Tento článek vznikl s podporou Grantové agentury České republiky, č. P401/12/0833.

Descartovy úvahy nastolily tak pevný rámec filosofického pojetí mysli, že i když se dnes pod vlivem vědeckého poznání téměř nikdo nehlásí k jeho verzi dualismu mysli a těla, představa mysli jako něčeho podstatně jednotného a nedělitelného se pořád zdá být tím, co vymezuje mysl tak, jak ji chápou filosofové. V konfrontaci s pohledy vědců, kteří „mysl“ zkoumají empiricky a vždy z hlediska nějakého dílčího aspektu či části, pak filosofové upřesňují, že jejich úvahy se týkají tzv. fenomenálního vědomí nebo též mysli nahlížené z hlediska první osoby (např. Chalmers 1995, Nagel 1974, McGinn 1989, Block 2001). Tím se upevňuje dojem, že kognitivní věda a filosofie mysli se zabývají dvěma různými věcmi, nebo aspoň dvěma podstatně odlišnými aspekty téhož. (Domnívám se přitom, že tento dojem pracuje především na podvědomé úrovni, neboť formulujeme-li jej takto explicitně, málokterý filosof se k němu bude znát, protože zavání nějakou formou dualismu.) Ačkoli je tento dojem do jisté míry opodstatněný, rozhodně to neznamená, že výsledky empirického zkoumání nemohou být relevantní pro filosofické zkoumání mysli.

Tento článek by měl představit některé teorie vzešlé z výzkumů v rámci v kognitivní vědy, které staví onen karteziánský pohled do nové perspektivy, v níž se jeví jako méně samozřejmý, než jak jsme z filosofické tradice zvyklí.

1. Modularita mysli

Empirické vědy zkoumají mysl v naprosté většině případů z hlediska určité její funkce, např. paměti, vizuálního vnímání, emocí, atd. Oproti filosofickým zkoumáním těchto dílčích funkcí se liší zejména tím, že při popisu těchto funkcí nemusí předpokládat reflektující subjekt, jemuž tyto funkce slouží a který je tedy sjednocuje. Řada empirických výzkumů se naopak zaměřuje na rozdíly mezi tím, co se člověk *domnívá*, že ví, vnímá apod., a co skutečně ví, vnímá apod.² V tradičním karteziánském pohledu přitom nedá-

² V rámci filosofie mysli jsou v tomto ohledu již za klasické považovány výzkumy pacientů s přetátným kalózním tělesem, které iniciovali R. Sperry a M. Gazzaniga (Sperry et al. 1969) a které pro filosofii zpopularizoval Nagel (1971). Další populární anomálií je tzv. syndrom opomíjení („neglect syndrom“), kdy si pacient není vědom, že opomíjí určitou část vizuálního pole (více viz např. Hillis 2006). V případě zdravých subjektů existuje řada výzkumů zabývajících se např. falešnými vzpomínkami (např. Roediger – McDermott 1995, Mather et al. 1997), primingem (např. Ferguson et al. 2005), či slepotou vůči změně (např. Simons – Levin 1997). Další výzkumy pak poukazují na případy, kdy si

vá smysl hovořit o nějakém obsahu mysli, který si neuvědomujeme, protože mysl je chápána jako transparentní a vymezená právě tím, co si uvědomujeme (myšlené obsahy). Na základě čeho tedy vědecké výzkumy mohou rozlišovat mezi tím, co je „skutečně“ přítomno v mysli a co je subjektivně zakoušeno?

Je tomu tak na základě pozorovaného chování, které je nejsnáze vysvětlitelné připíšeme-li subjektu neuvědomované obsahy mysli. Ve filosofii mysli jsou v této souvislosti často diskutované např. důsledky oddělení mozkových hemisfér, které vedou ke zřetelné diskrepanci mezi skutečnými a uvědomovanými důvody určitého chování (viz Nagel 1971). Nemusíme však zacházet do tak extrémních případů, abychom si uvědomili, že připisování nevědomých obsahů je běžnou součástí zaujetí tzv. intencionálního postoje. Často si například vysvětlujeme chování jiných lidí tím, že jednají z nevědomých motivů. V tomto kontextu nicméně máme tendenci chápat ono nevědomí v duchu psychoanalýzy jako obranný psychický mechanismus, který se sám řídí důvody, jež lze odhalit a přivést do vědomí. Věda však hovoří o nevědomých obsazích mysli i v silnějším smyslu, kdy tyto obsahy *nemohou* být přivedeny do vědomí. V jádru takového pohledu je teze o modularitě mysli.

Teze o modularitě mysli říká, že mysl se skládá z modulů, které lze v té nejobecnější formulaci chápat jako oddělitelné funkční jednotky. Často zmiňovanou analogií modulárního systému je hi-fi souprava, která se skládá z různých komponent, jež jsou na sobě funkčně nezávislé. Vzájemná nezávislost však neznamená, že moduly si nemohou vyměňovat informace nebo že jejich správná funkčnost nepředpokládá funkčnost jiného modulu. Funkčnost zesilovače předpokládá funkčnost zdroje elektřiny, přesto jde o různé komponenty. Zesilovač bude dále plnit svou funkci, pouze pokud získává nějaké vstupy (nezávisle na tom, zda z CD přehrávače, magnetofonu apod.) a převádí je na výstup vhodný pro další komponenty (reproduktory, sluchátka).

S takto obecnou a nekontroverzní analogií se však při snaze hlouběji porozumět mysli nelze spokojit. Jak totiž ve své kritice modularity mysli správně zdůrazňuje Prinz (2006), při obecné formulaci modulu jako oddělitelné funkční jednotky se teze o modularitě mysli stává v psychologii cel-

člověk není vědom toho, že něco v jistém smyslu ví – viz různé experimenty zkoumající osvojení si implicitních znalostí, o kterých však subjekt implicitně neví (např. Nissen – Bullemer 1987, Reber – Allen 1978). Dobře prozkoumaným jevem v této oblasti je také neuvědomované vnímání (přehled viz Merikle et al. 2001).

kem triviální myšlenkou, že mysl se skládá ze systémů, které se liší vykonávanou funkcí (Prinz 2006, 23). To, co činí modularitu myslí netriviální myšlenkou, jsou právě tvrzení o povaze těchto systémů.

Původní verze teze o modularitě, kterou představil J. Fodor (1983), obsahovala silnější pojetí modulu, než jaké je dnes přijímané. V původním pojetí byl modul nepřístupný, doménově specifický (domain specific) a informačně uzavřený (informationally encapsulated) systém zpracovávající informace, jehož vývoj je geneticky podmíněný.³ Nepřístupnost modulu spočívá v tom, že ostatní moduly nemají přístup k jeho interním procesům a datům – pouze k jeho výstupům. Doménově specifický systém je ten, jehož činnost spouští jen podněty určitého typu (např. obraz na sítnici, hlas apod.) a ty tvoří obor jeho vstupů. Informační uzavřenost znamená, že modul nemá přístup k jiným informacím mimo své vstupy. Z toho vyplývá, že organismus při zpracovávání informací fakticky využívá méně znalostí, než má jako celek k dispozici, protože informace přítomná v jednom modulu nemusí být dostupná pro jiný modul, přestože by pro jeho práci mohla být relevantní.⁴

Jednou z tradičních ilustrací fungování modulární architektury je tzv. Müller-Lyerova iluze. Přestože víme, že obě úsečky jsou stejně dlouhé, iluze různé délky přetrvává (je tzv. kognitivně neprostupná), což v termínech vysvětlených výše znamená, že náš iluzorní vním je výsledkem práce modulu, jehož vstupy jsou pouze vizuální data, je informačně uzavřený vůči našim přesvědčením (tj. v tomto případě zejména vůči přesvědčení, že obě úsečky jsou ve skutečnosti stejně dlouhé) a jeho činnost je nepřístupná, protože introspekci nemůžeme odhalit, proč vnímáme úsečku s šipkami jako kratší.

V tomto silném pojetí se za modulární považovaly pouze nižší kognitivní funkce vnímání, které tvoří rozhraní mezi světem a myšlením. Vyšší kognitivní funkce (volní rozhodování, jednání, uvažování, apod.) však pro své správné fungování vyžadují přístup k informacím z mnoha různých oblastí a při své činnosti se musí moci dotazovat jiných modulů, protože v různých

³ Mezi další vlastnosti, které Fodor modulům připisuje, patří to, že informace zpracovávají relativně rychle a automaticky (nepodléhají vůli), mají jednoduché výstupy a jsou v mozku jasně lokalizované. Tyto vlastnosti lze však chápat buď jako důsledky výše zmíněných hlavních vlastností nebo se jedná o vlastnosti, které nejsou z hlediska výzkumu architektury myslí tak podstatné. Více viz Prinz (2006), Fodor (1983).

⁴ Např. Cheng (1986) dokládá, že krysy se v novém bludišti orientují výhradně podle geometrických vlastností okolí a nejsou schopné využít uměle rozmístěných relevantních pachových značek, ačkoli je v jiném experimentálním kontextu umí rozpoznat.

situacích jsou relevantní různé druhy informací.⁵ Protože je podle Fodora modulární architektura vůdčím schématem, s nímž kognitivní věda přistupuje k výzkumu mysli (více viz Robbins 2010), lze jen obtížně odhalit fungování těch vyšších kognitivních funkcí, které jsou právě většinou středem pozornosti filosofie mysli.

Hlavní důvod, proč zkoumat mysl prismatem modulární architektury, však není metodologický v tom smyslu, že poskytuje jednoduché schéma pro konstrukci teorií různých kognitivních funkcí a jejich experimentálního ověřování. Teze o modularitě je totiž do značné míry podepřena evolučními a „inženýrskými“ argumenty. Modulární architektura má tu výhodu, že díky vzájemné nezávislosti modulů se změna či poškození jednoho modulu projeví jen ve výkonu té specifické funkce a funkčnost ostatních modulů zůstane nezměněna (pokud jejich vstupy netvoří právě výstupy poškozeného modulu). Tato výhoda je z evolučního hlediska důležitá nejen proto, že umožňuje organismu se snáze vyrovnat s lokálním poškozením kognitivního aparátu, ale zejména proto, že usnadňuje vývoj kognitivních schopností. Díky nepřístupnosti a informační uzavřenosti totiž změna způsobu zpracovávání informace nevyžaduje přizpůsobení zbytku kognice, čímž se zvyšuje šance adaptivní změny. Tento princip se uplatňuje i v objektovém programování, které se občas uvádí jako příklad modulární architektury: různí programátoři mohou efektivně vyvíjet komplexní program tak, že každý vyvíjí či upravuje určitou funkci s jasně definovanými vstupy a výstupy nezávisle na ostatních. Kód tak lze upravovat po částech a bez detailní znalosti celkového chodu programu.

Další zásadní výhodou modulární architektury je, že díky vzájemné nezávislosti mohou moduly zpracovávat informace paralelně, což umožňuje rychlejší reakce. Například i v případě rozumění mluvené řeči, což je proces, který se na první pohled zdá vyžadovat postupné zpracovávání (od fonologické úrovně přes syntaktickou až po sémantickou), se ukazuje, že vstupní informace (zvuk mluvené řeči) je zpracovávána paralelně na všech třech úrovních a k nutným korekcím dochází za běhu. Potřeba rychlé reakce plyne z toho, že organismus musí být schopen na své prostředí reagovat v reálném čase, tedy tak, aby reakce mohla být efektivní. To se samozřejmě může v různých druzích situací lišit – reakce na rychle se přibližující předmět musí být rychlejší než reakce na náhlý pokles teploty, tím pádem lze předpo-

⁵ Z toho důvodu jsou tyto vyšší kognitivní funkce občas označovány jako centrální, což však již předjímá jejich nedomulární architekturu.

kládat, že adaptace na první druh situace bude vyžadovat větší míru zapojení kognitivního aparátu. Podobně různé situace mohou vyžadovat různě diferencované reakce, takže zatímco instinktivní uhnutí před blížícím se předmětem může být dílem relativně nekomplikovaného motorického aparátu, rozumění řeči, přestože nemusí být tak rychlé, může být z výpočetního hlediska velice náročné. Podobné evoluční úvahy tedy tvoří určitý rámec pro možné koncepce architektury mysli tím, že ji omezují různými požadavky, které by daná architektura měla plnit.

2. Kritika modularity

V současné kognitivní vědě je poměrně živá otázka, do jaké míry je mysl skutečně modulární, resp. jaká je povaha oněch modulů, z nichž se mysl skládá.

Důležité přehodnocení Fodorova pojetí modulu předkládá Karmiloff-Smithová (1996), která z perspektivy vývojové psychologie argumentuje proti Fodorovu přesvědčení, že vývoj modulů je do značné míry geneticky určený a jejich fungování tedy vrozené. Její tezí je, že mysl se postupně modularizuje na základě interakce s prostředím a počátečních preferencí určitých druhů stimulů. Pro vysvětlení vývoje nějakého modulu tedy nemusíme předpokládat jeho funkční specifikaci na úrovni genomu, nýbrž stačí předpoklad nějaké obecné teorie učení a oboru stimulů, které budou od narození přitahovat relativně více pozornosti jedince. Slovy kognitivní vědy: vrozené dispozice k přijímání vstupů z určitých oborů stačí k tomu, aby se postupem času vyvinuly systémy zpracovávající data právě z těchto domén, které ponese charakteristické znaky modulů. Podle této představy je tedy kognice novorozence relativně nespécifická a díky selektivní pozornosti si dítě rychle vytváří implicitní teorie předmětů preferovaných oborů (řeč, sociální vztahy, neživé předměty, živé organismy, čísla, atd.).⁶

⁶ Je však třeba zdůraznit, že odmítnutím nativistického pojetí modulů se Karmiloff-Smithová ještě nepřiklání k Piagetovu konstruktivismu. Zdůrazňuje, že existuje silná empirická evidence ve prospěch vrozených doménově specifických predispozic, které jsou na sobě do značné míry nezávislé (což se projevuje nejzřejměji selektivními vývojovými poruchami). Důvodem proti přijetí silně nativistického stanoviska je však argument, že v rámci naší teorie mysli je potřeba udělat prostor pro flexibilitu kognitivního vývoje (což je z evolučního hlediska žádoucí), který nativismus umenšuje svým vysvětlujícím schématem s vrozenými principy vývoje pro každou určitou oblast. Více viz Karmiloff-Smith (1996, 4-12).

Samotný proces modularizace, jak jej popisuje Karmiloff-Smithová, je velmi zajímavý tím, že podstatná role je přisuzována aktivní přeměně reprezentací, které se utvořily na základě přímé interakce s prostředím. Má jít o proces, který „přeměňuje *implicitní* informace na *explicitní* znalost“ (Karmiloff-Smith 1996, 17). Můžeme si jej ilustrovat příkladem, kdy se děti učí vyvážit těleso ve vratké pozici na úzkém podstavci. Děti ve věku čtyř a osmi let tento úkol zvládají velice úspěšně, naproti tomu děti kolem šesti let často selhávají, protože těleso umístí na podstavec jeho geometrickým středem, i když je asymetrické a těžiště je tedy vychýlené. Nabízené vysvětlení je, že čtyřleté děti si rychle osvojí *techniku* vyvážení na základě zpětné vazby, kterou těleso proprioceptivně působí. Tato praktická technika posléze podléhá tzv. reprezentačního přepisu (representational redescription), kdy se v mysli re-reprezentuje praktická znalost (vědění *jak*) na jakousi implicitní teorii, která v tomto případě říká, že těleso je vyvážené na svém geometrickém středu. Teprve opakované selhání této teorie spolu s řadou dalších podobně re-reprezentovaných fyzikálních poznatků vede dítě k přechodu na další úroveň nyní již *explicitní* teorie, která je úspěšná.

Hypotézou Karmiloff-Smithové tedy je, že vývoj všech modulárních kognitivních funkcí až do stadia explicitní, artikulovatelné znalosti postupuje pomocí procesu reprezentačního přepisu, kdy se zpracovávají již nikoli vnější vstupy, nýbrž samotné vnitřní reprezentace mysli týkající se dané domény vstupů. Z filosofického hlediska by to znamenalo, že ryleovské „vědění jak“ vývojově předchází „vědění že“ a zároveň (a to je možná důležitější), že ono explicitní „vědění že“ týkající se nás samých není výsledkem nějaké obecné a epistemicky privilegované schopnosti introspektivně nahlédnout principy a důvody, jimiž se řídí naše chování, nýbrž že je zapotřebí proměny reprezentačního formátu pro daný obor. Tyto proměny se podle Karmiloff-Smithové odehrávají pro různé obory nezávisle na sobě a je tudíž možné mít explicitní znalosti v určité oblasti a pouze implicitní v jiné. Zároveň je dobře možné (a často se tak děje), že explicitně utvořená znalost není adekvátní teorií původního „vědění jak“, což jen vrhá další pochybnosti na představu, že mysl je sama sobě do značné míry transparentní. Piaget (1970) např. ukazuje, že děti si mohou osvojit určitou motorickou schopnost (použití odstředivého praku k zásahu cíle) a přitom nebýt schopny podat správné explicitní vysvětlení svého postupu v dané úloze – v tomto případě v jakém momentu je třeba uvolnit projektil.

Modulární pohled tedy na první pohled podporuje názor, že mysl není sama sobě nutně transparentní, protože modulární procesy jsou informačně

uzavřené a tudíž nepřístupné introspekci. Otázkou však zůstává, nakolik je mysl modulární a jaké to má přesně důsledky.

Zatímco Karmiloff-Smithová jen upřesňuje, jaká je povaha modulů, a proti samotnému modulárnímu přístupu k mysli nic nenamítá, J. Prinz (2006) je přesvědčeným odpůrcem modularity. Jeho stanovisko lze shrnout tak, že 1) v silném pojetí modulu není žádná funkce mysli modulární a 2) slabší pojetí je buď triviální, nebo tak vágní, že nemůže označovat žádnou přírodní kategorii. Argumentuje přitom jak empiricky, poukázáním na výsledky, které zpochybňují modularitu nižších kognitivních funkcí (zejména zpracování smyslových informací, které se většinou pokládá za nejlepší příklad modularity), tak i principiálně. Jeden z principiálních argumentů je namířen proti neurčitosti toho, co všechno se může počítat jako specifický obor (domain) vstupů. Prinz namítá, že chceme-li vidět mysl modulárně, stačí volit obory záměrně tak konkrétní, aby v posledku každý jedinečný kognitivní akt měl svůj modul. Nic nám tak podle něj nebrání předpokládat např. modul pro zacházení s filosofickým pojmem *kategorie*. Důsledkem této trivializace je, že nemáme nezávislé kritérium toho, co považovat za obor nějakého kognitivního systému, které by zároveň nebylo vágní.

Umírněnější stanovisko představuje Samuels (2006), který odmítá Prinzovu radikální kritiku, že modularita mysli je buď triviální a tudíž nezajímavá teze nebo prostě empiricky vyvrácená. Konkrétními argumenty ukazujícími Prinzovu neadekvátní interpretaci teze o modularitě se zde nemusíme zabývat, Samuelsův závěr je však takový, že za modulární lze jistě považovat funkce smyslového vnímání a jiné „periferní“ systémy, a naopak, že nemáme důvod se domnívat, že mysl je *masivně* modulární, tj. že mysl je modulární skrz naskrz, včetně vyšších kognitivních funkcí a vědomí.

3. Mysl masivně modulární

Právě teze o masivní modularitě mysli je filosoficky nejzajímavější, protože podle ní je i jednotné vědomí se svou (zdánlivě) centrální exekutivní funkcí a schopností integrovat mentální obsahy různých typů výsledkem činnosti specializovaného, informačně uzavřeného modulu. Radikalnost této teze se ukáže, zvážíme-li původní Fodorův argument proti modularitě vyšších kognitivních funkcí. Ten je založen na tom, že tyto funkce slouží k formování přesvědčení, což je zásadně holistický proces. Jakkoli tematicky vzdálená přesvědčení totiž mohou mít vliv na přijetí jiného přesvědčení,

takže tento centrální proces nemůže být informačně uzavřený ve výše uvedeném smyslu, což je pro Fodora *conditio sine qua non* modulu. Carruthers (2006), který je hlavním zastáncem teze o masivní modularitě, proto ve snaze obhájit možnost modulárního pohledu i na vyšší kognitivní funkce přichází se značně volnějším pojetím modulu. Především odmítá brát moduly jako zcela informačně uzavřené. Navrhuje chápat informační uzavřenost tak, že modul může během zpracovávání informace využívat i informace z jiných modulů, nemůže mít však aktuálně přístup ke všem najednou. To je však z výpočetního hlediska tak samozřejmé, že tento druh uzavřenosti lze stěží považovat za nějakou význačnou vlastnost. V jeho pojetí je pak modul chápán jako

isolable function-specific processing systems, all or almost all of which are domain specific..., whose operations aren't subject to the will, which are associated with specific neural structures (albeit sometimes spatially dispersed ones), and whose internal operations may be inaccessible to the remainder of cognition. (Carruthers 2006, 12)

Jak tedy Carruthers vysvětluje chod vyšších kognitivních funkcí, když je podle něj mysl „masivně modulární“? Nejprve uvádí empirické argumenty vedoucí k závěru, že to, co běžně chápeme jako jednu mentální funkci (např. paměť, učení, motivace, atd.) je ve skutečnosti zajištěno prací různých modulů. Například pohled na jednání, který zde Carruthers představuje, je takový, že pro určitou motivaci existují různá akční schémata (různé moduly) k dosažení daného cíle, která mezi sebou soupeří o možnost být vykonána a to nezávisle na nějakém nadřazeném rozhodovacím mechanismu. Zásadní otázkou tedy je, jak Carruthers vysvětluje, že naše jednání, myšlení atd. se *jeví* jako konsistentní a jednotné, přestože je založeno na paralelní práci různých modulů bez účasti nějaké vyšší organizační jednotky. V jádru jde o stejnou otázku, na kterou se snaží odpovědět Dennett (1991), když hledá způsob, jak se při zkoumání mysli vyhnout metafoře karteziánského divadla a vysvětlit vědomí na základě práce paralelních procesů.

Podobně jako Dennett, i Carruthers dospívá k tomu, že zásadní roli pro vědomí a vyšší kognici vůbec hraje jazyk. Východiskem Carruthersovy teorie je představa globálně vysílaných informací, což je klíčový aspekt vědomí jak jej původně vysvětluje B. Baars ve své vlivné knize *Cognitive Theory of Consciousness* (1988). Podle Baarse je naše vědomí také výsledkem činnosti různých specializovaných center, přičemž jednotný proud vědomí vzniká tím, že jednotlivé systémy spolu soupeří o přístup ke „globálnímu pracovnímu

prostoru“ („global workspace“, jinak též občas ztotožňováno s pracovní pamětí), díky němuž mohou své výstupy prezentovat všem ostatním systémům. Baars funkci této struktury přirovnává k funkci nástěnky, která informuje o tom, co se děje zajímavého a na co by se měla zaměřit aktivita. Je důležité si uvědomit, že podobně jako nástěnka ani globální pracovní prostor nemá exekutivní funkci, tedy sám o sobě nerozhoduje o tom, která globálně vysílaná informace bude mít vliv na jednání. Podle Baarse dochází k jednání, kdykoli se představa určitého cíle stane vědomou a nekonkuruje jí přitom představa nějakého jiného cíle (jako např. při ranním vstávání, kdy vůli vstát konkuruje chuť ještě chvíli spát). Podle této tzv. ideomotorické teorie vůle pak vědomá představa určitého cíle prostě spustí příslušné akční schéma. Zdá se tedy, že boj o přístup ke globální pracovní ploše představuje prakticky všechno to, co se ve filosofii tradičně připisuje aktivnímu subjektu.

Když se Baars snaží blíže popsat mechanismus tohoto boje, dospívá k takovému pohledu, podle něhož globálnost určité informace narůstá podle toho, jak moc nová či neobvyklá v dané situaci je. Globální pracovní prostor je tak brán jako hierarchizovaný a proces globálního vysílání je postupný – určitá zpráva, která je kandidátem na vyvěšení na globální nástěnku postupně získává podporu ostatních center podle toho, zda je podle nich relevantní nebo ne. Při překročení určité hranice je pak tato zpráva globálně vysílána a tudíž zpřístupněna všem ostatním centrům. Klíčovým kritériem, na jehož základě pak určitá informace získává dostatečnou podporu pro to, aby dosáhla vyšší úrovně na žebříčku globálního přístupu, je to, jak moc je informativní, což se chápe jako potenciál redukovat nejistotu ohledně možných stavů. Baarsovými slovy:

Consciousness seems to be a function of information, defined as a reduction of uncertainty within a stable context of receiving processor.
(Baars 1984, 204)

Baarsův model vysvětluje, proč se vědomí jeví jako lineární tok myšlenek nebo percepčí, přestože kognitivní systém pracuje na té základní úrovni paralelně – aby totiž mohl globální pracovní prostor plnit svoji funkci nástěnky, musí v daný okamžik poskytovat pouze jednu informaci. Z výpočetního hlediska je taková architektura sice pomalá, ale zato velice flexibilní v tom smyslu, že umožňuje inovativní řešení nových problémů (zatímco bez možnosti globálního vysílání a následné komunikace specializovaných systémů by jedinou možnou reakcí na novou situaci byla aplikace nějakého zažitého akčního schématu), neboť se na něm mohou podílet všechny specializované moduly.

Baarsův model tedy připisuje vědomí coby globálně vysílaným informacím tu důležitou kognitivní funkci, že umožňuje různým modulům přistupovat k informacím, které mohou být v určitém kontextu důležité, a přístup k nim přitom není běžnou součástí práce daného modulu. Tím dosahuje kognitivní systém jako celek již zmíněné flexibility, což se projevuje jako schopnost adekvátně reagovat na změnu kontextu a vyhnout se následování rigidních akčních schémat. Je však třeba si uvědomit, že Baarsův model je model určité funkční organizace mysli, který neříká nic o tom, jak jsou dané funkce realizovány. V tomto ohledu je Carruthers konkrétnější, když zásadní roli v integraci informací připisuje naší jazykové schopnosti. Jestliže se jazyk postupně vyvinul jako prostředek komunikace, tedy vzájemného informování se, pak se jeho efektivita odvíjí od toho, k jakým informacím má řečový modul přístup a jak rychle je dokáže artikulovat. Oba tyto důvody mají vysvětlovat to, že jsme v jazyce schopni vyjádřit velice úspornou formou (typicky jednou větou) informace pocházející z mnoha různých specializovaných modulů. Rovněž na straně zpracování řeči je důležité, aby význam zpracovaného výroku byl přístupný pro ostatní moduly, protože ty teprve z přijaté informace vyvozují důsledky (ať už praktické, v podobě jednání, nebo teoretické, v podobě vytvoření určitého přesvědčení). I v tomto masivně modulárním pojetí se však ukazuje potřeba tzv. jazyka myšlení (language of thought) coby komunikačního kódu mezi moduly.⁷ Základní myšlenka je tedy taková, že modul produkce řeči získává vstupní informace v nějaké obecné formě (jazyk myšlení) z jiných modulů a na základě syntaktických vlastností jazyka myšlení tyto informace integruje do nějaké jazykové jednotky.⁸

⁷ Je diskutabilní, nakolik je přesné nazývat tento komunikační kód jazykem. Jerry Fodor coby největší zastánce jazyka myšlení (viz např. Fodor 2008) hájí stanovisko, že tomuto komunikačnímu kódu lze připsat sémantiku (čistě referenční) a tudíž jde o smysluplný jazyk. Proti tomu pak stojí názory zdůrazňující normativitu či dvouúrovňovou sémantiku jako zásadní pro cokoli, co chceme nazývat jazykem (viz např. Brandom 2000). Ačkoli je předmětem sporu více než pouze otázka adekvátního použití slova „jazyk“, pro naši potřebu můžeme tento spor opomenout a chápat jazyk myšlení jako přímočarou hypotézu mající vysvětlit možnost globálního přístupu k informacím napříč moduly.

⁸ Nabízí se otázka, co považovat za jazykovou jednotku. Je zřejmé, že zde nejde o otázku, jaký jazykový útvar je ze sémantického hlediska primární (slovo / věta). Vzhledem k tezi, že jazyk umožňuje integrovat více informací do jednoho uchopitelného celku, se jako kritérium jazykové jednotky nabízí to, co subjekt může podržet jako jednu z omezeného počtu položek ve své krátkodobé paměti (7 ± 2 , podle známého Millerova zákona).

Vzhledem k funkci a povaze řeči se tedy jazyková reprezentace ukazuje jako dobrý prostředek umožňující sdílení a integraci informací mezi moduly. Mohlo by se však namítnout, že jestliže při vysvětlení této funkce řeči musíme předpokládat nějaký jazyk myšlení, pak právě tento jazyk myšlení by mohl tuto integrační a distribuční funkci plnit sám o sobě bez ohledu na to, zda jeho obsahy lze artikulovat v přirozeném jazyce. Můžeme však poznamenat, že pro vysvětlení oné funkce jazyka nemusíme předpokládat jazyk myšlení v tom silném smyslu, v jakém jej zavedl a později hájil J. Fodor, podle nějž myšlení spočívá v syntaktických operacích s reprezentacemi v tomto jazyce myšlení, který je atomistický, kompozicionální a jehož sémantika je čistě referenční. Pro plnění integrační a distribuční role totiž stačí předpokládat, že 1) jazykový modul je schopný převést výstupy z ostatních modulů do nějaké jednotné formy (kódu), na základě jejíž struktury může provést syntaktické operace nezbytné k oné integraci informací; a 2) forma výstupu jazykového modulu musí být srozumitelná pro ostatní moduly, aby jazykové reprezentace mohly sloužit jako jejich vstup. Není tedy nutné, aby obě formy byly shodné a tvořily tudíž nějaký univerzální jazyk myšlení s těmi vlastnostmi, které mu přisuzuje Fodor (i když lze argumentovat, že z inženýrského hlediska by to mohlo být výhodné řešení). Především se tím však ukazuje, že ona forma či kód, na jejímž základě lze integrovat informace, je potřebná pro možnost jazykového vyjádření, takže přirozený jazyk představuje důvod pro vývoj takové formy, která posléze umožňuje vyšší kognici. Bylo by asi příliš jednostranné tvrdit, že naše vyšší kognitivní schopnosti se vyvinuly až v důsledku vývoje jazyka, ale rovněž by bylo jednostranné tvrdit, že přirozený jazyk je jen artikulací jazyka myšlení, který se ustavil předtím.

Podobně jako Carruthers (2006) i Dennett (1991) argumentuje, že jazykový modul díky své potřebě přistupovat k informacím z různých modulů umožňuje nakonec jejich vzájemnou komunikaci. Na vysvětlenou nabízí evoluční příběh o tom, jak se z původně čistě externě užívaného jazyka stalo médium našich myšlenkových pochodů. Předpokládejme, že se v nějaké komunitě již vyvinul jazyk do té míry, že jej lze používat pro vzájemné získávání a poskytování informací, takže každý jedinec může hrát roli tazajícího i odpovídajícího. Zaslouchání otázky tedy působí jako stimulus pro hledání odpovědi napříč specializovanými moduly, které nemusí být vzájemně propojené, a zároveň potřeba dodatečné informace v určitém modulu vede k vyjádření odpovídající otázky. Může se tudíž stát, že část kognitivního systému jedince bude obsahovat informaci potřebnou v jiné části a jediný způ-

sob, jak tuto informaci předat je zprostředkovaně skrze jazykové vyjádření. Dennett následně dokresluje příběh tak, že jedinec bezděčně artikuluje otázku, a i když se v okolí nenachází nikdo jiný, vněm vlastní otázky bude adekvátně stimulovat nalezení požadované informace v té příslušné části kognitivního systému. Tento proces „autostimulace“ bude díky zřejmé výhodě posilován a posléze se tedy internalizuje do podoby vnitřního dialogu, který umožňuje sdílení informací napříč moduly a který prožíváme jako proud našeho vědomí.

V trochu jiné souvislosti hovoří Dennett (1991) o vědomí jako o virtuálním, sériovém programu, který běží na paralelním hardwaru mozku, který byl původně uzpůsoben k jiné práci. Všechny tyto pohledy se však snaží ukázat, že ono jednotné karteziánské vědomí vyvstává na základě souhry různých specializovaných systémů. Je dobré si připomenout, že ani Carruthers ani Dennett se ve svém vysvětlení vědomí neodvolávají na určitý jeden takový systém, který by naplňoval onu roli karteziánského subjektu. Dennett takový hypotetický systém nazývá „karteziánské divadlo“ a přesvědčivě argumentuje, že i když se vyhneme představě homunkula, který vyhodnocuje všechny zpracované stimuly a dělá rozhodnutí, a budeme hledat v mozku určitý systém, který plní všechny funkce karteziánského subjektu, žádný takový systém nenajdeme. Pro Carrutherse je zase přesvědčení o neexistenci takového systému jádrem jeho teze, že mysl je *masivně* modulární.

Z hlediska intuice, která stojí v základu karteziánského pohledu na vědomí, se podobné výklady jeví jako neuspokojivé, protože se okamžitě nabízí otázky typu „co tedy rozhoduje / uvažuje / reflektuje, apod.“ Odpověď vyplývající z oněch výkladů, že jde o špatně položenou otázku, neboť rozhodování a uvažování je výsledkem interakce různých modulů, z nichž žádný nemá ústřední roli, pak vede zastánce karteziánského pohledu k takové interpretaci těchto výkladů, že vědomí je jakási iluze či epifenomén vznikající na základě mozkových procesů. Je však zřejmé, že vědomí se bude jako iluzorní či epifenomenální ukazovat pouze tehdy, jsme-li a priori přesvědčeni o tom, že subjekt musí mít ztělesnění odpovídající svou strukturou karteziánskému pohledu, tedy že jeho mohutnosti budou soustředně orientovány. Pokud toto přesvědčení uzavorkujeme a zaměříme se místo toho na otázku, zda je možné, aby vědomí plnilo své funkce, aniž by bylo sjednocené v onom reálném subjektu, ukážou se výklady, jaké nabízí např. Baars (1988), Carruthers (2006) či Dennett (1991), jako mnohem smysluplnější.

Dalo by se namítnout, že zaujetí tohoto funkcionalistického pohledu na mysl už samo o sobě rozhoduje řadu otázek (zejména tím, že se v jeho světle

ukazují jako zbytečné) a tudíž je třeba pro něj nalézt nezávislé důvody.⁹ Je pravda, že především otázky týkající se tzv. subjektivně prožívaných (fenomenálních) kvalit mentálních stavů se tím vyloučí z diskuse, avšak stále zůstane ve hře řada otázek, které s karteziánským pohledem souvisí a na které nejsou jednoznačné odpovědi. Tou nejdůležitější je asi právě otázka, zda je mysl skutečně *masivně* modulární. Jak bylo řečeno výše, kognitivní vědci se totiž rozcházejí v otázce, do jaké míry lze mysl chápat modulárně. Carruthersův *masivně* modulární pohled je extrémní v tom, že řadu procesů, které doposud kognitivní věda chápala jako neredukovatelně globální (např. praktické uvažování), se snaží vysvětlit na základě interakce modulů. Ačkoli je posouzení argumentů obou stran mimo kompetence autora tohoto článku, ukazuje se tím přinejmenším to, že otázku, do jaké míry je určitá schopnost výsledkem práce systému s globálním přístupem (a tudíž metaforicky jako práce aktivního subjektu), lze posuzovat s ohledem na empirické poznatky.

Dodejme, že součástí oněch neintuitivních teorií mysli je i vysvětlení, proč se nám introspektivně mysl jeví právě tím způsobem, který lze po jeho teoretické reflexi označit za karteziánský. Poměrně detailní vysvětlení nabízí právě Carruthersova teorie, podle níž vlastní mysl vnímáme prismatem tzv. lidové teorie mysli, která se původně vyvinula za účelem předvídání chování druhých lidí. V duchu modulárního přístupu tak mluví o modulu „čtení mysli“, jehož funkcí je prisuzovat různým jedincům propoziční postoje (přání, přesvědčení, záměry, atd.) a na jejich základě odhadovat budoucí chování.¹⁰ V důsledku pak vědomí vlastních myšlenek není výraz toho, že by mysl byla sama sobě transparentní, nýbrž je výrazem nevědomé sebeinterpretace, čemuž nasvědčuje řada psychologických jevů, zejména tzv. konfabulace. Ačkoli se díky Freudovi tato myšlenka již nezdá být tak kontroverzní, je stále kontraintuitivní v tom, že upírá jedinci absolutní autoritu nad tím, jaké propoziční postoje zastává. Je-li totiž připsání přesvědčení výsledkem interpretace (či zaujetí intencionálního postoje, jak by řekl Dennett) a nikoli

⁹ Viz Chalmers (1995) a jeho tvrzení, že po zaujetí funkcionalistického pohledu se všechny problémy stávají „jednoduchými“, ale ten „těžký problém“, jenž spočívá ve vysvětlení kvality naší zkušenosti, zůstane stranou.

¹⁰ Více o modulu pro teorii mysli viz Baron-Cohen (1997). Podobné vysvětlení nabízí i Sellars (1956), když na smyšleném příběhu ukazuje, jak by se vědomí vlastních mentálních stavů mohlo vyvinout díky předcházející jazykové hře na připsování mentálních stavů ostatním jedincům. Zatímco však Sellarsův spekulativní příběh klade důraz na roli jazyka při tomto procesu sebeuvědomění, Carruthers na základě poznatků z vývojové psychologie kloní k názoru, že schopnost „čtení mysli“ je do značné míry vrozená.

přímého nazírání, pak sebeinterpretace a interpretace druhého jsou epistemicky na stejné úrovni, s tím rozdílem, že proces sebeinterpretace může mít lepší přístup k relevantním datům.

Takto zjednodušeně reprodukované vysvětlení by však mohlo vést k mylné představě, že se tím tvrdí, že člověk pozoruje své vlastní chování a na základě toho si domýšlí, jaké propoziční postoje vlastně zastává. Jelikož mezi vstupy modulu pro čtení mysli mohou patřit i vlastní perceptuální stavy stejně jako obsah vnitřního dialogu (řeč a vědomí toho, co ostatní vnímají, jsou důležité determinanty připsovaných propozičních postojů), není člověk při sebeinterpretaci odkázán pouze na pozorování svého vnějšího chování. Uvědomění si svých perceptuálních stavů a vnitřního dialogu pak podle Carrutherse pouze podléhá konceptualizaci, která je vlastní modulu pro čtení mysli a je tedy v jistém smyslu bezprostřednější než připsování si propozičních postojů, neboť vstupem pro onen modul jsou tyto stavy samotné.

V jistém smyslu je tedy mysl sama sobě v určité oblasti transparentní tím méně, čím více tato oblast podléhá nějaké teorii, jelikož člověk si je vědom pouze výsledku práce modulu ztělesňujícího danou teorii, nikoli však samotných principů, jimiž se práce toho modulu řídí. Zatímco např. vjem sklonu linie ve vizuálním poli je výsledkem relativně přímočarého zpracování informací v primární zrakové kůře (V1), reprezentace okolního 3D prostoru na základě obrazu na sítnici již podléhá řadě předpokladů, jejichž porušení může vést k iluzím, které mohou být kognitivně neprostopupné (viz výše). Nakolik je potom uvědomění si obsahů vlastní mysli také výsledkem práce specializovaného modulu, který zase pracuje na základě introspektivně nepřístupných principů, natolik je pravděpodobné, že při sebepozorování podléháme iluzím stejně jako v případě pozorování vnějšího světa.¹¹

¹¹ Jack a Shallice (2001) v této souvislosti tvrdí, že právě z tohoto důvodu jsou introspektivní výpovědi subjektů pro kognitivní psychologii obtížně použitelné, zvláště pokud je introspektivní kategorizace vlastních mentálních stavů silně idiosynkratická (což autoři naznačují). Zároveň však zdůrazňují, že empirický výzkum vědomí se bez introspektivních výpovědí neobejde a proto navrhuje metodologický rámec, v němž by bylo možné introspektivní výpovědi validovat na základě objektivně měřitelných výsledků (např. reakční čas).

4. Od klasické umělé inteligence ke vtělené kognici

Jedním z důvodů, proč se představa karteziánského divadla vkrádá do myšlení o mysli, je i to, že dlouhou dobu byla pro empirické zkoumání mysli paradigmatická představa mysli jako počítačového programu, podle níž myšlení je výpočetní povahy a je nezávislé na hardwaru, na kterém běží. Vysvětlení mysli či nějaké dílčí funkce myšlení pak spočívá v odhalení algoritmu, který z daných vstupů odvodí stejné nebo lepší výstupy. Tato tzv. komputační teorie mysli se rozvíjela především díky klasické umělé inteligenci, která považuje úroveň hardwaru a softwaru za jasně oddělenou a program chápe jako explicitně formulovaný algoritmus sestávající z postupně prováděných kroků, což odpovídá architektuře běžných počítačů. Jelikož všechny instrukce programu jsou postupně zpracovávány procesorem, je procesor (či nějaká nadřazená jednotka, která rozděljuje práci na instrukcích dalším procesorům) tím místem, kde se scházejí a vyhodnocují informace z celého systému a v analogii s myslí tedy jevištěm karteziánského divadla. Klasická umělá inteligence přitom usiluje zejména o řešení úloh, které vyžadují globální přístup k informacím, k čemuž právě pomáhá sériová architektura tím, že centrální rozhodovací jednotka postupně získává a vyhodnocuje informace z různých zdrojů. I když však připustíme, že globální úlohy lze optimálně řešit pomocí sériově zpracovávaných explicitních algoritmů, pro pochopení naší mysli je potřeba zohlednit omezení, která na zpracování informací klade prostředí. Důležitým omezením je, že kognitivní systém musí informace vyhodnotit v „reálném“ čase, tj. tak, aby ještě reakce mohla mít adekvátní následek, což může často být rychleji, než jak by toho bylo možné dosáhnout postupným zpracováním dat v jedné centrální jednotce (vzhledem k omezené rychlosti přenosu informací pomocí nervových vláken). Dalším omezením je, že náš mozek je evolucí uzpůsoben k paralelnímu zpracování informací v různých a na sobě často nezávislých oblastech, takže i když se naše kognitivní procesy často dělí do dvou systémů, z nichž jeden se vyznačuje nevědomým, paralelním a rychlým zpracováním informací, zatímco druhý systém pracuje vědomě, lineárně a pomalu, je i ten druhý systém realizovaný na paralelním hardwaru.

Postupem času se ukázalo, že klasická umělá inteligence může jen obtížně řešit určité typy úloh spočívající v tzv. rozeznávání vzorců, např. vizuální rozpoznávání objektů, nebo komplexní reakci na prostředí, např. chůze v nerovném terénu. Všechny tyto úlohy přitom spadají pod nižší kognitivní schopnosti, které ovládají i relativně jednoduché organismy. Některé z těch-

to úloh se ukázaly jako snáze řešitelné za použití tzv. konekcionistických sítí, které informace zpracovávají paralelně a místo explicitního algoritmu představují výsledek postupného učení. Tím vznikla na poli výzkumu umělé inteligence a odvozeně i myslí často dosti vyostřená debata mezi zastánci klasické umělé inteligence a právě konekcionismu. Hlavním argumentem konekcionistů bylo, že jejich výpočetní architektura více odpovídá způsobu, jakým informace zpracovává mozek, což se projevuje v tom, že řadu úloh právě konekcionistické sítě úspěšně řeší po krátkém období učení. Nevýhodou konekcionistického přístupu ovšem je, že způsob, jakým sítě řeší nějakou kognitivní úlohu, prakticky nelze interpretovat, protože způsob řešení se neskládá z různých, navzájem jasně rozlišitelných kroků (jako v případě explicitního algoritmu), nýbrž je holisticky ztvárněn na úrovni celé sítě. Jednotlivému uzlu v síti ani vahám na spojích mezi nimi nelze připsat nějakou výhradní roli. Kdybychom se tedy chtěli dobrat nějakých principů, podle nichž se v určité úloze řídí naše myšlení, konekcionismus nám příliš nepomůže, protože ze sítě úspěšně řešící danou úlohu nebudeme moci tyto principy odečíst. Rovněž existuje zase řada úloh, které je snazší řešit explicitním algoritmem než konekcionisticky – jedná se zejména o úlohy pracující s přesně vymezenými logickými vztahy, jako např. počítání, hraní šachu či samo logické odvozování. I když se tak nakonec zdá, že každý z přístupů je vhodný pro různé typy úloh, je konekcionistická alternativa přínosná v tom, že poskytuje novou perspektivu pro počítačnou teorii mysli.¹² Důležitým důsledkem této nové perspektivy je směr, který se nazývá „vtělená kognice“.

Ústřední motivací vtělené kognice, jejímž hlavním zastáncem je Clark (1996, 2001), je myšlenka, že schopnosti organismu se vyvinuly za účelem relativně specifických interakcí s prostředím, nikoli proto, aby poskytovaly obecná řešení abstraktně formulovaných úloh. Při řešení takto specifických úloh tedy příroda může využívat všechny triky, které má k dispozici, k nimž patří zejména to, že při řešení využívá vlastností prostředí a organismu samotného. Příkladem může být, že při řízení chůze může kognitivní systém prostě využít mechanických vlastností nášlapné plochy (např. tvarová přizpůsobivost či přilnavost) místo toho, aby komplikovaně vypočítával správné nastavení. Jiným příkladem, který dobře ilustruje rozdíl oproti klasické

¹² Více o vztahu klasické umělé inteligence a konekcionismu viz Garson (2010) či Clark (2001, kap. 4). Naopak dobré shrnutí argumentů ve prospěch klasické umělé inteligence poskytuje Carruthers (2006, 45-50).

umělé inteligenci, je propojení motoriky s vizuálním vnímáním. Zatímco klasická umělá inteligence by nejprve zpracovala vizuální vstupy, na jejich základě pak vytvořila 3D model okolního prostředí a na jeho základě zase řídila pohyb tímto prostředím, vtělená kognice používá „svět jako svůj nejlepší model“ a vizuální vnímání se už předem odvíjí od úmyslu tímto prostředím projít a může se tím pádem obejít bez prostředníka v podobě vnitřního 3D modelu okolí. Systém se tedy jinak dívá na prostředí tehdy, když jím chce projít a když v něm např. něco hledá.

Z hlediska vtělené kognice tedy porozumět kognitivním schopnostem nějakého organismu nelze bez přihlídnutí k jeho životní formě, konkrétně pak k jeho cílům a prostředí. Důraz na vztah organismu a prostředí nejen stírá dříve jasné rozlišení mezi vnímáním a jednáním (vnímání se ukazuje jako aktivní ohledávání světa za účelem možného jednání), ale zpochybňuje i představu, že myšlení je něco, co se děje výhradně v hlavě. Schopnost přizpůsobit si prostředí tak, aby se snížily kognitivní nároky na určité jednání, vede k tomu, že takto uzpůsobené prostředí je součástí řešení nějakého úkolu stejnou měrou, jako myšlenky, které aktuálně probíhají hlavou. Např. pomocné výpočty na papíře nebo samotný napsaný text představují proud myšlenek, který často umožňuje dospět k jinak nedosažitelným řešením. Zde právě vtělená kognice přichází s vysvětlením, jak může náš mozek být mimo jiné nástrojem postupného a abstraktního uvažování, když se původně vyvinul spojováním různých, paralelně pracujících oblastí určených k účelnému jednání ve světě. Díky využití prostředí jako tzv. vnějšího lešení (external scaffolding), kdy se dílčí kroky a mezivýsledky komplikovaného myšlenkového postupu mohou prostě ztvárnit a lze se tak k nim znovu vrátit, lze překonat jinak omezenou kapacitu pracovní paměti. Komplikované myšlenkové pochody se tak mohou rozdělit na řadu kroků, které lze postupně provádět. Samotná možnost využívat prostředí k rozšíření vlastní paměti a snižování výpočetní náročnosti by však nemusela stačit tam, kde je k nějakému syntetickému myšlenkovému kroku potřeba nahlédnout a podržet v mysli všechny předcházející mezivýsledky. Proto je důležité doplnit ideu využívání prostředí coby vnějšího lešení našeho myšlení o evoluční perspektivu. Počáteční úspěch této strategie totiž mohl vést k většímu přizpůsobení naší kognice právě k tomuto účelu.

Ústřední role v tomto postupném vývoji směrem k abstraktnímu uvažování se opět připisuje jazyku. Jazyk umožňuje snížit kognitivní nároky na práci s pojmy či reprezentacemi tím, že v jednoduše přístupné formě artikulovaného slova fixuje různá rozlišení, která mozek činí. Vezmeme-li pojem

v širším smyslu jako soubor rozlišení, ke kterým může mozek dospět na základě smyslových vjemů (např. malý žlutý objekt se specifickou vůní – citron), pak stabilní vztah mezi slovem a touto sadou rozlišení umožňuje použít slova k rychlé aktivaci těchto rozlišení bez nutnosti konstruovat celou sadu znovu na základě kontextu. Tato výhoda se zřetelně ukazuje např. v zavádění nových symbolů v matematice, které mají označovat struktury, jež lze pořád díky hierarchii definic vyjádřit v původním, holém axiomatickém jazyce. Hierarchie definic v rámci symbolického jazyka matematiky ilustruje i ten rys sepětí jazyka a myšlení, že stejně jako se komplikované struktury definují pomocí již definovaných symbolů, je i řada našich pojmů vázána na jiné pojmy, nikoli nutně na rozlišení pocházející ze smyslového vnímání. Jelikož však jazyk nemusí sloužit pouze k zastupování smyslových vjemů a zároveň se v mozku vztahy mezi pojmy mohou vytvářet i na základě asociace a nikoli výlučně logických či sémantických vztahů, je třeba výše zmíněnou analogii chápat s tím omezením, že přirozený jazyk, na rozdíl od umělého jazyka matematiky, nemusí být atomistický, tj. není nutné, aby každý pojem, s nímž naše kognice pracuje, bylo možné v posledku rozložit na sadu rozlišení pocházející ze smyslového vnímání.

Odtud pak pochází myšlenka, že jazyk je nástrojem kulturní evoluce, protože umožňuje v podobě jazykových her uchovávat nové pojmy a myšlenky pro další generace. Jazyk je tedy výsledkem práce mnoha generací, a jelikož s rozšiřujícím se pojmovým schématem se rozšiřují i naše kognitivní schopnosti, samotné osvojení si jazyka tak znamená postavit se na ramena velikánů a mít možnost dohlédnout dál než předchozí generace. Možnost radikálního rozšíření našich kognitivních schopností používáním jazyka motivuje diskuse ohledně toho, zda propast mezi lidskou a zvířecí inteligencí je důsledkem jedné malé evoluční adaptace, která pak odstartovala kulturní evoluci, nebo zda bylo zapotřebí komplikované souhry různých rysů. První pohled, který zastává např. Dennett nebo Clark, se opírá mimo jiné i o fakt, že z genetického hlediska se lidé liší od některých opic jen minimálně. Zdroj naší inteligence tedy pravděpodobně najdeme spíše ve vyvíjející se kultuře a naší schopnosti si ji rychle osvojit, k čemuž je jazyk zásadní.

5. Shrnutí

Empirický výzkum myslí vychází od dílčích schopností, které analyzuje nezávisle na sobě. Základním předpokladem tak je, že mysl, nakolik je pří-

rodním fenoménem vyskytující se v různé míře nejen u lidí, není jednotná a sestává z různě propojených schopností, které se původně vyvinuly ke specifickým účelům, přičemž žádné z nich nelze připsat všechny funkce, která se ve filosofii tradičně připisují myslícímu, reflektujícímu a rozhodujícímu se subjektu. Jednota mysli, kterou Descartes vztahuje na mysl prožívanou z perspektivy první osoby, se v interpretaci kognitivní vědy ukazuje buď jako strukturální rys vyšší kognice (přístup ke globálnímu pracovnímu prostoru) nebo jako výsledek sebeinterpretace, která vytváří dojem jednoty díky vrozené teorii mysli, jež se původně vyvinula k předvídání chování druhých lidí. Oba pohledy, které se vzájemně spíše doplňují, než vylučují, dále připisují podstatnou roli jazyku. Jazyková schopnost je díky svému účelu artikulace mentálních obsahů tím, co může vytvářet globální pracovní prostor a být tak základem našich vyšších kognitivních schopností. Trochu jiný pohled zase zdůrazňuje jazykem založenou možnost tvořit mentální stavy vyššího řádu, které tvoří proud našeho vědomí.

Předestřené pokusy o vysvětlení vědomí založené na empirických zkoumáních tedy nelze jednoduše odmítnout jako redukcionistická pouze proto, že zpochybňují transparentnost a jednotu mysli, což jsou definiční vlastnosti karteziánského pojetí vědomí nahlíženého z perspektivy první osoby. Ukazuje se totiž, že díváme-li se na mysl jako na vyvinuvší se přírodní fenomén, není charakterizace pomocí těchto vlastností adekvátní. Zároveň však i v tomto pohledu lze najít důvody k vysvětlení, proč se mysl introspektivně jeví jednotná a transparentní. A zaujmout stanovisko, že karteziánský (či spekulativně filosofický) a kognitivně vědecký pohled na mysl se vlastně zabývají různými věcmi označenými stejným slovem, znamená odmítnout možnost vzájemného poučení, což by byl spíše projev dogmatismu než kritického myšlení.

Literatura

- BAARS, B. J. (1988): *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- BARON-COHEN, S. (1997): *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.
- BLOCK, N. (2001): Paradox and cross purposes in recent work on consciousness. *Cognition* 79, No. 1, 197-219.
- BRANDOM, R. B. (2000): *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

- CARRUTHERS, P. (2006): *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Clarendon Press.
- CHALMERS, D. J. (1995): Facing up to the problem of consciousness. *Journal of Consciousness Studies*, No. 2, 200-219.
- CHENG, K. (1986): A purely geometric module in the rat's spatial representation. *Cognition* 23, 149-178.
- CLARK, A. (1996): *Being There*. MIT Press.
- CLARK, A. (2001): *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford – New York: Oxford University Press.
- DENNETT, D. C. (1991): *Consciousness Explained*. Penguin.
- DESCARTES, R. (2001): *Meditationes de prima philosophia: Meditace o první filosofii*. Praha: Oikúmené.
- FERGUSON, M. J. – BARGH, J. A. – NAYAK, D. A. (2005): After-affects: How automatic evaluations influence the interpretation of subsequent, unrelated stimuli. *Journal of Experimental Social Psychology* 41, 182-191.
- FODOR, J. (1983): *Modularity of Mind*. MIT Press.
- FODOR, J. (2008): *LOT 2: The Language of Thought Revisited*. Oxford: Clarendon Press.
- GARSON, J. (2010): Connectionism. In: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*. URL = <http://plato.stanford.edu/archives/win2010/entries/connectionism/>.
- HILLIS, A. E. (2006): Neurobiology of unilateral spatial neglect. *Neuroscientists* 12, 153-163.
- JACK, I. – SHALLICE, T. (2001): Introspective physicalism as an approach to the science of consciousness. *Cognition* 79, Nos. 1-2, 161-196.
- MCGINN, C. (1989): Can We Solve the Mind-Body Problem? *Mind* 98, 349-366.
- NAGEL, T. (1971): Brain bisection and the unity of consciousness. *Synthese* 22, 396-413.
- NAGEL, T. (1974): What is it like to be a bat? *Philosophical Review* 83, 435-450.
- KARMILOFF-SMITH, A. (1992): *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, Mass: MIT Press.
- MATHER, M. – HENKEL, L. A. – JOHNSON, M. J. (1997): Evaluating the characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition* 25, 826-837.
- MERIKLE, P. M. – SMILEK, D. (2001): Perception without awareness: Perspectives from cognitive psychology. *Cognition* 79, No. 1, 115-134.
- NISSEN, M. J. – BULLEMER, P. (1987): Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology* 19, No. 1, 1-32.
- PIAGET, J. (1970): *Psychologie et épistémologie*. Paris: Gonthier.
- PRINZ, J. J. (2006): Is the mind really modular? In: Stainton, R. J. (ed.): *Contemporary Debates in Cognitive Science*. Blackwell.
- REBER, A. S. – ALLEN, R. (1978): Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition* 6, 189-221.
- ROBBINS, P. (2010): Modularity of Mind. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*. URL = <http://plato.stanford.edu/archives/sum2010/entries/modularity-mind/>.

- ROEDIGER, H. L. – MCDERMOTT, K. B. (1995): Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21, 803-814.
- SAMUELS, R. (2006): Is the human mind massively modular? In: Stainton, R. J. (ed.): *Contemporary Debates in Cognitive Science*. Blackwell.
- SELLARS, W. S. (1956): Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science* 1, 253-329.
- SIMONS, D. J. – LEVIN, D. T. (1997): Change blindness. *Trends in Cognitive Sciences* 1, No. 7, 261-267.
- SPERRY, R. W. – GAZZANIGA, M. S. – BOGEN, J. E. (1969): Interhemispheric relationships: The neocortical commissures; syndromes of hemisphere disconnection. In: Vinken, P. J. – Bruyn, G. W. (eds.): *Handbook of clinical neurology*, Vol. 4. Amsterdam: North Holland, 273-290.

Fikce, skutečnost a radikální vyprávění

PETR KOŽÁTKO

Oddělení analytické filosofie. Filosofický ústav Akademie věd České republiky, v.v.i.
Jilská 1. 110 00 Praha 1. Česká republika
kotatko@flu.cas.cz

ZASLÁN: 21-09-2012 • AKCEPTOVÁN: 09-12-2012

ABSTRACT: The article focuses on the nature of the worlds of narrative fiction, ways of their representation, the status and identity conditions of fictional entities and correlatively on the role of singular terms in literary texts. According to the author, the basic question providing a proper framework for addressing such topics is: what does the reader have to do (to presuppose, to accept, to imagine) in order to allow the text of narrative fiction to fulfill its literary functions? The alternative is to start with the “text itself”, i.e. sentences with their linguistic meanings (in abstraction from their literary functions), and ask what kind of material does the text provide to the interpreter, what does it enable him/her to identify and determine and what does it leave principally unidentifiable and underdetermined. According to the author, such an approach blocks the access (or makes impossible the return) to the text’s literary functions. The author defends certain specification of the interpretative attitude required by the literary functions of a text of narrative fiction from its reader. Among other things, he attempts to demonstrate its general applicability by analyzing a highly non-standard type of narration (labelled “radical”).

KEYWORDS: Fictional entities – fictional worlds – literary functions – types of narration.

0. Úvod

Následující úvahy budou situovány na pomezí filosofie jazyka a literární teorie. Obě disciplíny mají, jak známo, řadu zajímavých průniků, ale nejenom to: domnívám se, že základní vztahy či situace, které by měly popsat

a vyložit, jsou v podstatných ohledech analogické. Ve filosofii jazyka jde o situace, v nichž promluvy nebo jazykové výrazy jako typy plní své *komunikativní funkce* a v rámci toho nabývají *význam*. V literární teorii jde o situace, v nichž texty plní své *literární funkce* a v rámci toho nesou *literární kvality*. Analogická nebo dokonce shodná jsou i pokušení a z nich plynoucí zkreslení, kterým musíme v obou disciplínách čelit. Jedním z nich je nutková potřeba zastavět scénu, tj. situace, o nichž je tu řeč, teoretickými konstrukty, které jsou vyžadovány našimi oblíbenými doktrínami nebo generovány našim technickým aparátem. Snadno pak vzniká dojem, že výkony účastníků těchto situací – mluvčích a posluchačů, autorů literárních textů a jejich interpretů – musí být obestavěny rozvětveným konceptuálním, případně normativním lešením, aby byly schopny generovat významy nebo literární kvality.

Odolat této představě by nemělo být těžké – vezmeme-li si poučení z (poměrně) nedávného vývoje analytické filosofie jazyka. Pokud jde o *významy promluv*, Donald Davidson (srov. zvl. Davidson 1986) ukázal, že je lze vyložit velmi přímočaře z prostého zájmu účastníků komunikace o *dorozumění* – ze snahy sladit komunikativní záměr mluvčího a adresátovu interpretaci. Tato shoda, k níž směřuje koordinované jednání mluvčího a adresáta (volba výrazů na jedné straně a jejich interpretace na druhé) konstituuje významy promluv nezávisle na tom, zda jí bylo dosaženo standardním nebo zcela nestandardním, nekonvenčním užitím výrazů a jejich interpretací. Díky Davidu Lewisovi zase známe způsob, jak vyložit samotný *konvenční význam* jazykových výrazů ze stejné reálné báze: z praktického zájmu členů společenství o koordinaci jejich komunikativního jednání. Tento zájem a jistý soubor dalších sdílených postojů (vymezených v Lewisově definici konvence) stačí k fixaci přiřazení mezi výrazy a významy bez intervence jakýchkoli norem.¹ To odpovídá intuici, že k ustáleným způsobům užívání jazykových výrazů se nepřipojujeme proto, abychom vyhověli normativní síle pravidel, ale proto, abychom se *domluvili*.

Paralela, kvůli níž jsem začal na půdě filosofie jazyka, zní takto: podobně jako v teorii významu musíme všechno podřídít otázce, *jaké podmínky musí být splněny, aby promluvy plnily základní komunikativní funkci* – tj. *aby účastníci komunikace dosáhli dorozumění* (protože žádný podstatnější zřetel tu není), ve sféře teorie literární interpretace musí výchozí otázka znít: *jaké pod-*

¹ To je důvod, proč Lewis odmítá označovat konvence jako *pravidla*. Srov. Lewis (1983, 179).

mínky musí být splněny (a především: co musí udělat čtenář – co musí přijmout, co si musí představovat), aby text mohl plnit své literární funkce?

1. Problematické nasazení: potenciál textu

Naopak je chyba začít „textem samým“, tj. větami s jejich jazykovými významy (v izolaci od literárních funkcí textu), a ptát se, jaký materiál nám text poskytuje, co nám umožňuje identifikovat či určit a co zůstává principiálně nedourčené či neúplné. Obávám se, že taková volba východiska nám zablokuje přístup k literárním funkcím textu: stěží můžeme doufat, že nás k nim v nějaké fázi dovede výklad, který od nich na začátku abstrahoval. Stejně beznadějně je sledovat obě hlediska (potenciál textu „an sich“ a nároky, které na nás kladou literární funkce textu) paralelně. V takovém případě snadno dospějeme k neslučitelným závěrům: ke zjištění, že text po nás vyžaduje něco, co nám principiálně nedovoluje udělat. Takový konflikt signalizuje špatné kladení otázek: neměli bychom ho přenášet na „věc samu“ a vydávat ho za rozpor, jímž je ze své podstaty zatíženo samotné literární dílo nebo jeho interpretace.

2. Správné odpovědi na špatně zvolené otázky

Vyzkoušejme si tedy způsob kladení otázek, jemuž jsem předem vyjádřil nedůvěru: začneme „textem samým“, dejme tomu textem Flaubertovy *Madame Bovary* a budeme se ptát, jaké nástroje identifikace nám jako čtenářům poskytuje a co z toho plyne pro jeho klíčové sémantické funkce.²

a) *Otázka:* Umožňuje nám text *identifikovat svět díla* – svět, v němž se odehrává Emin příběh?

Odpověď: Ne. Poskytuje nám deskriptivní určení stavů věcí (situací a událostí), která jsou splňována třídou možných světů. Všechny tyto světy (v nichž se stalo to, co se líčí ve Flaubertově textu) mají stejný nárok na označení „světy románu *Madame Bovary*“: neexistuje kritérium, které by umožnilo vyloučit všechny až na jeden.

² Odpovědi na následující čtveřici otázek odpovídají stanovisku, které najdeme v Currie (2003).

- b) *Otázka*: Umožňuje nám text *identifikovat osoby*, o nichž se vypráví – například osobu označenou výrazem „Ema B.“?

Odpověď: Ne. Jako v předchozím případě dostáváme jen komplex určení (typu: seznámila se s Karlem B., vzala si ho, žila s ním v Yonville atd.), který je v různých světech splňován *různými individui*. I když mezi těmito určeními najdeme identifikační deskripce (typu „manželka jediného lékaře, který vedl praxi v Yonville v letech...“), které ve světě, na nějž se aplikují, vyčlení přesně jedno individuum nebo nic, nefixují totéž individuum *napříč světy*. A protože nemáme způsob, jak vybrat jeden svět, k němuž bychom měli taková určení výlučně vztáhnout (viz a), nemůžeme identifikovat jedno individuum jako Emu B. Materiál, který nacházíme v textu, vymezuje jen *rolí Emy*, kterou v různých světech ztělesňují různá individua (ta, která v nich splňují všechna určení uvedená ve Flaubertově textu). Z logického hlediska se role chová jako (parciální) *funkce* přiřazující individua možným světům.

- c) *Otázka*: Můžeme výraz „Ema B.“, jak se vyskytuje ve Flaubertově textu, interpretovat jako *vlastní jméno* individua?

Odpověď: Ne. Neexistuje kritérium, jež by nám umožnilo rozhodnout, které z individuí splňujících roli Emy v různých světech má být uznáno jako referent výrazu „Ema B.“ Tento výraz (jak je užíván v textu románu) tedy neoznačuje individuum, ale *rolí*, chápanou jako funkce z možných světů do individuí (viz b).

- d) *Otázka*: Můžeme užití věty „Emu omrzelo venkovský život“ ve Flaubertově textu interpretovat tak, že vyjadřuje *singulární propozici*?

Odpověď: Ne. Věta neobsahuje žádný prostředek singulární reference (viz c).

3. Konflikt

Problém je v tom, že text po nás současně vyžaduje, abychom uvažovali *o osobách* – máme-li si z něj poskládat příběh, který by dával smysl. Bylo by zvláštní, kdyby nás namísto toho nutil vzít na vědomí, že *role Emy* se seznámila s *rolí Karla* a provdala se za ni, podvedla *rolí Karla* s *rolí Rudolfa* atd. Takové věci *role* (tj. funkce) nedělají: jsou někým splňovány, jsou navzájem kompatibilní či nekompatibilní atd., ale neuzavírají sňatky, nenasazují

zují si parohy atd. Nemá smysl namítat, že si (jako čtenáři) máme představovat, že tyto věci dělá individuuum splňující *rolí Emy*: takových individuí je, jak víme, nepřehledný zástup. Nemáme-li si představovat zástup žen paralelně vykonávajících tyto a další věci v různých světech, ale jen jednu z těchto žen, vzniká otázka, kterou – a ta je principiálně nerozhodnutelná (viz 2b).

Pokus zkonstruovat z nich jednu osobu (aby si mohla vzít Karla, podvést ho s Rudolfem atd.) by vedl k bizarním výsledkům. Buď ji poskládáme ze všech vlastností, které mají ženy splňující *rolí Emy* v jednotlivých světech (vzhledem k těmto třídám vlastností tedy půjde o sjednocení). Pak dostaneme inkonzistentní bytost: osobu, která má všechny vlastnosti, o nichž se mluví ve Flaubertově textu, a ve všech ostatních parametrech (obligatorních způsobech určenosti náležejících lidským bytostem) má soubory neslučitelných vlastností. Její oči jsou současně modré, hnědé, černé atd., počet jejích vlasů je v každém okamžiku jejího života současně sudý a lichý apod. Nebo ji (osobu, kterou konstruujeme jako Emu B.) vybavíme jen vlastnostmi, v nichž se ženy splňující v různých světech *rolí Emy* shodují (připíšeme jí tedy průnik tříd jejich vlastností). Pak to budou přesně ty vlastnosti, o nichž se mluví v textu románu (a které společně vymezují *rolí Emy*), a ve všech ostatních parametrech půjde o neurčitou bytost: její oči nebudou modré ani černé ani hnědé atd., počet jejích vlasů nebude (v žádném okamžiku jejího života) sudý ani lichý atd.³

Neměli bychom se zdržovat otázkou, které z těchto dvou monster je děsivější. Kdyby chtěl Flaubert stvořit něco takového (resp. zaměřit naše myšlenky a představivost na něco takového), stěží by mohl doufat, že tomu udělíme svoji empatii, a stěží by mohl aspirovat na status realistického spisovatele. Pokud k tomu došlo proti jeho vůli, vzniká otázka, jak si vysvětlit tak žalostné selhání.

4. Teoretické souvislosti: fikční světy a singulární reference

Je zde tedy zřejmý konflikt mezi intuicí o tom, co po nás text jako po čtenářích vyžaduje (co musíme předpokládat, resp. co si musíme představovat)

³ Analogickým způsobem konstruují Rescher a Brandom své nestandardní možné světy, v nichž neplatí klasické zákony sporu a vyloučeného třetího; srov. Rescher – Brandom (1980).

vat, má-li text plnit své narativní funkce), a výsledky teoretických úvah ve stylu G. Currieho o tom, co nám text umožňuje: závěr zní, že text nám znemožňuje dělat přesně to, co po nás vyžaduje. Zároveň se zde nabízí zajímavá konfrontace s teorií *fikčních světů*. Jednou z funkcí, které má tato teorie plnit, je zajistit fikčním textům předmětnou sféru (*universe of discourse*), do níž by mohla směřovat referenční funkce výrazů, jako je „Ema B.“. Srov. např. Doležel (2003, 30): „tím, že zavádí možné světy jako univerzum fikčního diskursu, poskytuje naše sémantika legitimitu pojmu fikční reference“. Právě jsme ale viděli, že teorie fikčních světů nedostane příležitost plnit takovou funkci, zvolíme-li postup, který uplatnil G. Currie, tj. začneme-li analýzou „samotného textu“ (větami s jejich konvenčními významy) a budeme se ptát, co nám text umožňuje identifikovat či určit a co zůstává principiálně neidentifikovatelné či nedourčené. V rámci takového způsobu kladení otázek nebude aplikace aparátu možných světů nástrojem, který nám pomůže objasnit možnost (a specifické podoby) singulární reference ve fikčních textech: naopak nutně vyústí v její popření.

Následující úvahy povedou, jak doufám, k obhajobě fikční reference proti tomuto způsobu uvažování a v tomto ohledu budou sledovat stejný cíl jako Doleželova teorie fikčních světů. Ve výkladu předmětné sféry, k níž se vztahují fikční texty narativního typu a do níž směřuje jejich singulární reference, ale nepůjdou stejnou cestou.

5. Paralela s nefikčními texty, vztahení textu k reálnému světu

Stojí za to si uvědomit, že postup, který jsme uplatnili po vzoru G. Currieho a který vedl ke konfliktu s naší čtenářskou intuicí i s aspiracemi teorie fikčních světů (viz kap. 2-4), má stejně problematické důsledky pro tzv. seriózní (tj. ne-fikční) texty – vědecká pojednání, policejní zprávy, novinové články, soukromé dopisy atd. Pokusíme-li se je interpretovat „samy o sobě“, tj. v izolaci od jejich komunikativních funkcí, dostaneme stejné výsledky jako v případě fikčních textů. Nebudeme schopni určit svět, k němuž je máme vztáhnout, protože v nich najdeme jen vymezení stavů věcí, která splňují různé možné světy. Nedokážeme identifikovat ani osoby, k nimž by v těchto textech referovaly výrazy typu „J. A. Komenský“. I když nám text poskytne identifikační deskripce, které budeme schopni spojit s takovými výrazy, nebude možné určit, ke kterému světu je máme vztáhnout, aby v něm vyčlenily jedno individuum (to, které je v něm jedinečně

splňuje) a aby toto individuum zafixovaly jako referent jména vzhledem ke všem možným světům. Nepomůže nám ani kripkeovské schéma řetězu užívání jména a jeho zakotvení v aktu křtu, nemůžeme-li určit svět, do něž máme situovat akt křtu (jako výchozí fixaci referentu výrazu „J. A. Komenský“) a z něho se odvíjející řetěz. Tento výraz pak tedy může označovat jen *rolí Komenského* – chápanou (tak jako role Emy) jako funkce z možných světů do individuí.

V takové bizarní pozici se při četbě běžných textů nikdy neocitáme, protože se k nim nestavíme tak, jak jsme to (vzhledem k fikčním textům) zkusili po vzoru G. Currieho. K teoretickým pojednáním, novinovým článkům, privátním dopisům atd. nepřistupujeme v nějakém neutrálním modu, lhostejném vůči jejich komunikativní funkci, ale typicky jako k potenciálnímu zdroji informací *o reálném světě* – nebo alespoň jako k textům, které se tak prezentují. Nemáme-li důvod k nedůvěře, čteme je, abychom se v nich něco dozvěděli o světě v těch parametrech, jichž se týkají propoziční obsahy jejich vět. Pokud jim nedůvěřujeme, přistupujeme k nim jako ke zdroji informací o tom, jaký obraz reálného světa v nás chce vyvolat autor. V obou případech je předem fixován svět, k němuž vztahujeme věty takového textu a jeho složky – například vlastní jména s jejich referenční funkcí. Tu chápeme jako zakotvenou v aktech křtu (přirazení jmen individuí), které se odehrály *v reálném světě*.

Teze: Přesně stejně také narativní fikční text zaměřuje naše myšlení a představivost k reálnému světu. Rozdíl je jen ve zvláštním modu tohoto zaměření: běžně ho vyjadřujeme obratem „jako by“ („als ob“).⁴

6. Text jako záznam promluv reálného vypravěče o reálném světě

V každém případě se zdá, že máme dobré důvody odmítnout postup, který jsme testovali v předchozích kapitolách. Místo abychom začali analýzou „samotného textu“ v abstrakci od jeho literárních funkcí, položíme si otázku, co musíme jako čtenáři udělat (myslet, akceptovat, představovat si)

⁴ V anglosaské literatuře se někdy ve stejném smyslu užívá výraz „pretence“, nebo (po vzoru Kendella Waltona) termín „make-belief“. Nadále budu modus *jako by* vyznačovat pomocí indexů připojených k názvům příslušných postojů či aktů: *vzít na vědomí_{JB}*, *předpokládat_{JB}*, *představovat si_{JB}* apod.

má-li pro nás narativní fikční text plnit své literární funkce. Nejobecnější odpověď, která, bude-li přijata, poskytne rámec pro všechny ostatní (a zároveň neponechá žádný prostor pro konflikty, které vyvstaly výše), zní takto:

Funkce narativního fikčního textu od čtenáře vyžadují, aby přistupoval_{JB} k jeho větám jako k záznamu promluv reálné osoby – vypravěče, který líčí běh událostí v reálném světě. K roli čtenáře zároveň patří, že promluvám vypravěče uděluje_{JB} apriorní důvěru – od níž ustoupí teprve tehdy, když se vypravěč projeví jako (v nějakém ohledu) nespolehlivý.

Z logického hlediska funguje *jako by* jako operátor a je třeba dbát na to, aby byl aplikován na správném místě. Nevztahuje se na autorovo užívání vět ani na jeho literární výkon:⁵ Flaubert přímočaře (v nemodifikovaném smyslu) užívá větu jako indikátor, že na daném místě máme předpokládat_{JB} promluvu vypravěče, v níž se užívá právě tato věta.⁶ Podobně neplatí, že máme myslet_{JB} vypravěče jako osobu, která předstírá, že mluví o reálných stavech věcí: máme předpokládat_{JB}, že mluví o reálných situacích a událostech (a v rámci toho referuje k reálným individuím). Jediný, jehož výkony se odehrávají v modu *jako by*, je tedy čtenář. Jak je snad zřejmé, tyto výkony tím nejsou *zpochybněny*, ale *modifikovány*: „předpokládat_{JB}“ čteme „předpokládat v modu jako by“, a ne „jako by předpokládat“.

7. Konfrontace s teorií „fiktivních promluv“

Podle tzv. „fictive utterance theorists of fiction“ (Gregory Currie, David Davies, Peter Lamarque, Stein Olsen aj.), tj. teoretiků, kteří vymezují povahu fikčního vyprávění pomocí pojmu „fiktivní promluvy“, náš postoj k textu jako k fikčnímu vyprávění předpokládá, že autorovi připisujeme (griceovskou) intenci vyvolat v nás soubor jistých přesvědčení_{JB} (tj. soubor „make-beliefs“), a to na základě našeho rozpoznání tohoto záměru; srov.

⁵ Tím spíše platí, že autor nic nepředstírá, pokud jde o sémantickou funkci jeho jazykových aktů: v tom se shodují s L. Doleželem (proti J. Searlovi a dalším): srov. Doležel (2003, 26, 41).

⁶ „Ema se probudila“ ve Flaubertově užití tedy nefunguje jako věta v „běžné“ komunikaci – ale je matoucí upírat tomuto útvaru status věty, jak to nacházíme např. v Martínez-Bonati (1981, kap. III). Není důvod popírat, že jde o tutéž větu (jako typ), kterou užije běžný mluvčí, když řekne „Ema se probudila“.

např. Davies (2007). V rámci interpretační vstřícnosti pak naplňujeme tuto autorovu intenci, tj. přijímáme příslušná přesvědčení_{JB}. V tomto pojetí tedy operátor *jako by* vymezuje modus postoje (k propozičním obsahům vyjádřeným větami románu), který v nás, jak předpokládáme, zamýšlí vyvolat autor textu. V pojetí, které zde obhajují já, se operátor *jako by* vztahuje na náš předpoklad o existenci vypravěče jako reálné osoby, která vypovídá o reálném světě.

Interpretační postoj čtenáře vyžadovaný narativním fikčním textem se zde tedy vymezuje dvojitým způsobem:

Davies: Čtenář věří, že autor zamýšlí způsobit, aby čtenář věřil_{JB}, že *p*. V důsledku toho (tj. v důsledku rozpoznání této intence) plus interpretační vstřícnosti čtenář věří_{JB}, že *p*.

P. K.: Čtenář věří_{JB}, že existuje reálná osoba (vypravěč), která tvrdí, že *p*; v důsledku toho plus presumpce_{JB} spolehlivosti vypravěče čtenář věří_{JB}, že *p*.

Ve druhém pojetí se, obecně řečeno, operátor *jako by* aplikuje na naše přesvědčení týkající se komunikativní a (v rámci toho) reprezentační funkce výskytu jednotlivých vět v textu, jinými slovy: operátor *jako by* určuje modus našeho přesvědčení o tom, jakým způsobem se propoziční obsah těchto vět uvádí do komunikace. Apriorní (ale ne neodvolatelný) předpoklad pravdivosti těchto propozic je pak zabudován do komplexního postoje, na nějž se vztahuje operátor *jako by*. Teprve v tomto rámci tedy může dojít ke konfrontaci přesvědčení v modu *jako by* a přesvědčení s tímž propozičním obsahem, ale v přímočarém (neposunutém) modu – například přesvědčení_{JB}, že Napoleon vtrhl do Ruska, a přesvědčení, že Napoleon vtrhl do Ruska.⁷

8. K myšlence „přesunu centra“ (konfrontace s M.-L. Ryanovou)

Na jednom z míst, v nichž Marie-Laure Ryanová objasňuje svůj známý pojem „recentering“, se říká:

⁷ K problému slučitelnosti „believing“ a „make-believing“ viz např. Gibson (2007, 166) a Friend (2008, 156).

Jaký má pro lidi smysl „předstírat reprezentaci skutečnosti“? Má-li pro nás fikce nějaký význam, je to proto, že předkládá představivosti svět a představivost nachází potěšení v tom, že se zaobírá tímto světem. Ale i když fikce reprezentuje cizí svět, reprezentuje ho, jako by byl aktuální, protože se vyjadřuje v indikativu spíše než v kondicionálu. Tím, že na sebe bere podobu faktičnosti (skutečnosti), vyzývá své uživatele, aby se v představivosti přemístili do tohoto cizího světa. Tento akt přemístění označuji jako „fikční přesun centra“. (Ryan 2010, 14)⁸

Myslím, že to, co se zde podává jako náš interpretační postoj k fikci, je zbytečně komplikovaná konstrukce. Pokud je fikční vyprávění schopné nás zaujmout díky tomu, že naší představivosti umožňuje zaobírat se jiným (než aktuálním) možným světem a poskytuje jí požitky, které z toho plynou, k čemu je pak zapotřebí, aby byl tento svět prezentován a přijímán jako aktuální? Proč je nutné, abychom se do něj ve své představivosti přenášeli i se svým pojmem „actual“ (který po tomto manévru bude odkazovat k tomuto „cizímu“ světu)? A obráceně: je-li pro interpretaci fikčního vyprávění podstatné, že je přijímáme jako vyprávění o aktuálním světě (v modu „jako by“, resp. „make-believe“), proč bychom u toho nemohli zůstat? K čemu je nutný manévr „přesunu centra“ a čím lze odůvodnit tvrzení, že „každá fikce vyžaduje přesun centra, má-li být správně pochopena“?⁹ Ryanová může odpovědět, že je to nezbytné kvůli určení světa, vzhledem k němuž se má věta užitá v literárním textu pravdivostně ohodnotit.¹⁰ Věty fikčního narativního

⁸ Srov. tamtéž: “When recentering takes place, the text is no longer regarded as making statements about the real world, or at least, not directly, and the fictional world is contemplated for its own sake.” Na jiném místě charakterizuje autorka povahu „recentering“ takto: “Consciousness relocates itself to another world and... reorganizes the entire universe of being around this virtual reality” (Ryan 2001, 104). Srov. též “[o]nce we become immersed in a fiction, the characters become real for us, and the world they live in momentarily takes the place of the actual world” (Ryan 1991, 21).

⁹ “All fictions require recentering to be properly understood” (Ryan 2010, 14). Srov. tamtéž: “... recentering is a logical operation which we deliberately perform whenever we read (or watch) a work of fiction.”

¹⁰ “The importance of the judgment of fictionality lies in the fact that it determines with respect to which world the information transmitted by the text should be evaluated. If the judgment says ‚fiction‘, this information concerns a non-actual possible world, where it is automatically true (unless the narrator is judged unreliable), since the world is created by the text. If the judgment says ‚non-fiction‘, the information de-

textu nevztahujeme (při určování reference jejich složek a při jejich pravdivostním ohodnocování) k reálnému světu, ale k odlišnému možnému světu zkonstruovanému autorem: to je (podle Ryanové) podstatný rys našeho postoje k fikci. Já se naopak domnívám, že rozdíl mezi postojem k textu jako k fikčnímu a jako k „serióznímu“ není v určení světa, k němuž text vztahujeme (tím je pokaždé reálný svět), ale ve způsobu tohoto vztahování: „seriózní“ (tj. ne-fikční) texty vztahujeme k reálnému světu zcela přímočaře, fikční texty v modu „jako by“. Viděli jsme, že Ryanová využívá pro rozlišení fikčního a ne-fikčního postoje jak rozdílů mezi světy, tak pojmu „jako by“ (resp. „make-believe“). Jak jsem už řekl, jednu položku z tohoto arzenálu pokládám za nadbytečnou: můj návrh zní obětovat první (a s ní i koncept „recentering“).

V každém případě se nevyhneme konfrontaci *dvojího pojetí interpretačního postoje k fikci*:

Ryanová: Předmětem myšlenkových a imaginativních výkonů interpreta je fikční svět přijímaný_{JB} jako aktuální. Jeho fakty jsou stavy věcí popsané v textu plus stavy věcí, které jsou fakty ve výchozím aktuálním světě, a které čtenář přenáší do fikčního světa v rámci jeho kompletace. Stěhovat lze pouze stavy věcí, které jsou slučitelné s tím, co se explicitně říká, implikuje nebo indikuje (ve smyslu Griceových implikatur) v literárním textu.

Předpokládají se zde tedy tyto pohyby: přesun do jiného světa (čtenář se v rámci „recentering“ situuje ve své představivosti do nového „centra“) + transport materiálu ze starého světa, který je v novém světě povolen (projde celní kontrolou nastavenou textem).

P. K.: Předmětem myšlenkových a imaginativních výkonů interpreta zůstává reálný svět našeho života.

Předpokládaný pohyb: čtenář upravuje_{JB} obraz reálného světa podle instrukcí textu (nahrazuje některé fakty jinými stavy věcí, které přijímá_{JB} jako fakty).

To lze shrnout ve dvou protikladných *instrukcích pro čtenáře fikce*:

scribes the actual world, but since this world exists independently of the text, it can be either true or false with respect to this world.” (Ryan 2010, 15)

Ryanová: Přesuň se (ve své představivosti) do fikčního světa a vezmi s sebou všechno, co ti dovoluje text!

P. K.: Zůstaň (ve své představivosti), kde jsi, a nehýbej s ničím, pokud to nevyžaduje text!¹¹

Tento konflikt se odráží v rozdílném postoji k problému *úplnosti/neúplnosti fikčního světa*:

Ryanová: Pozice interpreta narativního fikčního textu je v podstatném ohledu dvojznačná. Čtenář ví, že fikční svět konstruovaný textem je nutně *neúplný* (z důvodů, na něž jsme narazili v konfrontaci s G. Curriem, viz kap. 2). Zároveň k němu ale přistupuje¹² jako k úplnému¹² a kde to jde, vyplňuje (ve shodě s principem „minimální odchylky“) prázdná místa materiálem z reálného světa.¹³

P. K.: Taková schizofrenie není nutná. Fikční svět díla je pro čtenáře reálný svět našeho života (upravený podle instrukcí textu), a tedy něco principiálně úplného: o každém možném stavu věcí platí, že buď je, nebo není faktem v tomto světě.

¹¹ Tak jako výše nejde jen o to, co se explicitně říká v textu, ale i o to, co jeho věty implikují a co se v nich naznačuje (co se komunikuje na úrovni nepřímého významu).

¹² Srov. např. “The reader knows that fictional worlds are incomplete, but when he ‘plays the game’, when he submerges in a fiction, he pretends to believe that this world is complete” (Ryan, net 1). Srov. též: “From the point of view of the ‘actual actual world’ the worlds of fiction are discourse-created non-actual possible worlds, populated by incompletely specified individuals; but to the reader immersed in the text the TAW is imaginatively real, and the characters are ontologically complete human beings” (Ryan, net 2).

¹³ Srov. např. “This analysis entails a principle which has come to be recognised as fundamental to the phenomenology of reading. Various described as ‘the principle of minimal departure’ (Ryan), the ‘reality principle’ (Walton), and the ‘principle of mutual belief’ (Walton again), the principle states that when readers construct fictional worlds, they fill in the gaps... in the text by assuming the similarity of the fictional world to their own experiential reality” (Ryan, net 2). K souvislosti principu minimální odchylky s postojem k fikčnímu světu jako k úplnému viz tamtéž: “The principle of minimal departure presupposes that fictional worlds, like the PWs postulated by philosophers, are ontologically complete entities: every proposition *p* is either true or false in these worlds. To the reader’s imagination, undecidable propositions are a matter of missing information, not of ontological deficiency.”

9. Možné světy a modální intuice (S. Kripke)

V úvahách o vztahu narativního fikčního textu a reality vycházím (ostatně stejně jako Ryanová) z intuitivního předpokladu, že literární funkce takového textu ode mne vyžaduje víc, než abych vzal neutrálně na vědomí propoziční obsahy vyjádřené jeho větami nebo abych si neutrálně představoval situace a události, které jsou v nich vymezeny (tj. abych nechal jejich obrazy „volně plynout“ ve své mysli). Narativní funkce textu vyžaduje, abych propozice vyjádřené jeho větami interpretoval_{JB} jako obsahy promluv, jejichž mluvčí (vypravěč) přejímá záruku za to, že situace a události, které v nich popisuje, se *doopravdy* staly (jinými slovy, že jsou *skutečné*). A protože náš pojem „doopravdy“ (a náš pojem „skutečnosti“) je zakotven v *reálném světě* našeho života a odkazuje nás do něj, mohu to parafrázovat takto: literární funkce fikčního narativního textu si zpřístupním, budu-li jeho věty chápat_{JB} jako záznam promluv vypravěče jakožto obyvatele reálného světa, který mi sděluje, co se v tomto světě stalo, a budu-li ochoten akceptovat_{JB}, resp. představovat si_{JB}, že reálný svět je takový, jak mi ho vypravěč líčí. To zpravidla zahrnuje, že přijímám_{JB}, resp. představuji si_{JB}, že reálný svět je v některých ohledech jiný, než jaký je podle toho, co předpokládám mimo dosah operátoru *jako by* (na základě informací ze zdrojů, na něž se spoléhám v každodenním životě).

V žargonu možných světů to samozřejmě můžeme vyjádřit slovy, že uvažuji o jiném možném světě a přijímám ho_{JB} jako reálný. To je neškodné, pokud neztratím ze zřetele, že jde jen o parafrázi předchozí formulace, tj. o jiný způsob, jak říci, že uvažuji o jiném *stavu reálného světa* a přijímám_{JB} ho jako reálný. Problémy nastanou, nechám-li se svou terminologií či instrumentální stránkou svého pojmového aparátu svést k předpokladu, že se jako čtenář stěhuji (ve svém myšlení či ve své představivosti) do jiného světa a začnu-li uvažovat o jeho ontologickém statutu, o jeho vztahu k reálnému světu, o možnostech jeho identifikace, o jeho úplnosti či neúplnosti, o podmínkách identity postav napříč světy atd. To mě vrátí do situace ze samého počátku našich úvah (inspirovaného G. Curriem), obecně řečeno do stavu, kdy mi můj technický aparát podsouvá problémy, které sám generuje (viz kap. 2-5): jinými slovy, stávám se výkonným orgánem svého aparátu, místo aby mi sloužil k účelům, pro něž byl vytvořen.

Přesně tak chápu varování, které vyslovil Saul Kripke, původce moderní aplikace pojmu možných světů v logické sémantice (předchozí poznámku je možné číst jako vztažení jeho obecné výzvy na sféru literární interpretace):

Filosof „možných světů“ musí bezpochyby dbát na to, aby ho jeho technický aparát nenutil klást si otázky, jejichž smysluplnost není podložena našimi původními intuicemi o možnosti, od nichž aparát odvozoval své funkce.¹⁴

V této souvislosti Kripke konstatuje, že ve stejném smyslu jako termín „possible world“, ale s menším rizikem filosofických konfúzí, můžeme použít výraz „possible state (or history) of the world“ (Kripke 1980, 15). Obrat „možný stav světa“ může mít stěží přirozenější uplatnění než v kontextech následujícího druhu: uvažujeme o tom, že *stav reálného světa* by v nějakém ohledu, který nás zajímá, mohl být jiný. Například soudíme, že kdyby Blücher nedorazil včas na bitevní pole, Napoleon by porazil Wellingtona. V tomto bodě by tedy byl stav světa jiný, než aktuálně je, jinými slovy: toto místo v celkovém určení stavu reálného světa (vymezené otázkou „Kdo vyhrál u Waterloo?“) by bylo obsazeno jiným faktem. Ve stejném smyslu si klademe o individuích, která nás zajímají, dejme tomu o sobě, otázky typu: „Kde jsem dnes mohl být (co jsem dnes mohl mít, čemu všemu jsem se mohl vyhnout), kdybych ji tenkrát nepozval na skleničku?“ V takových úvahách neopouštíme reálný svět: uvažujeme o alternativních scénářích vývoje *tohoto* světa a o jejich důsledcích pro jeho obyvatele, aniž bychom je kvůli tomu přenášeli do jiných světů.¹⁵

Jsem přesvědčen, že *stejně povahy* jsou naše myšlenkové a imaginativní výkony při interpretaci narativního fikčního textu. Jeho narativní funkce nás zaměřuje na reálný svět a vyžaduje od nás, abychom předpokládali či abychom si představovali, že *s tímto světem* to je v některých ohledech jinak.¹⁶ Rozdíl je pouze ve specifickém modu tohoto předpokladu či této představy, který vyjadřujeme obratem *jako by*.

¹⁴ “Certainly the philosopher of ‘possible worlds’ must take care that his technical apparatus does not push him to ask questions whose meaningfulness is not supported by our original intuitions of possibility that gave apparatus its points” (Kripke 1980, 18).

¹⁵ Právě tak vyznívá Kripkeho příklad, v němž se v žargonu možných světů parafrázuje výpovědi o variantách, které mohou nastat při vrhání dvojice hracích kostek (Kripke 1980, 17).

¹⁶ V tomto duchu si vykládám Kunderova slova o „říši možností“ citovaná v mottu známé Doleželovy knihy (Doležel 2003): „Román nezkoumá skutečnost, nýbrž existenci. A existence není to, co se stalo, existence je říše lidských možností, všechno, čím se člověk může stát, všechno, čeho je schopen. Romanopisci kreslí *mapu existence* tím, že ob-

10. Důsledek pro chápání fikčních světů

V tomto pojetí je tedy fikční svět daného fikčního narativního textu ta podoba *reálného světa*, kterou musíme přijmout_{JB}, či představovat si_{JB}, má-li pro nás text plnit své literární funkce. V současné filosofii literatury tomu (zhruba) odpovídá teze Davida Daviese, že fikční vyprávění vyžaduje *reálný rámec* („real setting“), jehož existenci předpokládáme *v plném smyslu* (ne v modu *make-believe*). *O tomto reálném rámci* jako čtenáři věříme (v modu *make-believe*), že se v něm odehrává to a to.¹⁷

11. Uchování významů ve fikčním textu

Předpoklad_{JB}, že narativní fikční text zaznamenává reálné promluvy reálné osoby (vypravěče) a zaměřuje naše myšlení a představivost k reálnému světu, mi jako čtenáři umožňuje počítat zcela neproblematicky s tím, že výrazy užívané v tomto textu si v něm uchovávají své funkce z běžné komunikace v reálném světě. Mohu to předpokládat bez toho, že bych (ve svém myšlení či představivosti) přesouval naše komunikativní návyky a konvence ze světa, v němž se ustálily jako výsledek reálného vývoje komunikace, do jiných světů, aby tam byly k dispozici jejich obyvatelům (vypravěči a literárním postavám) pro jejich promluvy. I v tomto ohledu *zůstává vše na svém místě*. To mi umožňuje bez dalšího předpokládat, že vypravěč mluví stejným jazykem jako já (tj. česky, a ne fikčním duplikátem češtiny), případně že mluví francouzsky (francouzštinou, jak se vyvinula z latinských kořenů v našem světě, nikoli jejím fikčním duplikátem) a text románu je překladem jeho promluv

jevují tu či onu lidskou možnost.“ Takový výkon bezpochyby podává autor, který si klade otázky typu: co se ve světě, ve kterém žijeme (v poměrech vládnoucích v té a té době), stane s člověkem vybaveným takovým a takovým typem intelektu, vnímavosti, představivosti, emocionality a morální intuice, když se ocitne v tom a tom prostředí, dostane takové a takové příležitosti nebo je vystaven těm a těm druhům omezení?

¹⁷ “Arguably, all intelligible fictional narratives require a real setting of some kind, since their fictional worlds presuppose that the reader locates what is explicitly narrated in a wider framework. ... Given such a real setting for a fiction, we are prescribed to imagine something *of* that setting, but we are not prescribed to imagine *that* the real setting itself exists – this is something we are assumed to believe, not something we either imagine or are prescribed to imagine.” Srov. Davies (2012, 80).

do mé mateřštiny. Jak uvidíme v kap. 13, má to závažné důsledky i na úrovni reference (pro způsob identifikace fikčních entit).

12. Fikční jména

Pokud text obsahuje výrazy, které nejsou převzaty z reálné komunikace, pak jako čtenář automaticky předpokládám_{JB}, že v něm plní tytéž funkce, pro něž by byly způsobilé v běžné komunikaci (funkce, které indikují svou morfologii a svou roli v syntaktické struktuře vět). Například: vyskytuje-li se v textu výraz, který se syntakticky chová jako běžné vlastní jméno, dejme tomu „Ema B.“, přijímám to jako výzvu, abych na tomto místě předpokládal_{JB}, že vypravěč v dané fázi vyprávění užívá vlastní jméno, a to k označení osoby, které bylo toto jméno uděleno v reálném světě na počátku řetězu, jehož součástí je vypravěčova promluva.¹⁸

13. Identifikace fikčních postav

To mi (jako čtenáři) nabízí velmi přímočarý způsob identifikace fikčních postav. Mám-li určit referent vypravěčova užití jména „Ema B.“, mohu jednoduše odkázat_{JB} na řetěz právě uvedeného druhu nebo ho převést do podoby komplexní deskripce a identifikovat_{JB} Emu jako osobu, která v reálném světě splňuje deskripci: „osoba, které bylo jméno ‚Ema B.‘ uděleno na počátku řetězu, jehož součástí je *tato* vypravěčova promluva“. Takové deskripce, obsahující ve své metajazykové složce citaci jména samotného a odkaz na jeho užití vypravěčem, můžeme označit jako „parazitní“ či „deferenční“. Kromě toho mohu Emu identifikovat_{JB} pomocí „neparazitních“ deskripcí, které sesbírám při sledování vyprávění, například „manželka městského lékaře v Yonville v letech...“. Oba typy deskripcí spolu s důvěrou, kterou vkládám_{JB} v jejich splňování, mi nabízejí plnohodnotnou identifikaci osoby (samozřejmě v modu *jako by*): při interpretaci fikčního textu mohu postupovat tak, jako by byla jménem „Ema B.“ vyčleněna jedna osoba

¹⁸ Jak je zřejmé, chápu fungování vlastních jmen v každodenní komunikaci ve shodě s Kripkeho kauzální teorií a jsem přesvědčen, že literární funkce narativního fikčního textu od čtenáře vyžadují, aby právě takovým způsobem přistupoval ke jménům, která se v něm vyskytují. „Kripkeovský“ výklad role jmen v literárním textu obhájuje např. Thomas Pavel v Pavel (2012, 54n.).

v reálném světě – totiž osoba, která v tomto světě jedinečně splňuje uvedené deskripce. Mohu tedy s dostatečnou určitostí uvažovat_{JB} o Emě B. jako o reálné osobě, která má vlastnosti uvedené v textu a je stejně určitá ve všech ostatních parametrech, o nichž se v textu nemluví – tak jako v běžných úvahách předpokládám o reálných osobách, že jsou určité i v těch parametrech (v těch obligatorních způsobech určenosti lidských bytostí), které nejsem schopen obsadit žádným určením.¹⁹

To samozřejmě neznamená, že mezi úvahou o reálné osobě a úvahou o románové postavě *jako* o reálné osobě není podstatný rozdíl. Tvzení, že uvažuji_{JB} o Emě B. jako o reálné osobě, neimplikuje (mimo dosah operátoru *jako by*) *de re* tvrzení o Emě B. – totiž, že *o ní* platí, že *o ní* uvažuji jako o reálné osobě. Ze stejného důvodu tvrzení, že uvažuji_{JB} o Emě B. jako o reálné osobě, neumožňuje (mimo dosah *jako by*) existenční generalizaci, totiž že existuje entita, o níž uvažuji jako o reálné osobě. Jinými slovy, předmětná vztáženost mých úvah a představ nezakládá regulérní vztah mezi mnou a nějakou entitou, ale pouze intencionální *vztah bez konverze*.²⁰ Tvzení, že čtenář přistupující k textu způsobem, který jsem popsal (viz zvl. kap. 6), získává dostačující prostředky identifikace fikčních postav, je třeba číst tak, že má vše, co je třeba k tomu, aby uvažoval_{JB} o postavách s dostatečnou určitostí (s plnou identifikační silou): nikoli tak, že existují (ať už s jakýmkoli ontologickým statutem) postavy, které čtenář identifikuje v rámci jejich předmětné sféry.

V těchto úvahách si, jak je zřejmé, vystačíme s velmi slabým, ontologicky nenáročným pojmem literárních postav: pro každý narativní fikční text jsou to osoby, jejichž existenci v reálném světě musíme předpokládat_{JB}, má-li pro nás text plnit své literární funkce. Ve *Vojně a míru* musíme takto předpokládat_{JB} například Andreje Bolkonského a Napoleona jako obyvatele reálného světa. V tomto ohledu je jejich status stejný – na této úrovni je tedy zachován „princip ontologické stejnorodosti“, jehož respektování je podle L. Doležela „nezbytnou podmínkou pro spoluexistenci, interakci a komunikaci fikčních osob“ (Doležel 2003, 32). Jiná věc je, že o existenci jedné

¹⁹ To je v rozporu s obecně sdílenou tezí o principiální neúplnosti fikčních entit. Srov. např. Doležel (2003, 36): „Je zřejmé, že kdo popře neúplnost fikčních entit, zachází s nimi jako s entitami skutečnými.“ Mé stanovisko viz např. v Koťátko (2010).

²⁰ Pro standardní vztahy platí, že je-li A v takovém vztahu k B (například A je nalevo od B), pak je B v konverzním vztahu k A (B je napravo od A). Ke specifické povaze intencionálních vztahů jako „relations without converse“ viz zvl. Prior (1971, 136 aj.).

z těchto postav jsem zároveň přesvědčen i mimo dosah operátoru *jako by*. Je nepochybné, že také toto vědomí a aktivace historických znalostí, které jsem schopen spojit se jménem „Napoleon“, může hrát významnou roli v naplňování literárních funkcí textu (více o tom v Kot'átko 2010).

14. Paralela: jména v běžné komunikaci

Komu se nástroje identifikace fikčních postav, o nichž jsem právě mluvil, zdají příliš neurčité, měl by si uvědomit, že jde o stejný druh identifičnických prostředků, s jakým se často musíme spokojit v běžných komunikativních situacích. Dejme tomu, že jsem svědkem toho, jak můj učitel dějepisu opakovaně vyslovuje jméno „Bernadotte“. Chci-li mít něco z jeho výkladu, měl bych se spolehnout na jeho serióznost a předpokládat, že se svými promluvami zapojuje do řetězu užívání jména „Bernadotte“, který je zakotven v aktu křtu, v němž bylo toto jméno přiřazeno reálné osobě. Navíc se spoléhám na to, že deskripce, které posbírám z jeho výkladu, například „Napoleonův maršál, který se stal švédským králem“, jsou splňovány přesně jednou osobou v reálném světě – touž osobou, které bylo přiřazeno jméno „Bernadotte“ na počátku řetězu, jehož součástí je učitelovo užívání tohoto jména.

Při četbě Flaubertova textu činím (v modu *jako by*) přesně stejný druh předpokladů týkající se vypravěčova užívání jména „Ema B.“, abych umožnil textu plnit jeho literární funkce (abych si mohl poskládat příběh románu). Součástí mé role čtenáře narativního fikčního textu je spoléhání_{JB} na vypravěče, důvěra, kterou vkládám_{JB} do jeho promluv – dokud se neprojeví jako nespolehlivý. Tato důvěra_{JB} je tedy sice apriorní, ale nikoli neodvolatelná (vrátíme se k ní v kap. 16).

15. Singulární propozice v literárním textu

Zdá se tedy, že nám nic nebrání připisovat_{JB} vypravěči schopnost vyjadřovat singulární propozice – schopnost, která je *a priori* vyloučena, vyjeme-li z potenciálu „textu samotného“, jak jsme viděli v úvaze inspirované G. Curriem (viz kap. 2). Čtenář, jenž přijme a důsledně plní roli, kterou mu nabízí narativní fikční text, nemá problém přiřadit_{JB} promluvám vypravěče obsahujícím vlastní jména zcela určité singulární obsahy. Vyskytu-

je-li se v textu věta „Emu omrzel venkovský život“, připíše_{JB} čtenář vypravěči tvrzení, které je pravdivé právě tehdy, když osobu pokřtěnou jménem „Ema“ na počátku řetězu, jehož součástí je vypravěčova promluva, omrzel venkovský život. A k roli čtenáře, který si chce zpřístupnit literární funkce textu, a v rámci toho si chce především (jako podmínku všeho ostatního) poskládat dohromady příběh Emy, patří jeho spoléhání_{JB} na to, že existuje přesně jedna taková osoba a že udělala, resp. zakoušela to, co tvrdí vypravěč.

16. Důvěra ve spolehlivost vypravěče a její limity

Obecně řečeno, součástí interpretačního postoje čtenáře narativního fikčního textu je spoléhání_{JB} na vypravěče, důvěra, kterou *a priori* vkládá_{JB} do jeho promluv – dokud se neprojeví jako nespolehlivý.²¹

K tomu může dojít z důvodů zcela analogických těm, kvůli nimž pozbýváme důvěru ve spolehlivost svých protějšků v běžné komunikaci. V každodenní konverzaci občas dospíváme k závěru, že mluvčí záměrně zkresluje jisté události (například zveličuje svůj podíl na nich) nebo není schopen posoudit jisté souvislosti (například proto, že je v zajetí jistých předsudků, dejme tomu teorie všudypřítomného spiknutí). Zároveň s tím jsme zpravidla schopni poskládat si s velkou mírou jistoty, co se *doopravdy* stalo – bez toho, že bychom měli přístup k nezávislým informacím o dané věci. Obvykle stačí čerpat ze stejného zdroje, který nám umožnil odhalit nespolehlivost mluvčího: z toho, co nám poskytl on sám, z povšechné obeznámenosti s tím, jak to ve světě chodí, případně ze zkušeností s daným typem mluvčích.

Konstrukce literárního díla může počítat s tím, že tyto schopnosti nabyté v každodenní komunikaci přeneseme do interpretace textu a uplatníme je

²¹ Fenomén nespolehlivého vypravěče je častým předmětem literárněteoretických diskusí. Pro některé autory byl důvodem k zavedení jistých inovací do teorie literární interpretace (konstrukcí, v nichž vystupuje takzvaný *implikovaný* či *modelový* či *ideální* autor). Podle mého názoru jde o příklad situace, kdy se scéna (konkrétně konfigurace: autor – vypravěč – postavy – čtenář) zaplňuje teoretickým lešením, které nám spíše zakrývá, než odhaluje podstatné distinkce (srov. Úvod). Konstrukty tohoto druhu snadno generují vlastní distinkce a otázky, které nemají oporu v čtenářském postoji k textu – zkrátka spouštějí mechanismus, před nímž nás varoval Saul Kripke (viz výše kap. 9). Pokusil jsem se k tomu něco říci v Kožátko (2008).

při posuzování spolehlivosti vypravěče. V zamýšleném literárním účinku textu může hrát podstatnou roli diskrepance mezi tím, jaká tvrzení připisujeme_{JB} vypravěči a jak rekonstruuje_{JB} skutečný průběh událostí.

17. Kolaps vyprávění a performativní reprezentace

Prozatím byla řeč o *dílčí nespolehlivosti* zasahující jisté parametry vypravěčova výkonu (například jeho výklad souvislostí jistého druhu nebo jeho prezentaci své vlastní role v popisovaných událostech). Zároveň předpokládáme, že tato nespolehlivost je v principu *korigovatelná*: současně s jejím odhalením (a ze stejných zdrojů) jsme zpravidla schopni určit_{JB}, „jak se věci doopravdy mají“.

Od toho musíme striktně odlišit případ vyprávění, v němž se svět (*reálný svět*) prezentuje jako místo, ve kterém není prostor pro situace a události, o nichž by bylo možné podat koherentní, určité a spolehlivé svědectví. Jazyk, kterým se mluví v tomto světě, pak také nemá schopnost ani příležitost sloužit jako nástroj takových svědectví: nástroj k vyjadřování koherentních a určitých propozičních obsahů s plnou asertivní silou. Zahrneme-li vypravěče, pozici, z níž odvádí svůj výkon, a nástroje, kterých při tom užívá, do tohoto světa – do světa, o němž vypráví, ztrácíme možnost vkládat_{JB} čtenářskou důvěru do jeho promluv, tj. spoléhat se_{JB} na pravdivost vyjádřených propozic, protože samotná schopnost jeho promluv vyjadřovat koherentní a určité propoziční obsahy je vážně zpochybněna. Vyjadřování propozičních obsahů (vymežujících jisté stavy věcí jako podmínky své pravdivosti) pak nemůže být základním zdrojem reprezentace světa, o němž se vypráví. Reprezentační funkci naopak přejímá způsob, jakým se vyjadřování propozičních obsahů permanentně komplikuje a selhává – a tedy způsob, jakým se rozpadá vypravěčova snaha o kontinuální, koherentní a určité líčení běhu událostí. Tímto způsobem reprezentace světa v narativním fikčním textu, a tedy tímto typem vyprávění, který pokládám za literárně produktivní a mimořádně silný, chopí-li se ho velký autor (viz kap. 19), bych se chtěl na závěr zabývat podrobněji. Cílem je ukázat, že i zcela nekonformní vyprávění založené na destrukci narativních forem a rozbití obrazu spolehlivého vypravěče vyžaduje od čtenáře „standardní“ interpretační postoj, který jsme obecně charakterizovali v kap. 6.

18. Radikální vyprávění

Mluvíme tedy o typu vyprávění, v němž je narušena samotná schopnost vypravěčových promluv kontinuálně vyjadřovat určité a koherentní propoziční obsahy a v němž toto permanentní zadržávání a selhávání nemá indikovat vypravěčovu momentální indispozici. Má být prostředkem zobrazení, nebo spíše *předvedení* jeho trvalé situace; má charakterizovat pozici, z níž se pokouší odvést svůj vypravěčský výkon, a svět, v němž je tato pozice zakotvena. Tento typ vyprávění označuji jako *radikální* a to, co se v něm odehrává, lze shrnout takto:

- a) Pozice, z níž vypravěč odvádí svůj výkon, včetně narativních prostředků, které má k dispozici, je důsledně zahrnuta do světa, o němž se vypráví (to je něco jiného než vypravěčova účast v příběhu, který vypráví – nespadá sem tedy každý případ tzv. homodiegetického vypravěče).
- b) Vzhledem k povaze světa, o němž (a v němž) se vypráví, zahrnutí vypravěčovy pozice do tohoto světa má za následek, že jeho promluvy nemohou plynule plnit funkce známé z běžné komunikace, a že právě tak selhávají osvědčené narativní techniky.
- c) V důsledku toho základní rysy světa díla nejsou vymezeny v propozičních obsazích vypravěčových promluv a zpřítomněny imaginativním „vyplněním“ těchto obsahů (srov. Martínez-Bonati 1983), ale spíše předvedeny ve způsobech, jakým je vyjadřování propozičních obsahů permanentně narušováno, případně zcela selhává.

19. Příklad: Samuel Beckett

a) *Aspirace*. Nejsilnějším příkladem radikálního vyprávění, jehož jsem si vědom, je Beckettova románová trilogie (Beckett 1979) a jeho pozdější prozaické texty. Rozpad vyprávění a rozvrácení samotné schopnosti vět přenášet své konvenční významy do významů vypravěčových promluv zde není výsledkem jazykových či narativních experimentů; jejich cílem není testovat možnosti vyprávění nebo ukázat jeho meze. *Východiskem* je Beckettova výjimečná citlivost k těm stránkám života, které nám brání přijímat obraz světa jako řádu, v němž je místo pro smysluplné jednání a kontinuální živo-

topis.²² *Důsledkem* je Beckettova nikdy nekončící snaha najít literární formu, která nechá promlouvat chaos, a s tím spojené odmítnutí tzv. „realistických“ narativních technik, které (v Beckettových očích) slouží k zastírání chaosu a vytváření iluze řádu.²³

b) *Prostředky předvedení chaosu*. Podstatné je, že sám vypravěč, jeho narativní prostředky a jeho výkon jsou u Becketta důsledně zahrnuty do takto chápaného světa (to je definiční rys *radikálního vyprávění*, srov. kap. 18). Jinými slovy, narativní literární dílo, a v rámci něj vypravěčův výkon, není pro Becketta ostrůvkem řádu v moři zmatku, ale produktem a způsobem existence chaosu, a tedy také svědectvím o chaosu. Text, který je nositelem takového díla, pak nemůže reprezentovat chaos tím, že ho popíše standardními prostředky: Beckettův text činí chaos viditelným ve vypravěčových opakovaných pokusech o souvislé vyprávění a v jejich kolapsecích, v neustálých revokacích toho, co bylo řečeno, ve zpochybňování významů slov, v hromadění kontradikcí, které rozkládají propoziční obsahy vypravěčových promluv a podvracejí jejich asertivní sílu, v selhávání referenční funkce zájmena první osoby, a tedy i literární funkce *Ich-formy* atd.

To jsou jen některé způsoby rozkladu vyprávění, v nichž se ukazuje povaha světa, *v němž* a *o němž* se vypráví. Čtenář ji nejen registruje, ale zakouší ji sám ve svých pokusech o kontinuální čtení a v jejich selháních.

c) *Neúplnost Beckettova světa*. O světě, který se takto *ukazuje* v Beckettových textech, pak jistě platí, že je *neúplný*: není úplný v tom striktním smyslu, že by o každém možném stavu věcí platilo buď, že je, nebo není jeho součástí: naopak každý myslitelný stav věcí si v něm může dělat stejný nárok na fakticitu. Toto konstatování o neúplnosti světa Beckettovy *Trilogie* má zcela jinou bázi než *obecná* (a široce sdílená) teze o *principiální neúplnosti fikčních světů*, kterou jsme odmítli na začátku v konfrontaci s G. Curriem. Tato teze je, jak si vzpomeneme, výsledkem jistého způsobu aplikace aparátu možných světů na literární text nazíraný v abstrakci od jeho literárních funkcí (srov. kap. 2). Naproti tomu naše současné tvrzení zní, že jednou ze *specifických* literárních funkcí destrukce vyprávění v Beckettových textech je *zobrazení světa jako principiálně neúplného*.

²² V jednom z rozhovorů to Beckett shrnuje slovy: “I can’t see any trace of a system anywhere.” Srov. Shenker (1956).

²³ “To find a form that accommodates the mess, that is the task of the artist now” (Driver 1961, 23).

d) *Beckettův realismus*. Není to úrok z reálného světa do sféry čistých literárních konstrukcí: je to výsledek Beckettova úsilí předvést povahu světa, v němž žijeme, *bez* literárního konstruktérství, a tedy bez pomoci tradičních narativních technik (viz a). Jestliže Beckettovu vypravěči slova permanentně vypovídají službu, jestliže se ukazuje, že nejsou schopna propůjčit koherentní a určité významy vypravěčovým promluvám, není to produkt artistního hraní s jazykem, ale důsledné vyvozování konsekvencí z Beckettova odmítnutí obrazu světa (reálného světa našeho života) jako řádu.

To ukazuje, že ani *radikální vyprávění* se nevymyká z obecného požadavku, který na nás, jak jsem přesvědčen, vznáší literární funkce fikčního narativního textu. Tvrdil jsem, že takový text pro nás bude schopen plnit své funkce, jen když ho budeme (v modu *jako by*) interpretovat jako záznam promluv obyvatele reálného světa – vypravěče, který nám líčí, co se v tomto světě stalo (srov. zvl. kap. 6). Beckettův text může plnit svou funkci – *předvést* chaotickou povahu světa (a ukázat, co to znamená žít v tomto světě, pokoušet se v něm sledovat určitý cíl, snažit se říci v něm a o něm něco určitého atd.), pouze když důsledně situujeme vypravěče, jeho promluvy i jejich předmět do reálného světa. Jenom pak může jeho obsedantní snaha říci něco koherentního a určitého a její permanentní selhávání sloužit k vypovídání o povaze tohoto světa.

Beckett to shrnul slovy, která chápu jako vážně míněný výrok o naší situaci ve světě, v němž žijeme, spíše než jako charakteristiku literární konstrukce, pro niž se rozhodl, nebo artistní pozice, do níž se stylizuje:

There is nothing to express, nothing with which to express, nothing from which to express, no power to express, no desire to express, together with the obligation to express. (Esslin 1965, 17)

Beckettův vypravěč v *Nepojmenovatelném* vyjadřuje totéž slovy:

The fact would seem to be, if in my situation one may speak of facts, not only that I shall have to speak of things of which I cannot speak, but also, which is even more interesting, but also that I, which is if possible even more interesting, that I shall have to, I forgot, no matter. And at the same time I am obliged to speak. I shall never be silent. Never. (Beckett 1979, 268).²⁴

²⁴ Podstatná část tohoto textu se kryje s mým příspěvkem do kolektivní monografie věnované Lubomíru Doleželovi (Fořt 2012). Velmi oceňuji příležitost předložit jej

Literatura

- BECKETT, S. (1979): *The Beckett Trilogy (Molloy, Malone Dies, The Unnamable)*. London: Pan Books Ltd. (Picador edition).
- CURRIE, G. (1990): *The Nature of Fiction*. Cambridge: Cambridge University Press.
- CURRIE, G. (2003): Characters and Contingency. *Dialectica* 57, No. 2, 137-148.
- DAVIDSON, D. (1986): A Nice Derrangement of Epitaphs. In: Lepore, E. (ed.): *Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.
- DAVIES, D. (2007): *Aesthetics and Literature*. London: Continuum.
- DAVIES, D. (2012): Fictionality, Fictive Utterance and the Assertive Content. In: Currie, G. – Kořátko, P. – Pokorný, M. (eds.): *Mimesis: Metaphysics, Cognition, Pragmatics*. London: College Publications.
- DOLEŽEL, L. (2003): *Heterocosmica. Fikce a možné světy*. Praha: Karolinum.
- DRIVER, T. F. (1961): Beckett by the Madeleine. *Columbia University Forum* 4, No. 3, 23.
- ESSLIN, M. (ed.) (1965): *S. Beckett: A Collection of Critical Essays*. Prentice Hall: Englewood Cliffs.
- FOŘT, B. (ed.) (2012): *Heterologica. Poetika, lingvistika a fikční světy*. Praha: Ústav pro českou literaturu AV ČR.
- FRIEND, S. (2008): Imagining Fact and Fiction. In: Stock, K. – Thompson-Jones, K. (eds.): *New Waves in Aesthetics*. New York: Palgrave Macmillan.
- GIBSON, J. (2007): *Fiction and the Weave of Life*. Oxford: Oxford University Press.
- GRICE, P. (1989): Logic and Conversation. In: Grice, P.: *Studies in the Ways of Words*. Cambridge: Harvard University Press.
- KOŘÁTKO, P. (2008): Realismus a literární interpretace. *Svět literatury* 37 (XVIII), 25-39.
- KOŘÁTKO, P. (2010): Who Is Who in the Fictional Worlds. In: Kořátko, P. – Pokorný, M. – Sabatés, M. (eds.): *Fictionality, Possibility, Reality*. Bratislava: aleph, 89-100.
- KOŘÁTKO, P. (2011): Vyprávění jako performance: vypravěč v kondici a vypravěč v rozkladu. *Svět literatury* 43 (XXI), 57-65.
- KRIPKE, S. (1980): *Naming and Necessity*. Oxford: Blackwell.
- LEWIS, D. (1983): Languages and Language. In: Lewis, D.: *Philosophical Papers*, Vol. 1. Oxford: Oxford University Press.
- MARTÍNEZ-BONATI, F. (1981): *Fictive Discourse and the Structures of Literature: A Phenomenological Approach*. Ithaca: Cornell University Press.
- PAVEL, T. (2012): *Fikční světy*. Praha: Academia.
- PRIOR, A. (1971): *The Objects of Thought*. Oxford: Clarendon Press.

(v přepracované verzi) čtenářům *Organonu F* a děkuji oběma anonymním recenzentům za inspirativní polemické poznámky. Přestože na ně neodpovídám v této stati, těším se, že budu mít příležitost udělat to dodatečně v diskusi (pokud nějakou vyvolá).

- PUTNAM, H. (1975): The Meaning of "Meaning". In: Putnam, H: *Mind, Language and Reality*. Cambridge: Cambridge University Press.
- RESCHER, N. – BRANDOM, R. (1980): *The Logic of Inconsistency. A Study in Non-Standard Possible-Worlds Semantics and Ontology*. Oxford: Oxford University Press.
- RYAN, M.-L. (1991): *Possible Worlds, Artificial Intelligence and Narrative Theory*. Bloomington-Indianapolis: Indiana University Press.
- RYAN, M.-L. (2001): *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Digital Media*. Baltimore: Johns Hopkins University Press.
- RYAN, M.-L. (2010): Fiction, Cognition, and Non-Verbal Media. In: Grishakova, M. – Ryan M.-L. (eds.): *Intermediality and Storytelling*. Berlin/New York: De Gruyter, 8-26.
- RYAN, M.-L. (net 1): *From Possible Worlds to Parallel Universes*. <http://www.univ-paris-diderot.fr/clam/seminaires/RyanEN.html>.
- RYAN, M.-L. (net 2): *Possible Worlds*. <http://www.log24.com/log05/saved/050822-PossWorlds.html>.
- SHENKER, I. (1956): A Portrait of Samuel Beckett, the Author of the Puzzling *Waiting for Godot*. *New York Times*, May 6.

Na závěr úvah o intencionálních jsoucnech¹

STANISLAV SOUSEDÍK

Katedra filosofie. Katolická teologická fakulta. Univerzita Karlova v Praze
Thákurova 3. 160 00 Praha 6. Česká republika
sousedik.s@volny.cz

Rád bych v tomto příspěvku shrnul průběh diskuse probíhající v časopisu *Organon F* na námět mnou zastávané teorie intencionálních jsoucen a prozatím ji tím ukončil. Diskuse byla dlouhá, začala v r. 2008 a pokračovala pak s občasnými přestávkami až do posledního, minulého, čísla časopisu (obsahujícího důležitý příspěvek P. Cmoreje 2012). Je mým úmyslem naznačit, v kterých bodech jsem pod dojmem kritických připomínek oponentů své výchozí názory korigoval, a v kterých jsem naopak setrval na svém stanovisku. Rozdělím svůj výklad (v návaznosti na uvedený Cmorejův příspěvek) na několik tematických okruhů. Nejprve se vyjádřím k námětu nejzákladnějšímu:

1. Existují intencionální individua?

Moje přesvědčení o existenci intencionálních individuí vychází z předpokladu, že filosofie musí, nemá-li se proměnit ve fantastickou – byť třeba duchaplnou – spekulaci, vycházet z přirozeného jazyka a respektovat jej. Hned v počáteční fázi naší diskuse jsem to formuloval takto:

Předpokládám, že já i moji oponenti se shodujeme v tom, že při úvahách o problémech logické sémantiky musíme vycházet z přirozeného jazyka. Shodneme se jistě i na tom, že přirozený jazyk není nějaké nedotknutel-

¹ Tato stat' vznikla s podporou Grantové agentury České republiky (projekt P 401/12/1699).

né „sacrosanctum“, nýbrž je k němu třeba přistupovat s kritickou obezřetností. A i to je snad nesporné, že koriguje-li logik v něčem přirozený jazyk, musí k tomu mít i jasné, intersubjektivně dostupné důvody. Přirozený jazyk je tedy innocent until proved guilty. (Sousedík 2010)

Na pozadí tohoto přesvědčení jsem pak vícekrát uvedl, 1) že se o jakýchsi fiktivních individuích (např. v pohádkách) mluví, a že je v nich tedy o nich řeč, a 2) že je běžný uživatel jazyka v souvislosti s tím přesvědčen, že tato individua v rámci příslušného kontextu (např. v pohádkách) existují.

Protože je tomu tak, akceptuji ve shodě s běžnými uživateli jazyka existenci fiktivních individuí. Jejich existenci jsem pak v dalším odlišil od existence individuí reálných, a to na základě skutečnosti, že existenci fiktivních individuí verifikujeme způsobem, jenž je neslučitelný se způsobem, jímž verifikujeme existenci individuí reálných. Vycházel jsem při tom z principu, že jestliže přítomnost libovolné vlastnosti *F* na individuu *x* verifikuji způsobem, jenž je zásadně neslučitelný se způsobem, jímž verifikuji přítomnost libovolné vlastnosti *G* na individuu *y*, nemohou být vlastnosti *F*, *G* identické (podrobněji o tom viz Sousedík 2012, 243). Nazval jsem existenci, kterou se vyznačují fiktivní individua, existencí intencionální a začal jsem odtud mluvit o oněch fiktivních individuích jako o individuích intencionálních.

V mém právě uvedeném myšlenkovém postupu vzbudilo pochybnosti, 1) co rozumím slovem „individuum“, a 2) co výrazem „mluviti o...“ příp. „řeč je o...“. Pokud jde o „individuum“, rozumím jím v této souvislosti entitu, kterou nelze v určitém kontextu přisuzovat jiným, již však lze přisuzovat vlastnosti. Definice je, jak patrně, tak široká, že jí lze subsumovat jak reálná, tak intencionální individua.² Pokud jde o frázi „mluviti o...“, příp. „řeč je o...“ rozumím prvním z těchto výrazů („mluviti o...“) přisuzování nějakého atributu³ entitě, která je označována gramatickým podmětem příslušné věty, druhým („řeč je o...“) tuto entitu samotnou.

Moje výše uvedené vyvození existence intencionálních individuí vychází z přesvědčení, že přirozený jazyk spolu se způsobem, jímž mu běžný uživatel rozumí, nevyžaduje z hlediska logiky v tomto případě nějakou opravu. Pavel Tichý a v jeho šlépějích i jeho vykladač a pokračovatel, prof. P. Materna

² V jiné souvislosti považuji uvedenou definici individua za příliš širokou, protože její pomocí nelze odlišit přirozená individua od agregátů těchto individuí.

³ Slovo „atribut“ užívám ve smyslu užívaném Cmorejem (2002), totiž tak, že se jím označují vlastnosti i vztahy. Podle toho je tedy např. ve větě „Jan je větší než Eva“ řeč o Janovi, Eva je jen „zmiňována“.

(a mnozí další) ovšem soudí, že je tento můj předpoklad mylný. Připouštějí, že sice intencionální individua jsou *gramatickým* podmětem vět, jichž užíváme při vyprávění o intencionálních individuích, soudí však, že gramatický podmět v tomto případě není totožný s podmětem logickým. Logickým podmětem, tj. tím, o čem pomocí takových vět mluvíme, není podle jejich mínění intencionální individuum, např. Karkulka, nýbrž zvláštní druh intenze, kterému říkají „individuová role“, „úřad“ aj.

Tento názor stoupenců TILky znamená ovšem výraznou korekturu přirozeného jazyka. Jako takovou by ji bylo možno akceptovat pouze pod podmínkou, že je dobře, intersubjektivně dostupným způsobem, zdůvodněna. Materna v průběhu diskuse dokazoval, že takto zdůvodněna je, já jsem zastával názor opačný. Není mou věcí rozhodnout, či argumenty byly platné, to musím přenechat úsudku čtenáře. Tolik však snad smím bez arogance říci, že mě osobně Tichého a Maternovy názory v tomto bodu nepřesvědčily.

Odlíšné pochybnosti o způsobu, jímž dospívám k přesvědčení o existenci intencionálních individuí, formuloval P. Cmorej. Cmorej se domnívá, že 1) můj argument pro existenci intencionálních individuí obsahuje kruh (*petitio principii*), a 2) není v něm jasný klíčový termín „mluviti o“. Myslím, že jsem na pochybnosti, že se dopouštím „kruhu“, vlastně již odpověděl rekapitulací svého myšlenkového postupu uvedenou na začátku tohoto příspěvku. Vysvětlil jsem tam i to, co rozumím termínem „mluviti o“, a „individuum“. K přesvědčení o existenci intencionálních individuí nedocházím nějakým argumentem postupujícím od známého k neznámému, nýbrž spíše jen tím, že se ztotožňuji s názorem běžného mluvčího, jenž považuje existenci intencionálních individuí, např. pohádkovou existenci Karkulky, za samozřejmou. Můj argument by tedy bylo možno přehledně formulovat takto: *Nestojí-li tomu v cestě nějaký intersubjektivně prokazatelný důvod, je „intuice“ běžného mluvčího vždy pravdivá. Avšak: proti „intuici“ běžného mluvčího, že existují intencionální individua, nestojí žádný takový důvod.* Tudíž: atd.

Uvažuji tedy jako běžný uživatel jazyka a dělám-li ve srovnání s ním něco navíc, je to jen to, že jeho spontánní přesvědčení artikuluji, a že je hájím proti (podle mého mínění) nikoli dostatečně zdůvodněným námitkám, že je třeba přirozený jazyk, jak jej užíváme k vyprávění např. pohádek, korigovat. Opakuji: Nemyslím, že bych se tím dopouštěl nějakého argumentačního kruhu. Přecházím nyní k dalšímu námětu, v němž jde o problém:

2. Jaká je povaha intencionálních individuí?

Mezi tezemi, které jsem o intencionálních jsoucnech formuloval, je i teze označená římskou II, která zní: *Každé intencionální jsoucnó má pouze ty atributy, které mu přisoudíme a žádné další.*

Proti této tezi namítl Cmorej, že nemůže platit o „mých“ intencionálních individuích, která vedle vlastností, která jim přisuzujeme fiktivní predikací, musí i určité vlastnosti sama o sobě mít (takže jim je musíme připsat predikací „obyčejnou“⁴).

Považuji Cmorejovu kritiku své teze II za průkaznou a koriguji ji v souvislosti s tím v tom smyslu, že intencionální individua fungujeme tak, že mají sama o sobě (bez ohledu na to, zda ji jim připíšeme) tu vlastnost, že *jsou s to být spojována s abstraktními vlastnostmi, a že jsou individuální* (nechávám předběžně otevřenou otázku, jsou-li všeobecně vzato také identická se sebou samými, zda to totiž lze uvést do souladu se skutečností, že připouštím také intencionální individua s kontradiktorními vlastnostmi). Tato moje korektura má dílčí povahu, tj. nemá vliv na to, co jsem v průběhu diskuse o intencionálních individuích z teze II vyvodil ani na teorii intencionálních individuí uvažovanou v celku.

Dalším, posledním problémem, o který v naší diskusi šlo, lze vyjádřit otázkou:

3. Jsou intencionální jsoucna v rámci teorie, kterou zastávám, mentální nebo nementální?

Zastávám názor, že jsou tato jsoucna nementální, Cmorej naopak namítá, že jsou ve skutečnosti mentální.

Při řešení této otázky je třeba si především ujasnit, co v této souvislosti rozumíme termínem „mentální“, což Cmorej, pokud si vzpomínám, v průběhu diskuse výslovně neučinil. Pro svou osobu nazývám *mentální* něco, co nějak náleží k naší (mé, tvójí, kohokoli) mysli, a je tedy v nějakém smyslu její součástí, stránkou, akcidentem apod. Takovou povahu má např. každý

⁴ Rozlišení mezi „obyčejnou“ a fiktivní predikací je Cmorejovým přínosem, jež bez výhrad akceptuji.

kognitivní akt. Kognitivní akt (dále jen „akt“), je mentální, protože je úkonem naší mysli.⁵

Zastávám tezi, že akty (a vůbec všechny mentální entity) existují veskrze reálně. Usuzuji na to z toho, že jejich existence není závislá na tom, jsou-li či nejsou předmětem našeho poznání. Čtu-li např. napínavý detektivní příběh, bývám tak zaujat dějem, že si neuvědomuji akty, jimiž se to uskutečňuje (akty, jimiž rozlišuji písmena textu, akty, jimž chápu, že vytvářejí slova, smysluplné věty atd.). Tyto akty ovšem nepřestávají existovat, třebaže na ně nemyslím. Existují nezávisle na tom, zda jsou či nejsou předmětem mého poznání. Jsou to tedy jsoucna veskrze reálná.

Jinak je tomu však s předmětem těchto aktů, je-li jím intencionální individuum. To sice existuje jen v závislosti na těchto aktech, ale liší se od nich. To, co je co do své existence na něčem závislé, liší se totiž vždy od toho, co mu jeho závislou existenci poskytuje (předpokládat opak vede jasně ke sporu). Předmět aktů se tedy liší od aktů samotných. To zaprvé. A nyní zadruhé: Tento od aktů odlišný předmět existuje intencionálně, tj., jak jsem výše (doufám) dokázal, nějakou existencí odlišnou od existence reálné. Protože akt existuje veskrze reálně, nemůže s ním být intencionální předmět a) ani totožný, b) ani nemůže být nějakou jeho součástí, stránkou apod. Tudíž intencionální předmět není, nemůže být mentální.

Takto bych shrnul svou v průběhu naší diskuse při různých příležitostech naznačenou argumentaci ve prospěch teze, že intencionální předměty (zejména individua, o nichž v naší diskusi především šlo) nejsou nějakými mentálními entitami.

Cmorej uvádí ve svém posledním příspěvku pět pochybností (říkám „pochybností“, nejsou to totiž v přesném smyslu námitky) o tomto mém závěru. Proberu je postupně.

⁵ Zavedlo by nás příliš daleko probírat zde otázku, zda kognitivní akty jsou, a co jsou. Pokusím se to pouze přiblížit. Představme si, že jsme v zatemněné místnosti, kde nevidíme vůbec nic. Rozsvítíme-li světlo, vidíme pojednou, co jsme již dříve sice mohli vidět, ale vlivem temnoty jsme to ve skutečnosti neviděli, dejme tomu stěnu. Mezi naším stavem, kdy jsme v temnotě mohli vidět, ale neviděli, a stavem, kdy skutečně vidíme, je zřejmě jakýsi rozdíl, a ten běžně označujeme slovem „viděti“: viděti tedy znamená zakoušenou skutečnost něčeho, co bylo dříve jen možné. Tato skutečnost může být příjemná nebo nepříjemná, většinou je však neutrální. Má, krom toho, zvláštní vztahovou povahu: zpřítomňuje totiž naší mysli něco od jejího aktu odlišného, totiž viděnou věc.

První vzniká úvahou nad některými nezvyklými důsledky teorie, kterou zastávám, a vrcholí řečnickou otázkou: *Možno azda prostredníctvom introspekcie overiť existenciu nadindividuálneho kontingentného objektu?*

K tomu: Rozumím Cmorejovu termínu „nadindividuální“ takto: *Entita X je nadindividuální právě tehdy, když X může být předmětem poznávacího aktu většího počtu osob.* Je-li to míněno vskutku takto, musím na otázku odpovědět, že vylučně pomocí introspekce toho dosáhnout nelze. Že Hamlet, který je předmětem mého aktu je totožný s tím, který je předmětem Cmorejova aktu lze ověřit jen tak, že zjistím, že Cmorej připisuje „svému“ Hamletovi přesně tytéž vlastnosti, které já připisuji „svému“. Jenže zjistit to, přesahuje již rámec introspekce. Že by nicméně tato skutečnost měla nějak mluvit ve prospěch Cmorejovy teze o mentálnosti intencionálních individuí, nevidím.

Druhá Cmorejova pochybnost spočívá v upozornění na některé další, podle něho, obtížně přijatelné důsledky teorie, kterou zastávám. Jedna obtíž je spojena s tím, že připustíme-li existenci intencionálních objektů, nemáme nikdy jistotu, které a kolik prvků tvoří univerzum naší úvahy. K tomu: To, co Cmorej uvádí, je ovšem z jistého hlediska pravda, z jiného zase ne, ale ať tak či onak, není v tom, myslím, nic, co by nás muselo znepokojovat. Domnívám se totiž, že všeobecně vzato není taxativní výpočet individuí jediný (ba ani obvyklý) způsob, jímž univerzum zadáváme: většinou se spokojujeme pouze udáním kritéria, pomocí kterého jsme o každém prvku s to *princiálně* rozhodnout, zda do univerza náleží či nikoli. Tak je tomu i v případě intencionálních individuí. Principiálně můžeme s jistotou zjistit, kdo do univerza obsahujícího taková individua náleží. Dejme tomu, že je naším univerzem soubor osob vystupujících v dramatu Hamlet. Abych mohl udat, kolik z těchto osob nyní intencionálně existuje, musel bych např. zjistit, že Hamlet nyní intencionálně existuje, protože na něj právě myslí slovatný pražský anglista, prof. H., Ofélie rovněž existuje, neboť se její postavou právě v duchu zabývá její představitelka, členka ochotnického souboru v N., slečna R., chudák Polonius nyní neexistuje, alespoň v tom smyslu ne, že na něj právě nyní, kromě mne, nikdo nemyslí, atd. Toto vše lze principiálně zjistit, protože máme kritérium, umožňující nám rozhodnout, že např. Faust, na kterého právě myslí pan W., do hledaného univerza nenáleží a Ifigenie, kterou momentálně všichni pomíjejí, také ne. Divné by při tom ovšem bylo, kdyby se takovým zjišťováním někdo zabýval, protože by to byla velmi obtížná činnost a nadto, myslím, že i nepřilíš užitečná. Protože asi nebude nikoho, kdo by se do takového podniku chtěl pouštět, mů-

žeme připustit spolu s Cmorejem, že není jisté, které a kolik prvků do našeho univerza náleží.

Dále v rámci této druhé pochybnosti Cmorej správně upozorňuje, že jsou pro mne intencionálními předměty nejen individua, ale i vlastnosti, vztahy, matematické entity atd., a vyslovuje přesvědčení, že to vede k nepřekonatelným obtížím. Objasňuje pak (pod čarou) pouze jednu z těchto obtíží, a to otázkou: „Je konkrétně jablko sladké iba vtedy, keď niekto myslí na jeho sladkosť?“ (Cmorej 2012, 518). Na otázku si máme zřejmě dát zápornou odpověď. Smysl námitky není zcela jednoznačný, ale vykládám si ji takto: Je-li vlastnost „sladkosť“ intencionální objekt, pak, není-li v okamžiku t nikoho, kdo by měl tuto vlastnost na mysli (takže abstraktní vlastnost sladkosti v okamžiku t neexistuje), je obtížné pochopit, jak by mohlo být v okamžiku t nějaké konkrétní jablko sladké (jak to běžný mluvčí předpokládá). Odpovídám: I když abstraktní vlastnost (= pojem) sladkosti neexistuje a nikdo mu ji tudíž nemůže *připisovat*, má konkrétní jablko bez ohledu na to konkrétní vlastnost sladkosti. My, kteří pojem sladkosti máme, můžeme tedy např. říci, že jablko na stromě v ráji bylo již před stvořením Adama sladké, ovšem že to tehdy (na rozdíl od dneška) žádný člověk pomocí pojmu sladkosti nemohl konstatovat.

Třetí Cmorejova pochybnost spočívá v tom, že se výroky o intencionálních individuích vyznačují ve srovnání s výroky o reálných individuích některými zvláštnostmi v oblasti verifikace: Výrok „Intencionální objekt ‚O‘ existuje“ nelze totiž podle jeho mínění falzifikovat, protože každý pokus o jeho ověření je spjatý s porozuměním tomuto výroku, při němž znovu chápeme mentální korelát objektu O. Toto Cmorejovo zjištění je správné, ale neúplné. Zaujal jsem k této problematice již dříve dosti podrobně stanovisko (srov. Sousedík 2012, 91–92), jež však pravděpodobně uniklo pozornosti. Stručně: musíme rozlišovat nejméně dvojí existenci: 1) tu, kterou objekt O díky našemu fingování má, 2) tu, kterou mu fiktivně připisujeme. Polonius, jenž je (neklame-li mne paměť) ve 3. dějství Hamleta zavražděn, existuje sice nadále intencionální existencí, kterou, pokud na něj někdo myslí, má, ale tu, kterou jsme mu na počátku hry fiktivně připsali, v důsledku své smrti ztrácí: platí tedy, že Polonius ve 4. a 5. dějství již neexistuje_{ft}, i když na něj třeba myslíme. Nevidím v tom všem nic, co bych měl považovat za průkaznou kritiku teorie, kterou zastávám.

Čtvrtá pochybnost je podobného rázu. Dotvrzuje pouze Cmorejovu kritiku mé teze II, k níž jsem se již, a to pozitivně, vyjádřil v předchozím.

Pátá Cmorejova pochybnost obsahuje více zajímavých postřehů, mám nicméně za to, že jsem na ně poskytl odpověď v rekapitulaci své argumentace ve prospěch existence intencionálních individuí, jež je uvedena na počátku prvního bodu této mé repliky.

* * *

Na základě dohody, k níž došlo mezi redakcí a účastníky, se tímto mým příspěvkem dosavad probíhající diskuse o intencionálních jsoucnech prozatím končí. Výzva, abych ji ukončil právě já, a měl v ní tím pádem předběžně i poslední slovo, mně uvedla do nemalých rozpaků, protože vím, že by prof. Cmorej či nějaký jiný účastník našeho dialogu měl k tomuto mému „poslednímu“ slovu jistě ještě co říci. Ustoupil jsem posléze, nejsem si jist, zda oprávněně, argumentu, že se v diskusi jednalo o mnou nadhozený námět. Ať jakkoli, vyslovil bych na závěr rád svůj dík redakci *Organonu F* za (mimořádně benevolentní) poskytnutí prostoru k výměně názorů a váženým pánům kolegům, účastníkům diskuse, za početné podněty, jimiž se cítím nemálo obohacen: A. Dolákovi, že diskusi rozvířil, P. Maternovi a J. Ráclavskému, že v ní prezentovali hledisko transparentní intenzionální logiky, P. Cmorejovi, že mě pozorně sledoval na neprošlapaných cestách,⁶ po nichž se ubírám, a vytrvale mě při tom upozorňoval na úskalí a scestí, jimž jsem se při svém putování nedokázal vždy vyhnout.

⁶ Vývoj nauky o intencionálních jsoucnech je dosud málo probádanou oblastí. Textový základ u Aristotela je poměrně malý, rozvinuta byla teprve ve středověku, nejprve pravděpodobně arabskou, později křesťanskou scholastikou. K jejímu velikému rozvoji došlo však, jak se zdá, teprve u scholastiků 17. stol. Rozvětvená nauka o tomto námětu, s níž se tu setkáváme, zasahuje do mnoha oblastí a je často velmi subtilní, překvapuje však, že staří autoři pomíjeli, pokud dosud známo, námět intencionálních individuí. Pokusil jsem se doplnit starší (především tomisticky zaměřenou) nauku právě o toto téma. V tom smyslu je tedy třeba chápat mé vyjádření, že jsem putoval po „nevyšlapaných cestách“. V našem prostředí se dějinami nauky o intencionálních jsoucnech (převážně v 17. stol.) zabývá pracovník theologické fakulty Jihočeské univerzity, Daniel Novotný. Jeho monografie o tomto námětu má být zveřejněna v dohledné době ve Spojených státech.

Literatura

- CMOREJ, P. (2002): *Úvod do logické syntaxe a sémantiky*. Přel. T. Machula. Praha: Triton.
- CMOREJ, P. (2012): Existencia a mentálnosť intencionálnych individuí a predikácia ich vlastností. *Organon F* 19, č. 4, 506-522.
- SOUSEDÍK, S. (2010): Kritická poznámka k sémantickým východiskům transparentní intenzionální logiky. *Organon F* 17, č. 3, 358-365.
- SOUSEDÍK, S. (2012): Platónské a aristotelské pojetí intencionálních jsoucn. *Organon F* 19, č. 1, 84-95.

K logickej analýze lingvistického pojmu synonymie

MAREK MIKUŠIAK

Katedra slovenského jazyka. Filozofická fakulta Univerzity Komenského v Bratislave
Gondova 2. 814 99 Bratislava. Slovenská republika
marek.mikusiak@gmail.com

V tomto článku sa budem zaoberať predovšetkým knihou Barbory Geistovej Čakovskej *Vztáhy synonymie: logická sémantika verzus lingvistické teórie o synonymii*. Na úvod uvedenú monografiu stručne predstavím a následne budem s niektorými, dá sa povedať, ťažiskovými riešeniami teórie polemizovať.

Knihou B. Geistovej Čakovskej nepochybne predstavuje zaujímavý pokus o transdisciplinárne preskúmanie javu synonymie. V prvej časti monografie skúma autorka vzťah synonymie z pohľadu logickej sémantiky, explikuje pojem kognitívneho významu, definuje tzv. LOGsynonymiu a dospieva k záveru, že LOGsynonymia „splňa všetky požiadavky ekvivalenčného vzťahu“ (Geistová Čakovská 2010, 50). Uvedené konštatovanie pokladá autorka za dôležité v súvislosti s tým, ako chápu synonymiu lingvisti, pretože pojem synonymie, s ktorým pracuje jazykoveda, požiadavky ekvivalenčného vzťahu nespĺňa. Samozrejme, lingvisti si obyčajne nekladú otázku, či synonymia je, alebo nie je ekvivalenčný vzťah. Prvá časť monografie venovaná LOGsynonymii preto nielen objasňuje vzťah synonymie z pohľadu logickej sémantiky, ale zároveň ponúka vypracovaný konceptuálny aparát, pomocou ktorého možno tzv. LINSynonymiu skúmať a rozlišovať *rovnosť, blízkosť, rôznosť, prípadne hyponymiu a hyperonymiu*. Súčasťou tohto aparátu je Cmorejova koncepcia semivýrazov, Tichého koncepcia náležitostí a logické pojmy ako reflexívnosť, symetrickosť, tranzitívnosť, L-ekvivalencia, koreferenčnosť a pod. V druhej, rozsiahlejšej časti knihy sa tento konceptuálny aparát uplatňuje pri skúmaní vzťahu LINSynonymie.

Zvolený metodologický rámec pomáha autorke preklenúť problémy spojené s faktom, že „lingvistika neposkytuje uspokojivé koncepcie pre explikáciu kognitívneho významu slov“ (Geistová Čakovská 2010, 13). Kognitívny význam autorka pokladá za myšlienkový obsah výrazu (Geistová Čakovská 2010, 21), ktorý umožňuje komunikáciu a ktorý možno ďalej explikovať ako intenziu, čiže ako funkciu (Geistová Čakovská 2010, 23-24). Pokiaľ dve slová vyjadrujú ten istý kognitívny význam, t. j. pokiaľ signifikujú tú istú intenziu, môžeme ich podľa autorky pokladať za LOGsynonymné. Definícia LOGsynonymie teda vychádza zo sémantiky intenzionálnej logiky. Napriek tomu, že sa niektoré logické nástroje uplatňujú aj pri skúmaní LINsynonymie, analýza kognitívneho významu ako intenzie nie je v rámci skúmania LINsynonymie uspokojivá, pretože neumožňuje skúmať čiastočnú podobnosť významov, ktorá je pre mnohých lingvistov postačujúcim kritériom synonymie. Teoretickým nástrojom takzvaného rozštiepenia kognitívneho významu má byť Tichého koncepcia náležitosti, pomocou ktorej sa autorka pokúša analyzovať kognitívny význam slov jemnejšie, než to umožňuje analýza významu ako intenzie. Okrem faktu, že vzťah LINsynonymie nepredpokladá nevyhnutnú totožnosť kognitívnych významov, autorka zdôrazňuje aj ďalšie dôležité odlišnosti tohto vzťahu od LOGsynonymie. Lingvisti predovšetkým problematiku neredukujú na otázku totožnosti kognitívnych významov, ale zohľadňujú aj nekognitívne významy, ktoré autorka chápe ako štylistické príznaky príslušného slova. Nemenej dôležitým postrehom je fakt, že v lingvistike sa synonymia obyčajne nepokladá za reflexívny vzťah (lingvista obyčajne neuvažuje o tom, že slovo „dom“ je synonymné so slovom „dom“).

Ako som už spomenul, autorka v rámci LINsynonymie rozlišuje rovnoznačnosť a blízkoznačnosť, ktoré sú v opozícii k nesynonymnému vzťahu rôznoznačnosti. Rovnoznačnosť predstavuje to, čo niektorí lingvisti pokladajú za tzv. absolútnu synonymiu; dva rôzne výrazy sú rovnoznačné, keď majú totožný kognitívny aj nekognitívny význam. Pojem blízkoznačnosti je o čosi zložitejší, pretože sa takpovediac rozpadá na tri typy (A, B a C) podľa toho, či majú blízkoznačné slová

A. blízky kognitívny význam a totožný nekognitívny význam

alebo majú

B. blízky kognitívny význam a odlišný nekognitívny význam

alebo majú

- C. kognitívny význam totožný a líšia sa v nekognitívnom význame (pozri Geistová Čakovská 2010, 102).

Autorka v monografii ponúka aj jednotlivé príklady slov, na ktorých sa pokúša demonštrovať uvedené vzťahy. Treba zdôrazniť, že adekvátne metóda skúmania vzťahov synonymie má spočívať predovšetkým na testovaní vzájomnej nahraditeľnosti synonym vo vetnom kontexte. Práve na základe tohto testu je potom podľa autorky možné odhaľovať konkrétny typ vzťahu medzi príslušnými slovami. Ak teda chceme potvrdiť napr. rovnoznačnosť, tak po nahradení slova za jeho rovnoznačný korelát nesmie dôjsť k zmene kognitívneho významu, ale ani k zmene štylistického príznaku vetného kontextu.

V tejto súvislosti si však kladiem otázku: Čo navrhovanou procedúrou vlastne testujeme? Ide skutočne o testovanie, ktoré môže priniesť netriviálne výsledky? Tu treba pripomenúť, že vzhľadom na koncepciu semivýrazov a jazykových výrazov sa každá zmena v kognitívnom význame jazykového výrazu prejaví ako zmena výrazu, pretože identita jazykového výrazu je podľa prijímanej teórie daná spojením semivýrazu s určitým kognitívnym významom ($S = \langle s, k \rangle$). Pokiaľ však máme skúmať vzťah medzi dvoma jazykovými výrazmi, zrejme by sme mali vedieť, o ktoré výrazy ide, a to znamená, že by sme mali poznať aj kognitívny význam daných slov. V takom prípade sa potom stačí opýtať, či sú dané kognitívne významy totožné, a nepotrebujeme nič testovať.

Pravda, pri rovnoznačnosti musíme brať na zreteľ aj štylistické príznaky, ktoré sa na identifikácii výrazu nepodieľajú. V takomto prípade má, zdá sa, test opodstatnenie: Podmienkou rovnoznačnosti je to, aby „po vzájomnom nahradení dvoch rovnoznačných slov nedošlo k zmene na úrovni príslušnosti jazykového kontextu do istej štylistickej vrstvy jazyka“ (Geistová Čakovská 2010, 83-84). Treba však povedať, že testovanie rovnoznačnosti vo vetnom kontexte má zmysel a opodstatnenie iba vtedy, ak pripustíme určitú formu kontextualizmu, podľa ktorého (nekognitívny) význam slov, a teda aj uvažované synonymné páry určujeme až v kontexte. Pokiaľ by sme totiž štylistický príznak určitého vetného kontextu určovali na základe štylistických príznakov jednotlivých slov, tak test synonymie nepotrebujeme, pretože sama možnosť testovať by z hľadiska teórie predpokladala poznanie toho, čo má byť informatívnym výsledkom testovania. Inými slovami, ak nepripustíme, že – ako sa nazdáva Mistrík – okrem slov s konštantnou štylistickou hodno-

tou existujú aj slová, ktoré môžu nadobúdať tzv. kontextovú štylistickú hodnotu (Mistrík 1997, 401-402), tak kontextový test synonymie nepotrebujeme. Na druhej strane sa však zdá, že kontextualizmus z hľadiska koncepcie celej knihy nie je prípustný. Pod koncepciou knihy mienim predovšetkým teoreticko-metodologický základ umožňujúci tzv. logickú analýzu, ktorá má byť „v mnohom nápomocná pri teoretickom cizelovaní problematiky blízkoznačnosti a LINsynonymie vôbec“ (Geistová Čakovská 2010, 118).

Už som spomínal, že analýza vzťahu blízkoznačnosti vedie autorku k tomu, aby nejakým spôsobom rozčlenila intenzie blízkoznačných slov na časť, ktorá predstavuje spoločný kognitívny význam (sk), a na časti, ktoré sú síce súčasťou kognitívneho významu oboch slov, avšak sú navzájom odlišné (ok_1/ok_2). Teoretickým nástrojom umožňujúcim formálne zachytenie uvedenej predstavy je Tichého koncepcia náležitostí. Autorka zdôrazňuje, že náležitosti vyplývajú z kognitívnych významov slov a v prípade blízkoznačných slov (typu A a B) musia z čiastočne odlišných kognitívnych významov vyplývať niektoré spoločné a niektoré odlišné náležitosti. Problém je však v tom, že aj v prípade dvojice slov, ktoré nepokladáme za synonymné, môžeme odvodiť spoločnú náležitosť. Teda: Koľko spoločných náležitostí odvodených z rôznych slov musíme odhaliť a akú musia mať povahu, aby sme mohli oprávnene konštatovať, že ide o blízkoznačné a nie rôznoznačné slová? Podľa autorky rozdiel medzi blízkoznačnými (typ A a B) a rôznoznačnými slovami spočíva v tom, že v prípade blízkoznačných slov z ich kognitívnych významov vyplývajú aj také náležitosti, ktoré sú dostatočne konkrétne na to, aby zabezpečili nevyhnutný neprázdny prienik rozsahov daných slov (Geistová Čakovská 2010, 133). V tomto prípade sa autorka odvoláva na istú úroveň konkrétnosti, avšak bližšie ju nešpecifikuje. Namiesto toho predkladá príklad dvojice slov *matematik* a *hudobník* s nasledujúcim vysvetlením: Napriek tomu, že môžu mať tieto slová spoločný prienik rozsahov referencie (je možné, aby bolo jedno individuum aj matematikom aj hudobníkom), nie sú blízkoznačné, pretože ak nejakému individuu „pripisujeme vlastnosť ‚(byť) matematik‘, ide kognitívne o niečo úplne iné, ako keď mu pripisujeme vlastnosť ‚(byť) hudobník““ (Geistová Čakovská 2010, 135). Uvedený argument podľa všetkého vychádza z určitej jazykovej intuície. Nepokladám podobné argumentačné stratégie za márne, otázka však znie, ako nám problém blízkoznačnosti a všeobecne LINsynonymie pomáha riešiť koncepcia náležitostí? Zdá sa totiž, že rozlišovanie medzi blízkoznačnosťou

t'ou a rôznoznačnosťou je v konečnom dôsledku založené na jazykovej kompetencii a intuíciiach.¹

Ako som naznačil, autorka okrem rovnoznačnosti, blízkoznačnosti a rôznoznačnosti uvažuje aj o vzťahoch hyponymie a hyperonymie. Píše:

Hyponymia a hyperonymia sú asymetrické vzťahy medzi slovami, ktoré sa prejavujú aj tým, že rozsah hyponymného slova je úplne obsiahnutý v rozsahu hyperonymného slova. Príkladom môže byť dvojica slov *človek* a *žena*, pričom slovo *človek* je hyperonymom slova *žena* a slovo *žena* je hyponymom slova *človek*. (Geistová Čakovská 2010, 139)

Podľa autorky je dôležité odlišovať vzťahy hyponymie a hyperonymie od vzťahu blízkoznačnosti, „pretože slová v týchto vzťahoch nespĺňajú stanovené podmienky o vzájomnej čiastočnej odlišnosti kognitívnych významov“ (Geistová Čakovská 2010, 140). Nie som si istý, či uvedenej poznámke správne rozumiem, pretože kognitívne významy slov *človek* a *žena* sa podľa môjho názoru vzájomne (aspoň čiastočne) odlišujú. V každom prípade autorka ponúka aj iné rozlišovacie kritérium – tvrdí, že rozsahy slov, ktoré sú vo vzťahu hyponymum-hyperonymum „majú nevyhnutne spoločné prvky, ako to bolo v prípade blízkoznačných slov, nemajú však navzájom odlišné prvky, čím sa od blízkoznačných slov líšia“ (Geistová Čakovská 2010, 141). To, že rozsahy slov *človek* a *žena* nemajú navzájom odlišné prvky, znamená, že hoci do rozsahu slova *človek* patria aj prvky, ktoré sú odlišné od prvkov rozsahu slova *žena*, naopak to neplatí, všetky prvky rozsahu slova *žena* totiž patria nevyhnutne aj do rozsahu slova *človek*. Uvedené kritérium je však podľa môjho názoru taktiež problematické, nepomáha nám totiž rozlíšiť hypero-hyponymický vzťah od blízkoznačnosti typu C, keďže ani rozsahom C-blízkoznačných slov nenáležia navzájom odlišné prvky.

Ťažkosti s rozlišovacími kritériami, ktoré som naznačil, nemožno zanedbať, pretože autorka hypero-hyponymické vzťahy od LINsynonymie prísne odlišuje (Geistová Čakovská 2010, 139). Nenašiel som explicitný argument v prospech potreby takéhoto „prísneho rozlišovania“, avšak v tejto súvislosti je zaujímavý predovšetkým prekvapujúci záverečný obrat autorky, ktorá napokon pripúšťa, že „môžeme... pokladať hyponymiu a hyperonymiu v súlade s niektorými lingvistami tiež za LINsynonymné vzťahy a podporovať prítomnosť takých slov v synonymických slovníkoch“ (Geistová Čakov-

¹ J. Dolník v tejto súvislosti píše o tzv. synonymickej kompetencii. Pozri Dolník (2003, 96).

ská 2010, 143). Jednak sa teda autorka vzdáva „prísneho rozlišovania“ a jednak, zdá sa, pripúšťa možnosť, že aj dvojice slov *človek* a *žena*, prípadne *živočích* a *muž*, možno pokladať za synonymá.

Musím povedať, že nemám celkom jasno v tom, aký *teoretický problém* stojí v pozadí diskutovanej štúdie. Autorka uvádza, že sa usiluje o „riešenie synonymie“ (Geistová Čakovská 2010, 100), bližšie však nešpecifikuje, v čom má prípadné riešenie spočívať. Okrem toho sa zdá, že voľba teoretického aparátu ako nástroja analýzy LINsynonymie nebola príliš šťastná. Mnohokrát totiž logická analýza takpovediac „beží naprázdno“. Uveďme príklad:

Nekognitívne významy slov *sopka* a *vulkán* teraz necháme bokom a budeme sa venovať analýze ich kognitívnych významov. Vieme, že ich vzájomne totožný kognitívny význam môžeme v intenzionálnej logike explikovať ako intenziu vyjadrenú napr. ako Sop/Vulk. Na základe čoho však tvrdíme, že ide o jednu a tú istú intenziu? Ako by sme lepšie ukázali napríklad cudzincovi, ktorý sa učí slovenčinu, že tieto dva výrazy majú skutočne totožné kognitívne významy? Uvádzaním spomínaných intenzií by sme didakticky veľmi nezapôsobili, cudzinec by musel jednoducho veriť nášmu tvrdeniu.

Jedným z možných spôsobov by bolo poukazovanie na nevyhnutnosť totožnosti rozsahov daných výrazov. Pre všetky svetamihy a všetky indivíduá platí, že ak tieto indivíduá patria v nejakom svetamihu do množiny indivíduí s vlastnosťou ,(byť) *vulkán* teda do rozsahu predikátu *vulkán*, patria aj do množiny indivíduí s vlastnosťou ,(byť) *sopka* teda do rozsahu predikátu *sopka*, a naopak. V nijakom svetamihu neexistuje také indivídium, ktoré by bolo *vulkánom*, a zároveň by nebolo *sopkou*, a naopak. Nevyhnutná totožnosť rozsahov skúmaných výrazov nás vedie k tomu, že tieto výrazy signifikujú tú istú intenziu. (Geistová Čakovská 2010, 93)

Nepredpokladám, že by cudzinca táto odpoveď potešila, preto uvediem autorkine alternatívne vysvetlenie:

Ďalším zo spôsobov vyjadrenia totožnosti daných kognitívnych významov je použitie koncepcie náležitostí, ktorú sme si priblížili v kapitole II.5. ... Náležitosti, ktoré by sme uviedli v prípade vlastnosti Sop, by sme uviedli aj pre vlastnosť Vulk, a naopak. Preto je Sop a Vulk jedna a tá istá vlastnosť, intenzia. Pre intenziu Sop neexistuje z nej vyplývajúca vlastnosť, ktorá by nevyplývala aj z intenzie Vulk, a naopak. Sop a Vulk

majú tie isté náležitosti, čo je dané totožným kognitívnym významom príslušných slov. (Geistová Čakovská 2010, 93-94)

Z posledného tvrdenia musí byť jasné, že východiskom predchádzajúcich úvah je totožnosť kognitívnych významov rovnoznačných slov *sopka* a *vulkán*. Predchádzajúce úvahy však, ak tomu dobre rozumiem, boli motivované otázkou: „Ako by sme lepšie ukázali napríklad cudzincovi, ktorý sa učí slovenčinu, že tieto dva výrazy majú skutočne totožné kognitívne významy?“ (Geistová Čakovská 2010, 93)

Okrem toho, že nástroj analýz pracuje na niektorých miestach knihy takpovediac naprázdno, treba zdôrazniť aj fakt, že prijatie koncepcie semivýrazov v istom zmysle „podmýva“ adekvátne skúmanie synonymie – samozrejme, pokiaľ má byť tento vzťah analyzovaný skutočne z lingvistického hľadiska. Dôvod je jednoduchý: Prijatá koncepcia semivýrazov, resp. prijatá koncepcia slova nepripúšťa existenciu viacvýznamových slov. To nevyhnutne vedie k opomínaniu známych lingvistických pojmov homonymie a polysémie, ktoré sú v jazykovedných teóriách obyčajne vymedzené aj vo vzťahu k synonymii. Prv než naznačím, prečo je táto poznámka dôležitá, uveďme definíciu LINsynonymie a porovnajme ju so štandardným lingvistickým vymedzením synonymím:

Slovo S_1 z jazyka L je LINsynonymné so slovom S_2 z jazyka L (pričom nemôže nastať prípad, že $S_1 = S_2$) vtt S_1 z L má taký istý alebo blízky, príbuzný význam ako S_2 z L . (Geistová Čakovská 2010, 55)

Naproti tomu v lingvistike sa za synonymá štandardne považujú:

a) formálne odlišné lexikálne jednotky s úplnou zhodou významu (absolútne synonymá) ako i b) formálne odlišné lexikálne jednotky, ktoré nie sú významovo identické, ale len podobné (čiastočné synonymá). (Orgoňová – Bohunická 2011, 58)

Podstatný rozdiel medzi uvedenými vymedzeniami spočíva v tom, že v prvom prípade je formulovaná podmienka odlišnosti slov, kým v druhom prípade je formulovaná podmienka *formálnej* odlišnosti slov, t. j. podmienka odlišnosti semislov. Podľa autorkinej koncepcie napr. v prípade slova *piť* vo význame 1. prijímať tekutinu (*piť* vodu, mlieko) a vo význame 2. cicať (dojča hltavo pilo) ide v konečnom dôsledku o dve rôzne, hoci významovo podobné slová *piť*₁ a *piť*₂. To znamená, že tieto slová sú vo vzťahu LINsynonymie. Z lingvistického hľadiska by však mala byť v tomto prípade reč

skôr o polysémii. Okrem toho môžeme uvažovať taktiež o slove *pit*₃ vo význame 3. piť alkohol (pije ako dúha), ktoré je hyponymom vo vzťahu k slovu *pit*₁. Pokiaľ na základe našich jazykových intuícií odhalíme „kognitívnu príbuznosť“ daných slov, budú taktiež LINsynonymné. V rámci teórie je uvedená možnosť prípustná, lingvisti však podobné prípady klasifikujú ako polysémiu. (Porovnaj tiež možnú analýzu vzťahov medzi výrazmi *nemčinár*₁ vo význame 1. učiteľ nemčiny a *nemčinár*₂ vo význame 2. študent nemčiny; alebo medzi výrazmi *vojak*₁ 1. príslušník armády a *vojak*₂ 2. príslušník armády bez hodnoti.)

Na záver možno konštatovať, že B. Geistová Čakovská spracovala zaujímavú a nesmierne komplexnú problematiku synonymie spôsobom, ktorý môže byť v mnohom podnetný prinajmenšom ako pokus rozlíšiť rozličné stupne či typy synonymných vzťahov. Monografia *Vzťahy synonymie: logická sémantika verzus lingvistické teórie o synonymii* však zároveň podnecuje k otázke, či je teoreticko-metodologický základ skúmania adekvátny vo vzťahu k tomu, čo má byť predovšetkým skúmané, t. j. vo vzťahu k lingvistickej synonymii. Je zrejmé, že v diskutovanej monografii sa uplatňuje takzvaný reduktívny prístup, avšak metodologicky opodstatnené zúženie problematiky, zdá sa, neprináša očakávané výsledky. Možno sa teda pýtať na dôvod nasledujúcej redukcie:

Zdôrazňujeme, že kognitívny význam je primárnym významom v celej oblasti rôznych vzťahov synonymie, pretože dva výrazy nemôžu byť synonymné (LINsynonymné ani LOGsynonymné) iba na základe nekognitívneho významu! (Geistová Čakovská 2010, 74)

Možná odpoveď na moju otázku znie: Uvedené tvrdenie vychádza z definície LOG a LINsynonymie. Problém však spočíva v tom, že jednak sa v definícii LINsynonymie nehovorí o *kognitívnom* význame, a jednak – a to predovšetkým – táto definícia predstavuje akýsi *návrh* ako uchopiť (lingvistickú) synonymiu. Tvrdím, že tento návrh je reduktívny (ako sú reduktívne napokon aj mnohé lexikologické teórie), pretože neposkytuje teoretický priestor na uchopenie synonymných dvojíc slov, ktoré nemajú kognitívny význam (resp. klasickú sémickú štruktúru). Mám na mysli dvojice slov ako napr. žiaľ/bohu/bohužiaľ (časť.), napokon/nakoniec (prísl.), hoci/aj keď (spoj.) a pod. Mimochodom, informáciu o tom, že do synonymických radov vstupujú aj častice, príslovky či spojky, nájdeme na 71. strane diskutovanej knihy.

Literatúra

- DOLNÍK, J. (2003): *Lexikológia*. Bratislava: Univerzita Komenského.
- GEISTOVÁ ČAKOVSKÁ, B. (2010): *Vztáhy synonymie. Logická sémantika verzus lingvistické teórie o synonymii*. Bratislava: aleph.
- MISTRÍK, J. (1997): *Štylistika*. 3. vyd. Bratislava: Slovenské pedagogické nakladateľstvo.
- ORGOŇOVÁ, O. – BOHUNICKÁ, A. (2011): *Lexikológia slovenčiny*. Bratislava: Stimul.

Marek Picha: *Kdyby chyby. Epistemologie myšlenkových experimentů*
Olomouc: Nakladatelství Olomouc 2011, 195 strán

Známe olomoucké nakladatelství vydalo v roku 2011 dve monografie analyticky orientovaných filozofov. Prvou z nich je súbor štúdií Jiřího Raclavského *Individua a jejich vlastnosti*, zameraný na metafyziku, druhou kniha Mareka Pichu *Kdyby chyby*, venovaná – ako je zrejmé z podtitulu – prevažne epistemológii. S prvou publikáciou sa čitatelia *Organonu F* iste budú môcť bližšie zoznámiť v niektorom z najbližších čísel; tú druhú podrobnejšie predstavím na nasledujúcich riadkoch.

Problematika myšlienkových experimentov a ich uplatnenia vo vede, resp. vo filozofii, sa najmä v posledných dvoch desaťročiach stala predmetom živého záujmu epistemológov i filozofov vedy. No hoci v bývalom Československu už pred vyše štyridsiatimi rokmi vznikla práca Václava Černíka (1972), na ktorú Picha upozorňuje v predslove, od tých čias bolo u nás príspevkov k tejto téme porskromne. Svojimi zámermi – a tiež, ako sa pokúsim ukázať, ich realizáciou – však *Kdyby chyby* túto medzeru zaplňajú a zároveň domácemu publiku kriticky predstavujú výsledky zahraničných diskusií.

Prvoradou ambíciou Pichovej práce je odpovedať na otázku, či myšlienkové experimenty (ME) majú epistemický význam – jednoducho povedané, či ich treba brať vážne, alebo nie. Pochopiteľne, odpoveď bude závisieť od mnohého, no predovšetkým od toho, čo všetko sa rozhodneme považovať za ME, akú silnú schopnosť zdôvodňovať presvedčenia im pripíšeme, a na čom bude založená. Týmto základným problémom sú venované prvé štyri kapitoly, v ktorých sa postupne rysuje kladná odpoveď na uvedenú otázku – hoci, ako uvidíme, s viacerými „ale“. Piata, záverečná kapitola potom načrtáva rámec na posudzovanie argumentácie myšlienkovými experimentmi.

V prvej kapitole sa Picha venuje problému identity ME, teda otázke „Čo sú zač?“. Ako ukazuje, koncepcie, ktoré jednotlivý ME stotožňujú takpovediac s celou jeho formuláciou, takže za jeho súčasť považujú nielen konkrétny „scenár“, ale aj všetky dôsledky jeho premyslenia, narážajú na vážne námietky: Ako je napríklad možné, že ten istý ME môže viesť (napríklad u rozličných experimentátorov) k rôznym výsledkom? Alebo: Ak zmeníme ciele, ktoré argumentáciou pomocou ME sledujeme, ide o nový experiment, hoci podstatný obsah scenára sa nezmenil? Zoči-voči týmto problémom navrhuje Picha rozlišovať

medzi „širokým“ a „úzkym“ chápaním ME, pričom podľa druhého je identita experimentu určená výlučne konkrétnym scenárom, bez ohľadu na všeobecné dôsledky, ktoré z neho majú vyplývať, či širšiu argumentáciu, v ktorej scenár figuruje. Úzke chápanie, z ktorého Picha ďalej vychádza, sa uvedeným ťažkostiam vyhýba a umožňuje, aby takto chápaný jednotlivý ME vystupoval v rôznych argumentoch s rôznymi cieľmi. Tým sa zároveň darí zabrániť „inflácii“ myšlienkových experimentov, ku ktorej vedie široké chápanie.

Úzke chápanie navyše dovoľuje úspornú a transparentnú analýzu ME pomocou veľmi jednoduchých nástrojov, ktorá má tú intuitívnu výhodu, že ich pripodobňuje reálnym experimentom. Základom každého ME je podľa Pichu predstavovanie si situácie, pri ktorej vychádzame z určitého známeho (resp. scenárom ozrejmeneho) vzťahu „vstupnej konštelácie“ k „výstupnej konštelácii“ a (v duchu) sledujeme, k akej novej výstupnej konštelácii povedie zmena pôvodnej vstupnej konštelácie. Každý ME možno teda analyzovať do nasledujúcej podoby:

$$\begin{array}{ccc} i_1 & \longrightarrow & o_1 \\ \downarrow & & \\ i_2 & \longrightarrow & o_2 \end{array}$$

Pri každom ME sa najprv dozvedáme, že určitá vstupná konštelácia (i_1) vedie (\rightarrow) k výstupnej konštelácii (o_1); podstatou myšlienkového experimentu je potom skúmanie, k akej výstupnej konštelácii (o_2) povedie modifikácia (\downarrow) pôvodnej vstupnej konštelácie (i_2). Takúto rekonštrukciu ME nazýva Picha „destilátom“. Napríklad pre známy Descartov myšlienkový experiment z *Meditácií o prvej filozofii* predkladá Picha nasledujúci destilát (s. 174):

$$\begin{array}{ccc} (i_1) \text{ menej mocný klamúci démon} & \longrightarrow & (o_1) \text{ nemôže klamať o niektorých} \\ & & \text{mentálnych obsahoch} \\ \downarrow & & \\ (i_2) \text{ veľmi mocný klamúci démon} & \longrightarrow & (o_2) \text{ nemôže klamať o existencii} \\ & & \text{klamaného subjektu} \end{array}$$

Popri službách, ktoré tento prístup poskytuje autorovmu skúmaniu v knihe, má i ďalšiu prednosť. Umožňuje prehľadne a jednoznačne rekonštruovať ľubovoľný ME. Okrem jednotlivých kapitol to Picha demonštruje v rozsiahlom glossári, ktorý sa nachádza na konci knižky. Ten obsahuje pôvodné formulácie (v prirodzenom jazyku) až dvadsiatich klasických filozofických a vedeckých myšlienkových experimentov s odkazmi na pramene, ďalej destilát každého experimentu, prípadne i všeobecný „princíp“, ktorý má ME zdôvodniť alebo diagramatickú rekonštrukciu celého širšieho argumentu. Domnievam sa, že vďaka

tomu Pichova knižka predstavuje veľmi cenný materiál pre prácu s niektorým z rekonštruovaných experimentov (napríklad na seminári k Descartovi, Leibnizovi, filozofii mysle atď.), ale tiež pre výučbu teórie argumentácie. Ako autor viackrát zdôrazňuje, mnohé ME sú vynikajúcimi didaktickými pomôckami.¹ Neviem, či Picha počítal s pedagogickým využitím svojej práce – mohla by tomu nasvedčovať voľba termínu „destilát“, ktorý si mnohí študenti a študentky filozofie iste rýchlo obľúbia. Tak či onak, *Kdyby chyby* sa vo svojich didaktických prednostiach podobajú dobrému myšlienkovému experimentu.

Vráťme sa však k autorovmu postupu. Z uvedeného destilátu, ktorým som trochu predbehol autorov výklad, nie je celkom zrejmé, čo presne pre Pichu znamená experimentovať v myšlienkach. Podľa Pichu je „jadrom myšlienkového experimentovania“ imaginácia. Pri každom ME ide podľa neho o to, aby si experimentátor prostredníctvom predstavivosti vytvoril zmyslovú skúsenosť objektu pri absencii reálneho objektu. Myšlienkový experiment teda definuje ako „súbor pokynov určujúcich, akú konkrétnu situáciu si predstaviť s cieľom niečo zistiť“. Táto definícia je dostatočne široká na to, aby zachytila faktúálne ME (v ktorých ide o posúdenie pravdepodobnosti experimentálneho výsledku) i hodnotniace ME (v ktorých ide o posúdenie napr. morálnej hodnoty modifikovanej vstupnej konštelácie), známe najmä z etiky. Šírka Pichovej definície má však jednu nevýhodu, na ktorú autor sám upozorňuje. Nezahrňa totiž len tie ME, ktoré sa uskutočňujú s *teoretickými* cieľmi (a ktoré, mimochodom, boli výlučným predmetom spomínanej Černíkovej práce), ale doslova akékoľvek predstavovanie si konkrétnej situácie s ľubovoľnými *kognitívnymi* cieľmi. Z teoretického hľadiska nezaujímavý ME, pri ktorom sa usilujem zistiť, ktorým smerom musím otočiť segment Rubikovej kocky, aby farby zapadli, a teoreticky významný ME typu Galileovej úvahy o rýchlosti voľného pádu zloženého telesa, sa potom od seba podľa Pichu nelíšia kvalitou, ale výlučne použitím v argumentácii. Tá hrá v Pichovom prístupe rozhodujúcu úlohu, ku ktorej sa ešte vrátim.

To, že ME vedú u experimentátora k prijatiu presvedčenia (napr. časť Rubikovej kocky musím otočiť doprava; klamúci démon by ma nemohol klamať o mojej existencii), ešte vôbec neznamená, že existujú dôvody, pre ktoré by sme takéto presvedčenia mali považovať za zdôvodnené či pravdivé. Pichova obhajoba vlastného epistemického významu ME sa začína druhou kapitolou, kde predstavuje niekoľko konkurujúcich si „optimistických“ koncepcií. Tie svorne tvrdia, že ME majú epistemický význam, pričom každá z nich predkladá vlastnú odpoveď na otázku, prečo je to tak. Spomedzi apriorizmu (Brown), psychologizmu (Mach), konceptualizmu (Kuhn), asocianizmu a mentálneho modelova-

¹ Vo filozofii to azda platí dvojnásobne. Preto napríklad učebnica Schick – Vaughn (2002) uvádza čitateľa do základných filozofických problémov práve pomocou ME.

nia (Nersessianová) sa Picha najpodrobnejšie venuje posledným dvom menovaným. V niektorých koncepciách sa optimizmus spája s problematickými záväzkami (napr. Brownov platonizmus a anti-empirizmus), iné pri otázke o epistemickom význame ME síce pritakávajú, no kladnú odpoveď podrobnejšie nezdôvodňujú (Kuhn). Spočiatku sa pri porovnaní koncepcií zdá, že asociанизmus, ktorému autor venuje najväčší priestor, predznačuje ďalší postup v knižke. Čitateľ, ktorý takto naletí, bude na konci druhej kapitoly možno prekvapený: Picha tu ohlasuje zámer zakotviť epistemický význam ME bez toho, aby prijal ktorúkoľvek zo spomínaných „optimistických“ koncepcií. Teda bez toho, aby musel vysvetliť, prečo sú ME zdrojom zdôvodnenia.

Všetkým optimistickým koncepciám je totiž spoločná argumentácia, že ME sú založené na istom epistemicky dôveryhodnom procese (napr. pamäti, percepcii, mentálnej simulácii), ktorý sám osebe považujeme za zdroj zdôvodnenia. Schopnosť slúžiť ako zdroj zdôvodnenia sa potom automaticky prenáša aj na ME. Keďže východisko tejto argumentácie – epistemická dôveryhodnosť daného procesu – je ľahko napadnuteľné, Picha volí takpovediac „minimalistický“ postup, založený na pojme *prima facie* zdôvodnenia (pf-zdôvodnenia). Presvedčenie *p* definuje ako pf-zdôvodnené, ak „existujú epistemicky relevantné dôvody pre *p* a zároveň dané presvedčenie nie je vyvrátené“ (s. 71); inými slovami, pokiaľ neexistujú iné, silnejšie dôvody proti presvedčeniu *p*. Epistemicky relevantný dôvod pritom definuje jednoducho ako taký, ktorý vedie k prijatiu (akceptácii) presvedčenia. Ak teda ME vedie u experimentátora k prijatiu určitého presvedčenia, pričom toto presvedčenie nie je iným spôsobom vyvrátené, možno ME považovať za zdroj pf-zdôvodnenia. Takto zdôvodnené presvedčenie je, samozrejme, vždy vyvrátiteľné inými, silnejšími dôvodmi. Picha chápe pf-zdôvodnenie ako „nulový stupeň epistemického zdôvodnenia“ (s. 86), pričom ním podopreté presvedčenia stoja „na začiatku zdôvodňovacích reťazcov, vedľa perceptuálnych, introspektívnych a inštinktívnych presvedčení“ (s. 159).

Pichova argumentácia sa teda na jednej strane neopiera o (skeptikom vždy spochybniteľnú) epistemickú dôveryhodnosť napr. imaginácie, no na druhej strane ešte nerieši problém, či brat' ME vážne, alebo nie. To, že experimentátor prijme na základe ME určité presvedčenie a môže ho považovať za pf-zdôvodnené (resp. epistemológ ho na základe prijatej definície môže takto opísať), ešte nikoho nemusí viesť k tomu, aby prijal rovnaké presvedčenie. Tu do autorovho postupu opäť vstupuje úloha ME v argumentácii. Hoci presvedčenia, ku ktorým experimentátor dospeje prostredníctvom ME, nestačia na to, aby presvedčili jeho oponenta v diskusii, situácia sa môže zmeniť, keď sa z oponenta stane experimentátor. Podľa Pichu majú myšlienkové experimenty „dialektickú funkciu“ (s. 81) v tom zmysle, že za istých okolností dokážu v spore preniesť dôkazné bremeno na oponentovu stranu. Ak oponent v kritickej diskusii

dospeje na základe vykonania ME k prijatiu presvedčenia, ktoré pomocou predloženého ME bráni jeho protistrana, je týmto výsledkom ďalej viazaný. Spor sa buď končí, alebo oponent musí uviesť nové dôvody proti prijatému presvedčeniu. Akceptovaný záver ME však už nemôže ignorovať. Epistemický význam je takto zakotvený bez odkazu na epistemickú dôveryhodnosť procesu experimentovania: podstatná je akceptácia presvedčenia a roly účastníkov diskusie. Ide o originálne a dôvtipné riešenie, ktoré je, pokiaľ viem, v literatúre o myšlienkových experimentoch novinkou.

Kým v druhej a tretej kapitole sa Picha negatívne vymedzil voči optimistickým koncepciám, v štvrtej kapitole sa sám zámerne dostáva do defenzívy. Predstavuje tu eliminativistickú argumentáciu, ktorej cieľom je redukovať epistemický význam ME na význam iných, menej problematických postupov. Podľa Nortonovho eliminativizmu sú ME z hľadiska zdôvodnenia všeobecného záveru, ku ktorému vedú, zbytočné: To podstatné v nich možno redukovať na všeobecný argument bez partikulárnych detailov príznačných pre ME. V prvej časti kapitoly Picha sleduje diskusiu medzi Nortonom, Brownom a Gendlerovou, ktorá sa krúti okolo epistemického významu Galileovho „pisánskeho experimentu“.² Picha pomocou podrobnej rekonštrukcie ukazuje, že obrana optimistov voči Nortonovmu redukcionizmu neobstojí. Obrana jeho vlastnej koncepcie, ktorú navrhuje, opäť úzko súvisí s úlohou ME v argumentácii.

Picha zavádza rozlíšenie medzi dvoma druhmi myšlienkových experimentov. Na jednej strane stoja tie, ktoré fungujú ako *ilustrácie* všeobecného princípu (ilustratívne ME, „i-xperimenty“), ktorý obe strany v diskusii vopred akceptujú. Napríklad Galileov experiment je z hľadiska (väčšiny) dnešných experimentátorov ilustráciou vopred známeho princípu. Ako taký je redukovateľný na argument, medzi ktorého premisami bude daný princíp; formulácia takého argumentu v podobe ME má len didaktické prednosti, no nie svojbytný epistemický význam. „I-xperimentovanie“ nevedie k prijatiu nového presvedčenia, môže viesť len k silnejšej akceptácii existujúceho presvedčenia. Na druhej strane stoja podľa Pichu myšlienkové experimenty, ktoré fungujú ako *exemplá*, teda ako dôvody pre prijatie všeobecného princípu, ktorý experimentátor vopred neprijíma (exemplárne ME, „e-xperimenty“). Picha demonštruje, že ich epistemický význam je neredukovateľný: e-xperiment vedie od premyslenia jednotlivého prípadu k akceptácii nového, všeobecnejšieho záveru. Neredukovateľnosť ME teda autor zakladá nie na jeho obsahu, ale na argumentačnej role princípu vo vzťahu ku konkrétnym premisám, ktorú mu experimentátor pripisuje. Ten istý expe-

² Niektoré argumenty v tejto kapitole budú pravidelným čitateľom *Organon F* známe z Pichovej štúdie (2011).

riment môže v závislosti od experimentátora byť i-xperimentom alebo e-xperimentom. Len v druhom prípade možno hovoriť o epistemickom význame.

Posledná kapitola sa venuje hodnoteniu myšlienkových experimentov. Autor, vychádzajúc zo svojho chápania „úzkych“ ME, najprv predstavuje tri druhy chýb, ktoré postihujú samu štruktúru ME: neprijateľný, nedôveryhodný a nedosiahnuteľný výsledok. Je pozoruhodné, že Picha v tejto časti reflektuje čoraz populárnejšiu „experimentálnu filozofiu“ (*x-phi*), ktorá postupne preniká aj do našich končín. Ukazuje, že experimentalistická kritika dôveryhodnosti predstavivosti môže slúžiť ako upozornenie na obozretnosť pri používaní ME, no nemá vplyv na ich pf-zdôvodňovaciu silu, ako ju chápe autor. Tá totiž, ako som už uviedol, nestojí na dôveryhodnosti imaginácie.

V druhej časti autor skúma tri chyby, spojené s použitím ME v argumentoch: nerelevantnosť, nesprávna generalizácia (na strane toho, kto argument predkladá) a „netrpezlivosť“ (na strane oponenta). V tretej časti poslednej kapitoly Picha predkladá schému argumentu myšlienkovým experimentom. Ten chápe ako prezumpatívny, teda taký, v ktorom vzťah medzi záverom a premisami nie je ani deduktívny, ani induktívny. Záver predstavuje len domnienku, ktorá sa za daných okolností javí plauzibilná. V navrhutej schéme, znázornenej pomocou diagramu, Picha identifikuje miesta, v ktorých môže oponent takpovediac nasadiť páku a pokúsiť sa argument spochybnit'. Ako glosár, o ktorom som sa už zmienil, tak aj táto časť knihy by mohla nájsť využitie pri výučbe teórie argumentácie. Schému, s ktorou pracuje Picha, možno okamžite uplatniť napríklad pri skúmaní úryvkov v glosári.

Celkovo vzaté, Pichova knižka predstavuje obranu myšlienkových experimentov akoby od základnej čiary. Stavia pritom na kombinácii jednoduchého pojmu zdôvodnenia a na funkcii ME v argumentácii. V závere, kde v podobe otázok a odpovedí zhrňa stanovisko, ku ktorému dospel, autor naznačuje, že sa možno opýtať: A nie je to málo? Dôvody pre presvedčenia, ktoré môžu ME poskytovať, sú podľa Pichu „veľmi slabé“. V kontexte vedeckého bádania preto ME (v zmysle e-xperimentu) podľa neho nemôžu slúžiť ako dostatočná podpora všeobecných tvrdení; v oblasti filozofických skúmaní sa ME síce môže javiť ako „najlepší možný dôvod pre tvrdenie“, no stále platí, že ide o veľmi slabú podporu. Niektorí čitatelia budú možno z týchto „skromných“ záverov Pichovej práce rozčarovaní. Je to však cena za vážnosť, s akou autor pristúpil k problémom existujúcich koncepcií a k výhradám eliminativizmu, a tiež za vysoké nároky, ktoré uplatnil na vlastnú analýzu. Keďže *Kdyby chyby* chceli pre ME predovšetkým dobyť nedotknuteľný epistemický terén, nezostalo v nich ani miesto na podrobnejšiu diskusiu napríklad o tom, aké úlohy – a ako – plnia vo vedeckej metodológii, prípadne v niektorej z jednotlivých disciplín. No čitateľ, ktorý sa po prečítaní Pichovej knihy bude zaujímať o podobné otázky, môže sa obrá-

tit' k prácam diskutovaných filozofov a filozofiek, napríklad k práci Nancy Nersessianovej (2008).

Juraj Halas

juraj@halas.tk

Literatúra

- ČERNÍK, V. (1972): *Myšlienkový experiment a produkcia ideí*. Bratislava: Pravda.
- NERSESIAN, N. J. (2008): *Creating Scientific Concepts*. Cambridge: MIT Press.
- PICHA, M. (2011): How to Reconstruct a Thought Experiment? *Organon F* 18, No. 2, 154-188.
- SCHICK JR., T. – VAUGHN, L. (2002): *Doing Philosophy: An Introduction Through Thought Experiments*. London: McGraw-Hill.

Barbora Geistová Čakovská: *Vzťahy synonymie: logická sémantika verzus lingvistické teórie o synonymii*
Bratislava: aleph 2010, 156 strán

Barbora Geistová Čakovská sa problematikou synonymie zaoberala dlhodo-
bo, a to najmä počas svojho doktorandského štúdia. Po istom čase sa k téme
vrátila a v roku 2010 vydala monografiu s názvom *Vzťahy synonymie: logická sé-
mantika verzus lingvistické teórie o synonymii*, ktorá je prepracovanou verziou jej
dizertačnej práce z roku 2007.

Prv než podrobnejšie rozoberiem obsah knihy, uvediem niekoľko poznámok
k štýlu knihy. Hoci ide o knihu od autorky venujúcej sa analytickej filozofii,
neznamená to, že je napísaná formálnym jazykom plným symbolov. Práve na-
opak. Autorka sa v celej monografii vyhýbala technickému jazyku, pokiaľ jej to
obsah aspoň trochu dovolil. Navzdory tomu je text jednoznačný, prehľadný, ba
až transparentný, pretože čitateľ prakticky vždy má pred sebou centrálnu prob-
lematiku a vývoj jej spracovania v rámci celej monografie. Autorka využíva jed-
noduchý, čitateľný jazyk bez úsilia nadbytočne využívať príklady či opakovať
myšlienky. Čitateľovi je vždy zrejmé – a autorka to občas zdôrazňuje – čo sú zá-
kladné stavebné kamene práce. Jednotlivé definície sú pochopiteľné a ich úloha
je z textu vždy zrejmá. Definície aj základné termíny sú v texte vždy zvýraznené.
Ich počet postupne narastá, no práve vďaka prehľadnému štýlu monografie nie
je problém nájsť relevantné pasáže. Autorka vychádza čitateľovi v ústrety aj
tým, že tieto termíny nepoužíva najprv skusmo, v intuitívnom význame, ale vždy

až v jasne stanovenom teoretickom význame. Každý nový termín, s ktorým sa čitateľ stretne, sa prakticky ihneď vysvetlí. Takto je vždy zrejmé, akú úlohu hrá, prečo ho treba používať, aj potenciálne nástrahy a problémy, ktoré vzniknú jeho ignorovaním. Pri čítaní monografie získate pocit, že s autorkou vediete rozhovor, v ktorom vám pregnantne a zároveň nenásilne sprostredkúva výsledky svojej práce. Vzhľadom na to, že práca sa pohybuje na rozhraní viacerých disciplín, je napísaná výborným ústretovým štýlom, ktorý veľmi zjednodušuje prípadné diskusie odborníkov i nadšencov z rôznych oblastí. Z tohto hľadiska by mohla pre autorov či študentov slúžiť aj ako príručka o tom, ako písať zrozumiteľný text.

Teraz prejdem k obsahu práce. Ako už názov knihy naznačuje, neexistuje iba jediný vzťah synonymie. Autorka sa zaoberá najmä vzťahmi synonymie v logickej sémantike a lingvistických teóriách, pričom prvé označuje termínom LOGsynonymia a druhé LINsynonymia. V každej z týchto oblastí existuje množstvo odbornej literatúry, ktorá sa zaoberá explikáciou synonymie. Prednosť recenzovanej monografie spočíva v tom, že je interdisciplinárna. Autorka pomocou logickej sémantiky skúma a rozlišuje rôzne vzťahy synonymie ukryté pod jedným termínom používaným v lingvistike. Pracuje ako biológ pri skúmaní organizmov pomocou rôznych technických nástrojov. Kniha je aj preto rozdelená na dve hlavné kapitoly. V prvej sú predstavené nástroje skúmania. Ako aj sama autorka tvrdí, čitatelia oboznámení so základmi problematiky logickej sémantiky v tejto kapitole nenájdu nič nové. Táto prvá, menšia kapitola slúži hlavne pre čitateľov z iných disciplín, pretože v nej sú ústretovo vysvetlené základné termíny logicko-sémantickej analýzy ako *možný svet*, *rozsah* či *intenzia*. Druhá kapitola už predstavuje originálne výsledky autorkinej práce. Skúma vzťahy synonymie deklarované v lingvistických prácach týkajúcich sa najmä slovenského jazyka. Hoci empirické dáta pochádzajú najmä zo slovenského jazyka, resp. slovenských synonymických slovníkov, neznamená to, že výsledky sú obmedzené len na slovenský jazyk. Pomocou logicko-sémantickej analýzy pri skúmaní lingvistických teórií synonymie autorka odhalila niekoľko odlišných pojmov. Ako sama v závere tvrdí, „pod termínom synonymia či synonymita sa skutočne skrýva podstatne väčšie množstvo rôznych vzťahov, ako by používateľ jazyka očakával, a dúfame, že sme odhalili a odlišili väčšinu z nich alebo aspoň naznačili, akou cestou sa možno za týmto účelom vybrať. Tento bod pokladáme za najväčší prínos tejto práce“ (s. 146).

Prejdime si jednotlivé kapitoly podrobnejšie. Prvá kapitola má názov *Synonymita slov z pohľadu logickej sémantiky – LOGsynonymia*. Autorka rozlišuje termín *LOGsynonymie* a termín *(byt) LOGsynonymý s...*; podobne bude neskôr rozlišovať medzi termínmi *LINsynonymia* a *(byt) LINsynonymný s...* Prvé z dvojice termínov sú v supozícii *de dicto* a druhé v supozícii *de re*. Termín *LOGsynonymia* sa vzťahuje všeobecne na daný vzťah. Termín *(byt) LOGsyno-*

nymný s... sa týka vzťahu dvoch konkrétnych jazykových výrazov. V prípade LOGsynonymie to môžu byť aj dva jazykové výrazy z dvoch rôznych jazykov. Podľa definície sú dva jazykové výrazy LOGsynonymné vtedy a len vtedy, keď (vtt) majú v jazyku, ktorého sú súčasťou, ten istý kognitívny význam. Vzhľadom na to, že empirickými dátami, ktoré autorka skúmala, sú empirické výrazy v synonymických slovníkoch, a na to, že vzťah synonymie v lingvistike (LINSynonymia) je vzťahom medzi dvoma slovami, rozhodla sa autorka použiť prostriedky intenzionálnej logiky. Kognitívny význam jazykového výrazu explikuje pomocou pojmu intenzie. Jazykové výrazy signifikujú svoju intenziu a referujú na svoj rozsah. Rozsah jazykového výrazu je hodnotou intenzie v konkrétnom svetamihu. Napríklad predikát *učiteľ* signifikuje vlastnosť „(byť) učiteľom“ a jeho rozsahom je v závislosti od svetamihu trieda individuí s touto vlastnosťou. Identita jazykového výrazu sa v monografii explikuje pomocou Cmorejovho rozlíšenia medzi *jazykovými výrazmi* a *semivýrazmi*. Na identifikáciu jazykového výrazu potrebujeme poznať okrem semivýrazu aj jeho kognitívny význam (intenziu). Jazykový výraz je daný spojením semivýrazu a kognitívneho významu. Toto elementárne rozlíšenie medzi jazykovým výrazom, jeho semivýrazom a kognitívnym významom umožní v druhej kapitole autorke prehľadne formulovať jej výsledky. Zároveň je zřejmé, že vzťah LOGsynonymie nezávisí od nekognitívneho významu jazykových výrazov.

V tejto kapitole sa pomocou spomenutých technických prostriedkov postupne zavádzajú pojmy koreferencie, L-ekvivalencie a LOGsynonymie empirických výrazov. Empirické predikáty sú vo vzťahu koreferencie v nejakom svetamihu vtt majú v danom svetamihu ten istý rozsah. Vo vzťahu L-ekvivalencie sú dva empirické predikáty vtt majú ten istý rozsah vo všetkých svetamihoch. LOGsynonymné sú dva jazykové výrazy vtt pri spomínanej explikácii kognitívneho významu signifikujú tú istú intenziu. Podstatnú časť prvej kapitoly tvorí aj formulácia testu na skúmanie LOGsynonymie jazykových výrazov. Dostatočným testom LOGsynonymie dvoch jazykových výrazov je možnosť ich vzájomného a vzájomného spätného nahrádzania v každom jazykovom kontexte bez zmeny kognitívneho významu daného jazykového kontextu. Kognitívny význam jazykového kontextu sa explikuje pomocou známeho pojmu intenzionálnej štruktúry. Vzhľadom na definície L-ekvivalencie a LOGsynonymie autorka dokazuje, že oba tieto vzťahy sú reflexívne, symetrické, tranzitívne (t. j. ekvivalenčné).

V druhej kapitole s názvom *Lexikálna synonymia z pohľadu lingvistiky – LINSynonymia* sa porovnávajú vlastnosti LOGsynonymie a LINSynonymie. V úvode sa uvádza zásadná všeobecná definícia LINSynonymie, ktorá tvorí východisko ďalšieho skúmania. LINSynonymné sú slová s totožným významom, ale aj s blízkym či príbuzným významom. Z definície LINSynonymie vyplýva, že tento termín sa vzťahuje na viacero rôznych druhov vzťahov medzi jazykovými

výrazmi. Preto autorka rozlíši vzťahy rovnoznačnosti a blízkoznačnosti. Termín „význam“ má však podľa Geistovej v lingvistike mnohé rôzne použitia, pričom ako sama píše, lingvistické teórie významu „neposkytujú... uspokojujúcu explikáciu či aspoň priblíženie pojmu kognitívneho významu jazykového výrazu, ktorý by sme mohli využiť pri našom skúmaní“ (s. 58). Následne opíše štyri rôzne skupiny lingvistických koncepcií synonymie, pričom nepovažuje tento zoznam za vyčerpávajúci.

Veľmi zaujímavú časť monografie predstavuje kapitola o nekognitívnom význame slov. Okrem úsilia o technické uchopenie daného termínu považujem za podnetný spôsob, ako sa autorka vyrovnala s mierou neurčitosti a subjektivity, ktoré sú spojené určovaním nekognitívneho významu slov. Východisko tvoria štylistické príznaky slov uvádzané v synonymických slovníkoch. Hoci nie každý musí súhlasiť s týmto kritériom, pre skúmanie vplyvu nekognitívneho významu slov na ich synonymiu (LINSynonymiu) je voľba nejakého kritéria zásadná. Prípravné pasáže druhej kapitoly uzatvára krátka explikácia Tichého termínu „náležitosti“, ktorý sa využíva v ďalších častiach monografie.

Najprv autorka skúma vzťah rovnoznačnosti. Uvádza jeho definíciu pomocou termínov kognitívneho a nekognitívneho významu. Následne formuluje test rovnoznačnosti dvoch slov založený na vzájomnom a vzájomnom spätnom nahradzovaní slov v jazykových kontextoch, pričom použitie nových termínov demonštruje na konkrétnych príkladoch. Na to, aby autorka zistila, či sa vzťah rovnoznačnosti, ktorý patrí medzi vzťahy LINSynonymie, odlišuje od vzťahu LOGsynonymie, analyzuje, či ide o ekvivalenčný vzťah. Vzhľadom na definíciu rovnoznačnosti nie je rovnoznačnosť ekvivalenčným vzťahom. Rozsiahlejšiu časť monografie tvorí skúmanie vzťahu blízkoznačnosti, pri ktorej Geistová Čakovská identifikovala tri rôzne druhy. Rozdiely medzi druhmi blízkoznačnosti sú dané rôznymi podobami vzťahov, do ktorých môžu vstupovať kognitívne a nekognitívne významy slov. Keďže okrem podobnosti zohráva úlohu aj odlišnosť medzi významami, autorka ukazuje, že treba uviesť okrem pozitívnych aj negatívne jazykové kontexty, v ktorých sa aplikuje test vzájomnej a vzájomnej spätnej nahraditeľnosti. Kým v pozitívnych jazykových kontextoch môžeme vzájomne nahrádzať blízkoznačné slová, v negatívnych jazykových kontextoch tak robiť nemôžeme. Pri klasifikácii vzťahu blízkoznačnosti sa rozlišujú spoločné časti kognitívneho významu slov od rozdielnych častí, na čo sa využíva koncepcia náležitostí. Aj pri vzťahoch blízkoznačnosti sa ukázalo, že nejde o ekvivalenčný vzťah. Druhú kapitolu uzatvára krátka časť, v ktorej sa odlišujú vzťahy hyponymie a hyperonymie od vzťahov LINSynonymie a LOGsynonymie.

Recenzovaná monografia predstavuje výsledky využitia nástrojov intenzívnej logiky, Cmorejovho rozlíšenia medzi jazykovým výrazom a semivýrazom či Tichého koncepcie náležitostí pri analýze vzťahov synonymie v lingvistike.

Aby však nevzniklo nedorozumenie, musím upozorniť, že autorka uznáva, že intenzionálna logika má svoje obmedzenia, pričom rešpektuje prednosti hyperintenzionálnych systémov ako TIL. Intenzionálnu logiku si zvolila, pretože jej prostriedky sú podľa nej dostatočné pre jej prácu. Vzhľadom na charakter prezentovaných výsledkov sa ukázala daná voľba ako oprávnená.

Často vzniká problém spôsobený tým, že si autor alebo vedec neuvedomí obmedzený obor platnosti predpokladov a zákonov nejakej teórie a bezbreho ich začne používať aj mimo tento obor. Podobný problém vzniká, ak autor pri predkladaní svojich téz neupozorní na obor ich platnosti. Recenzovaná monografia je z tohto pohľadu výnimočná. Autorka vždy dôrazne upozorňuje na ohraničený obor svojho skúmania. Rovnako dôrazne upozorňuje, že zvolené nástroje logicko-sémantickej analýzy nemusia byť a nie sú vždy tými najlepšími. Naproti tomu však ukazuje, že tieto nástroje použité adekvátnym spôsobom na ohraničenom obore skúmania vedú k zaujímavým výsledkom. Monografia Barbory Geistovej Čakovskej je toho dôkazom.

Miloš Kosterec

milos.kosterec@gmail.com

Stephen Hawking – Leonard Mlodinow: *The Grand Design*
London: Bantam Press 2010, 208 pages

Stephen Hawking has become a notable icon of contemporary cosmology and theoretical physics. One of his indisputable contributions is that he is systematically trying to bring science to the general public and attempting to range scientific areas into a broader philosophical framework. Despite this, or rather because of this, his claim to be a philosopher raises some embarrassment. There is no doubt that Hawking is a leading physicist, but what kind of philosopher is he?

I will set out from his latest book, *The Grand Design*, written together with physicist Leonard Mlodinow (Czech edition: Hawking – Mlodinow 2011). At the first sight its name strongly evokes the concept of Intelligent Design. Is Hawking actually advocating intelligent design, or is it just a marketing gimmick? Of course, the second is correct. However, there arises a question, to which extent are other Hawking's and Mlodinow's formulations rather "marketing gimmicks".

But let us approach Mlodinow and Hawking's philosophizing. At the very beginning, the authors surprise us by claiming that "philosophy is dead now...".

This categorical statement provoked irritated reactions, and not only among philosophers (among others Jaroslav Peregrin; cf. Peregrin 2011). I am aware that many branches of philosophy have nothing to say to physics, and perhaps to science as a whole. However, any adept of philosophy knows that there is no “one philosophy”. There are just teachings of various philosophers, philosophical schools and movements. Do Hawking and Mlodinow believe that all philosophical effort belongs to the past? That all philosophical questions mentioned in the book – existence of the world, meaning of life, role of Creator, etc. – can be solved by “living” physics instead of “dead” philosophy?

The whole text unfortunately exhibits real contempt for philosophy. And this is not because of the authors’ knowledge of philosophy, but because of their ignorance.¹ The parts dealing with ancient philosophy are full of inaccuracies and misinformation. Some examples: Epicurus is said to disagree with atomism, but the opposite is true (p. 32 in English / p. 30 in Czech edition). Epicurus was a proponent of atomism; he learnt about it from a certain Nausifanes, accepted it and developed it. He disagreed only with its hidden absolute determinism. (This is why he introduced a kind of uncertainty which is nice reminiscent of the indeterminacy of quantum mechanics. This fascinating fact, however, is omitted in the book.) The justification for Aristotle’s approach to atomism is also misleading. Aristotle did not reject the concept of atoms because “they could not reconcile with the notion that human beings are made up of inanimate objects without a soul”, as written in the book (p. 32/30). It cannot be so because the atomists accepted soul (they even claimed that the soul is composed of small round atoms). The opinion that the Greeks did not discover the rules of scientific procedure is at least controversial. Perhaps all modern science is built on the Aristotelian form of thinking, and Aristotle’s formal scheme has become a model for writing of scientific texts. The assertion that Aristotle rejected the idea of science based principally on observation is also not valid. On the contrary he was engaged in natural science that was based on observations. Observations also served as a basis for his astronomical conceptions. Aristotle is further said to have accepted geocentrism for certain “mystical reasons” (p. 55/48). Where does this view come from? Geocentrism is nothing but extended egocentrism, and egocentrism is inherent to all living

¹ Also many statements referring to physics are so oversimplified that they are not correct. For example in the part describing initial inflation of the universe (p. 167 in English / p. 138 Czech edition) it is written that expansion was “much faster than light speed”. However, speed of expansion cannot be compared with speed of light, the first being relative speed (sec^{-1}), the second absolute ($\text{m}\cdot\text{sec}^{-1}$).

creatures, including us. At first glance it looks simpler, and more natural than any other “centrism”. No “mystical reasons” are needed.

Furthermore, Laplace is said to have been the first to clearly formulate “scientific determinism” (i.e., absolute determinism). It is not commonly known – not only among physicists – that the same idea was formulated by the Stoics two thousand years before – see fragment SVF II, 944 (e.g., Long 1986).

“There is no concept of reality which would be independent of any concept or theory” (p. 57/50). This approach, popularized some years ago by the Matrix movie, had also been formulated in antiquity. We should particularly mention the Sophists, then Immanuel Kant, and from contemporary period we should mention W.V.O. Quine (“ontological relativity”), Gregory Bateson, and others. It is also unacceptable to claim that Plato “first questioned the essence of reality”. Long before him, this had been done by Parmenides (“Consideration and being is the same...”) and many Sophists. Of course, also by the Buddhists, but the East is generally ignored on the West.

The claim about the death of philosophy is framed by a text containing inaccuracies and misleading statements. Philosophers who oppose the verdict on the death of philosophy furthermore point out that Hawking with Mlodinow are also philosophizing and repeating mistakes committed by philosophers before them.

In particular, Hawking’s radical and inadequately self-reflected reductionism is hidden in the central concept of the “theory of everything”. Taken literally (not only as a marketing gimmick) it is logically inconsistent, if there is no specification of what “everything” means. “All” or “everything” does not form any set (in mathematical sense) that can be unambiguously defined and summarized. It is a contradictory pseudo concept, as was demonstrated, e.g., by the Russell paradox more than one century ago.

Further, Hawking’s final declaration that our universe was created out of nothing (p. 227/187) can sound impressively and looks also in accordance with the Book of Genesis (or rather with its later interpretation from the second century AC). However, Hawking obviously assumes not creation from “real nothing”, but “creation out of empty space”. Nevertheless, since the days of general relativity and quantum electrodynamics, empty space (vacuum) has not been “nothing” (really empty, i.e. a void), but “something” – a gravitational field, a sea of virtual particles, etc. Spontaneous creation is therefore not real “creation”, but only “mutation”, as Thomas Aquinas would say. Therefore it cannot be the ultimate answer to the question “Why is there something rather than nothing”. It only shifts the question to “How did the empty space arise?” (Similar is true for divine creation, because “How did God emerge?”).

One of the main aims of Hawking's "non-philosophical philosophy" was, perhaps, to show that thanks to M-theory (the importance of which is overestimated according to some physicists) there is no need of Creator for an explanation of physical world. However, a similar opinion was formulated two centuries ago by Pierre Simone Laplace in his dialogue with Napoleon held over his "Celestial mechanics". Hawking's considerations are, of course, based on a deeper level of physical reality. But did Hawking really think that the existence or non-existence of God can be proved by means of M-theory or other physical theory?

Fortunately, not all branches of philosophy are dry and barren, but our authors are not acquainted with them. However, there is one thing in which Mlodinow and Hawking are unfortunately right. The gap between philosophy and science is widening. Most philosophers do not understand science. And *The Grand Design* testifies that even prominent physicists do not understand philosophy.

Despite my critical remarks, I would like to stress that the Hawking and Mlodinow's book is definitely interesting and worth reading (especially for not very demanding reader). I only regret that such respected authors spread into the public domain views that degrade philosophy as a whole and support narrow scientism.

Peter Zamarovský

zamarovs@fel.cvut.cz

References

- HAWKING, S. – MLODINOW, L. (2011): *Velkolepý plán*. Transl. by Jan Klíma. Praha: Argo, Dokořán.
- LONG, A. A. (1986): *Hellenistic Philosophy*. Duckworth. (Czech edition: *Hellénistická filosofie*. Praha: Oikúmené, 2003.)
- PEREGRIN, J. (2011): Je filosofie „mrtvá“? *Filosofie dnes* 3, č. 2, 91-93.
<http://filosofiednes.ff.uhk.cz/index.php?journal=hen&page=article&op=viewFile&path%5B%5D=86&path%5B%5D=93>

50 rokov Katedry logiky a metodológie vied Filozofickej fakulty Univerzity Komenského v Bratislave

Pred päťdesiatimi rokmi, presne 1. septembra 1962, vznikla na Filozofickej fakulte Univerzity Komenského Katedra logiky. V porovnaní so starovekými gréckymi filozofickými školami, z ktorých niektoré trvali viac ako 500 rokov, je to málo. V porovnaní s dobou trvania Academie Istropolitana je to zhruba dvojnásobok. Z hľadiska generácií bádateľov na katedre pôsobí už tretia generácia – žiaci prvých učiteľov už majú svojich žiakov ako kolegov. Tretia generácia môže ťažiť z výsledkov prvých dvoch generácií a nemusí prekonávať detské choroby vzniku inštitúcie a odboru. Jej predstavitelia sa môžu postaviť na plecía predchádzajúcich členov katedry, a aj keby to neboli plecía obrov, určite môžu uvidieť viac.

Patríme k najmenším katedrám na fakulte a náš počet osciluje okolo sedmičky statočných. História katedry do roku 2000 máme podrobne spracovanú a publikovanú najmä vďaka iniciatíve Jozefa Viceníka. Dnes na katedre pôsobí jeden profesor, dvaja docenti, traja odborní asistenti a školíme dvoch doktorandov. Učíme predmety z logiky a metodológie vied na štyroch fakultách Univerzity Komenského – na Filozofickej fakulte, Právnickej fakulte, Fakulte sociálnych a ekonomických vied a na Fakulte managementu. V porovnaní s fakultným pomerom je naša publikačná aktivita nadpriemerná a podobne je to aj so získavaním grantov.

Hlavným vedeckým záberom našej katedry v oblasti logiky je logická analýza prirodzeného jazyka, logická sémantika, intenzionálna a modálna logika, epistemická logika, filozofické problémy logiky a aplikácia logiky v právnej teórii a praxi. V oblasti metodológie vied je to najmä problematika vedeckého zákona, otázky vedeckého vysvetlenia a metodológia sociálnych vied. Často však začíname aj do problematiky analytickej filozofie. V personálnej únii sa podieľame na úspešnom vydávaní časopisu pre analyticky zameranú filozofiu a logiku *Organon F*.

Na rozdiel od mnohých iných pracovísk sa katedra vyznačovala tým, že umožňovala tradične otvorené diskusie na rozmanité témy – odborné, politické, ale aj iné témy; katedrové zasadnutia boli krátke a zápisnice stručné. Túto tradíciu sa usilujeme udržiavať.

Pri príležitosti 50. výročia vzniku katedry sme zorganizovali vedeckú konferenciu na tému „Systémy dedukcie: od extenzionálnej logiky k hyperintenzionálnej logike“, ktorá sa uskutočnila 17. a 18. 9. 2012 v Kaštieli v Dolnej Krupej.

Prvé dva bloky konferencie boli venované teórii dedukcie v Transparentnej intenzionálnej logike, ktorej pôvodcom je Pavel Tichý. Hlavný príspevok mala Marie Duží (VŠB – Technická univerzita, Ostrava) na tému *Dedukce v TIL: Přejchod od jednoduché k rozvětvené hierarchii typů*. Jiří Raclavský (Masarykova univerzita, Brno) prezentoval tému *Dedukce a logická analýza v Tichého logice*; Ivo Pezlar (Masarykova univerzita, Brno) mal príspevok s názvom *Tichého dvou-dimenzionální pojetí inference*; Karel Šebela (Univerzita Palackého, Olomouc) hovoril o Tichého teórii dedukcie. František Gahér a Lukáš Bielik (obaja Univerzita Komenského v Bratislave) mali dve spoločné prezentácie s témami *Prečo len (nutné) pravdy ako predpoklady deduktívnych úsudkov?* a *Využitie pojmu „zhoda“ v hyperintenzionálnej dedukcii*.

V ďalších dvoch – tematicky rôznorodejších – blokoch vystúpil Pavel Materna (Akademie věd České republiky, Praha) s príspevkom *Expresivita logické analýzy přirozeného jazyka*; Jaroslav Peregrin (Akademie věd České republiky, Praha) s prezentáciou na tému *Odkud se berou axiomy logiky?*; Ludmila Dostálová (Západočeská univerzita, Plzeň) mala príspevok s názvom *Aristotelská logika jako deduktívni systém: extenzionální nebo intenzionální?*; Miloš Kosterec (Univerzita Komenského v Bratislave) hovoril o problematike anaforického reťazca; Igor Sedlár (Univerzita Komenského v Bratislave) mal vystúpenia s názvom *Poznanie a subštruktúrne logiky* a Igor Hanzel (Univerzita Komenského v Bratislave) s prednáškou nazvanou *Predstava, myseľ a produkcia myšlienok*.

V záverečnom monotematickom bloku mal prezentáciu Martin Homola (Univerzita Komenského v Bratislave) na tému *Ontologie v informatike a v reprezentácii znalostí*; Martin Baláž (Univerzita Komenského v Bratislave) hovoril o argumentácii a reprezentácii znalostí; Jozef Šiška (Univerzita Komenského v Bratislave) mal príspevok na tému *Logické programovanie*; Michal Čertický (Univerzita Komenského v Bratislave) prezentoval tému *Modely akcií a ich indukcia* a Ján Šefránek (Univerzita Komenského v Bratislave) mal vystúpenia s názvom *Reprezentácia znalostí a nemonotónne usudzovanie*.

Na konferencii sa zúčastnili okrem súčasných členov katedry jej zakladajúci členovia Augustín Riška a Pavel Cmorej i ďalší bývalí členovia. Otec-zakladateľ (prvý vedúci katedry) akademik a bývalý rektor Vojtech Filkorn pred niekoľkými rokmi zomrel. Konferenciu pozdravil v mene dekana fakulty prodekan FiFUK doc. Anton Eliáš, vedúca Katedry filozofie a dejín filozofie UK doc. Mariana Szapuová, vedúci Oddelenia logiky Filozofického ústavu AV ČR v Prahe prof. Jaroslav Peregrin, vedúci Katedry logiky Filozofickej fakulty Karlovej Univerzity v Prahe doc. Vítězslav Švejdar.

Katedra spolupracuje najmä s Katedrou filozofie a dejín filozofie Filozofickej fakulty UK a s Oddelením analytickej filozofie SAV. Z Čiech je to najmä Oddelenie logiky Filozofického ústavu AVČR v Prahe a Katedra logiky Filozofickej

fakulty Karlovej Univerzity v Prahe. Spolupracujeme tiež s kolegami z katedier filozofie filozofických fakúlt Masarykovej Univerzity v Brne, Západočeskej univerzity v Plzni, Univerzity Palackého v Olomouci, s kolegami z Katedry informatiky Fakulty elektrotechniky a informatiky Vysokej školy banskej – Technickej univerzity v Ostrave, ale aj s inými pracoviskami, čoho príkladom je aj účasť kolegov z Katedry aplikovanej informatiky z Fakulty matematiky, fyziky a informatiky našej univerzity.

Dúfam, že sme úspešne nadviazali na dielo otcov zakladateľov, že aj tí po nás budú katedru rozvíjať a že sa dožije aspoň toľko rokov, koľko trvala platónska akadémia.

František Gaheř

Frantisek.Gaheř@uniba.sk

25 Years In Contradiction (University of Glasgow, 7-9/12/2012)

It has been 25 years since Graham Priest initiated a revolution in logic (and philosophy) and published a book called *In Contradiction – A Study of the Transconsistent*. The book provides a comprehensive defense of a philosophical stance dubbed dialetheism. Although still controversial, it is now considered as one of the most important philosophical works in these days. Due to its influence on logicians and philosophers all over the world, partisans of dialetheism decided to celebrate its quarter-of-a-century existence by organizing a conference that would reflect the actual status of the debate. The conference – named *25 years In Contradiction* – took place at the University of Glasgow in December, 7-9, 2012.

It was clear at the beginning that the list of speakers was about to guarantee a unique philosophical experience. The first keynote called ‘Rejection, Denial: A Negative Appraisal of Dialetheism’ was given by Alan Weir (Glasgow). Quite humorously, though, the talk outlined the most discussed points of Priest’s book, including the principal impossibility to refuse dialetheism, the problems concerning the classical recapture, reasoning from classical situations, falsity preservation in dialetheism or the entailment as such. Ben Burgis (Miami) presented a paper (co-authored with Otavio Bueno) called ‘Liars with Curry: Why Dialetheists Violate the Principle of Uniform Solution’. As Burgis argued, given the dissimilarity between the Curry Paradox and the Liar Paradox, dialetheists face a dilemma: either the so-called Inclosure Schema – basically, a pattern

common to all paradoxes – describes the structure of Curry Paradox, or not. If it does, so much the worse for Priest's reliance on the Inclosure Schema. If it does not, Burgis concludes, so much the worse for the schema as a universal tool to paradoxes. The last Friday's talk was given by Gareth Young (Glasgow). In his 'Dialetheism and Expressive Limitations', Young outlined two putative limitations of dialetheism (originally presented in Steward Shapiro's *Simple Truth, Contradiction and Consistency*). By means of difference between just false, false and not true sentences, he considered analogical distinction between just invalid, invalid and not valid inferences in order to avoid certain unwanted consequences. He stayed, however, rather skeptical on the issue.

The Saturday's session commenced with Sebastiano Moruzzi's (Bologna) presentation of the paper entitled 'Relativism in Contradiction' (co-authored with Annalisa Coliva). His contribution concerned the question whether relativism can be a descriptive project. Since disputes in the subjective domains appear to involve judgments with incompatible contents, he claimed, dialetheist relativism is worth considering. Can Baskent's (Paris) 'Dialetheism and a Game Theory Paradox' discussed two approaches to Brandenburger – Keisler Paradox, namely Non-well-founded set theoretic approach and paraconsistent approach. The morning session concluded the second keynote by JC Beall (Connecticut) with his 'Free of Detachment: Logic, Rationality, and Gluts'. Since glut theory involves the invalidity of modus ponens and modus ponens is central to our theoretical reasoning, the theory seems to be implausible. However, as Beall (very interestingly) tried to show, we still can successfully pursue rational theorizing despite the invalidity of modus ponens. Karin Verelst (Vrije Universiteit Brussel) in her 'Zeno's Paradoxes. A Cardinal Problem' demonstrated that Zeno's Paradoxes of Motion and Paradoxes of Plurality come down to the same and whatever it was that Zeno's opponents refuted was certainly not Zeno. Franz Berto's (Aberdeen) 'Exclusion vs. Explosion: Going Predicative?' explores a 'foothold on undisputed ground' that we can rely on when we accept true contradictions. One way of getting such a foothold is, for Berto, to consider exclusion as primitive, meaning that exclusion should not be defined at all. The third keynote 'Inconsistency and Incompleteness, Revised' by Steward Shapiro (Ohio State) was partly a reminder of his previous paper 'Incompleteness and Inconsistency' in which he discusses reliance on an informal provability argument, partly a quite original contribution to the logic of inconsistency.

The last day of the conference began with Aaron Cotnoir's (St. Andrews) (co-authored with Zach Weber) 'Inconsistent Boundaries of Ordinary Objects: Toward a Paraconsistent Mereotopology'. As Cotnoir pointed out, the existence of boundaries of everyday objects is problematic, as in classical mereotopology three conditions – boundaries exist, space is topologically connected and

discrete entities can be in contact – are mutually inconsistent. However, as he concluded, the problem can be met with a paraconsistent mereotopology. The paper ‘Remarks on Naïve Set Theory Based on LP’ by Hitoshi Omori (Kobe/CUNY) showed, that although systems of paraconsistent logic with classical negation were thought to be quite problematic, there are several possibilities in formulating theories if we have both paraconsistent negation and classical negation. The penultimate talk, by Diego Tajer, (Buenos Aires) was called ‘Dialetheism and the Curry-Validity Paradox’. As Tajer maintained, Validity-Curry paradox shows that if validity is contractive, it is hard to represent it in a dialethic framework. Thus, the inconsistency of validity is a better explanation for Curry-Validity Paradox in the framework.

Finally, Graham Priest (CUNY/Melbourne). At the very beginning of his talk, Priest confessed that the book was rejected by ten major publishers before one accepted it, and thus if somebody had told him about the 25 anniversary of the book, he would not have believed. Also, he acknowledged that the debate on dialetheism has increasingly improved over the years and expressed deep acknowledgement to all those authors who had contributed to the conference. Although Priest’s talk was called ‘Contradiction, Language and the World’, he spent the whole of his time addressing several objections and suggestions being raised during the conference. Given the notes he was making thorough every talk, Priest precisely responded to all the presented papers. He stressed several details that either had been overlooked, or underestimated and thus provoked an exciting discussion. It was especially interesting to see how detailed and delicate any philosophical debate must be in order to avoid potential confusions and misunderstandings.

To conclude, I dare to say that the conference met the highest criteria for philosophical conferences and we can only hope that such conferences will be more frequent in the near future.

Martin Vacek

martinvacekphilosophy@gmail.com

Projekt Analytická teologie

Záměrem této zprávy je upozornit na zrod analytické teologie z bouřlivého rozvoje analytické filosofie náboženství a představit mezinárodní projekt *Analytická teologie*.

Nejprve se ptejme: co je to analytická filosofie náboženství a analytická teologie? Obě zkoumají význam, konzistenci, pravdivost a rozumnost náboženských výpovědí. To jsou zejména výpovědi o existenci a povaze nadpřirozena, v první řadě Boha a posmrtného života. Liší se však v premisách, jež v provozu dané disciplíny není třeba zdůvodňovat. Pojednání z analytické filosofie náboženství by nemělo jednoduše předpokládat, zda se Bůh zjevil, ani co zjevil. (Může ovšem zvažovat, co plyne z takových předpokladů, případně argumentovat v jejich prospěch či neprospěch.) Oproti tomu pojednání z analytické teologie, jakožto pojednání teologické, prostě předpokládá jako pravdivé, že se Bůh zjevil, případně co zjevil. Tolik k rozdílu. Na druhou stranu je mezi oběma disciplínami shoda v analytickém postupu, jenž se vyznačuje poměrně častými a netriviálními aplikacemi logiky či sémantiky ve výkladu pojmů, tvrzení a argumentů. Tato shoda v metodě slibuje snadnou komunikaci mezi analytickými filosofy náboženství a analytickými teology. Slibuje i dělbu práce: zástupci první strany mívají lépe osvojené filosofické nástroje, zatímco zástupci strany druhé zase lepší znalosti náboženských textů a praktik.

Třeba uznat, že analytická teologie se pořád teprve rodí (Crisp 2009, Crisp – Rea 2009, Flint – Rea 2009, McCall – Rea 2009, Rea 2009, Taliaferro – Meister 2009, Marmodoro – Hill 2011). Přitom navazuje na předchozí úspěchy o dost zavedenější analytické filosofie náboženství. Ta jistě v první polovině 20. století neměla na různých ustláno, mimo jiné kvůli tehdy populárnímu názoru, že smysluplné jsou jen věty rozhodnutelné přírodovědecky či lingvisticky. Od 50. let ale sílila, s tím jak víc a víc analytických filosofů odhazovalo spolu s podobnými kritérii smysluplnosti i skrupule vůči metafyzice, metafyziku nadpřirozena nevyjímaje (srov. MacIntyre – Flew 1955). Tomuto obratu už v 30. letech předcházely první ukázky analytické scholastiky a analytické filosofie náboženství, produkované filosofy tzv. Krakovského kruhu (viz Pouivet 2011). Vedle kolapsu verifikacionismu lze do mezníků analytické filosofie náboženství počítat zahájení aplikace epistemologie na otázky racionality náboženských přesvědčení (Plantinga 1967), otevření problémů chápání a konzistence vlastností Bohu tradičně přisuzovaných nezávisle na jakémkoli údajném zjevení (Swinburne 1977), zohledňování fyzikální kosmologie (Craig 1979, Leslie 1989), zahájení využívání induktivní logiky (Swinburne 1979) a otevření problémů chápání a konzistence vlastností Bohu tradičně přisuzovaných v závislosti na údajném zjevení (Plantinga 1984, Morris 1986). Dnes již analytická filosofie náboženství v angloamerických zemích přitahuje značnou pozornost. Její hlavní časopisy (*Religious Studies*, *International Journal for Philosophy of Religion*, *Faith and Philosophy*, *Sophia*, *Philosophia Christi*) publikují příspěvky předních filosofů; některá prestižní nakladatelství, jako třeba Oxford University Press, v několika letech posledního desetiletí vydala více knih o filosofii náboženství než o filosofii jazyka, vědy

i myslí (Smith 2001). Vycházejí objemní průvodci analytickou filosofií náboženství (Quinn – Taliaferro 1999, Martin 2007, Craig – Moreland 2009). Z různých náboženství se nejvíce pozornosti dostává křesťanství. Traktovány jsou však i specifické výpovědi judaismu, islámu, buddhismu, hinduismu, konfucianismu, taoismu a afrických náboženství. Odkud ale tolik zájmu o náboženství? Důvod je asi dvojitý: náboženská témata jsou naléhavá existenciálně a navíc úrodná teoreticky. Pravdivostní hodnota některých náboženských výpovědí má zjevně dalekosáhlé důsledky; jde v nich o počátky vesmíru, smysl života, posmrtnou existenci a základy veškeré etiky. Náboženská témata jsou dále polem aplikace nových idejí i jejich vzniku. Na straně jedné analytická filosofie náboženství vstřebává podněty z ontologie, teorie poznání, logiky, etiky, filosofie jazyka, filosofie vědy, kognitivní vědy, sociálních věd, historie, práva, přírodních věd a matematiky. Na straně druhé analytická filosofie náboženství zrodila specifické přístupy k identitě objektů, k jednoduchosti hypotéz, k abstraktním objektům, k modálním pojmům, k metaetice, k poznání a racionalitě, k teorii rozhodování a k epistemologii svědectví a nesouhlasu (viz Nagasawa – Wielenberg 2008, Nagasawa 2012, Chandler – Harrison 2012).

Ačkoli analytičtí filosofové náboženství analytické teologii připravili půdu, stále je málo teologů ochotných ji provozovat. (Výjimkami jsou William J. Abraham, David Brown, Vincent Brümmer, Oliver D. Crisp, Bruce D. Marshall, Marilyn McCord Adamsová, Alan G. Padgett, Humphrey Palmer, Richard Sturch a Denys Y. Turner.) Nezájem je však často oboustranný. U analytických filosofů náboženství se projevuje povrchní obeznámeností s klasickou teologií a skoro nulovou znalostí teologie současné.

Na tento stav reaguje čtyřletý mezinárodní projekt *Analytická teologie* (*Analytic Theology: The Convergence of Philosophy and Theology*, 2010 – 2014), financovaný americkou Nadací Johna Templetona, jež přispěla celkovou částkou převyšující 1 mil. USD. Projekt cílí na posun v náboženských otázkách skrze nástroje analytické filosofie. Na tento cíl jsou zaměřeny konference, workshopy, přednášky, stipendia a z nich vzešlé knižní a časopisecké texty. V řešitelských týmech jsou stejným dílem zastoupeni filosofové a teologové. Organizační střediska sídlí na třech vysokých školách a jsou vedena třemi tamními filozofy: Michael C. Rea (Univerzita Notre Dame, USA), Georg Gasser (Univerzita v Innsbrucku, Rakousko) a Yoram Hazony (Shalem Center, Izrael). Každé středisko koordinuje množství zeměpisně blízkých badatelů i týmů složených z filosofů a teologů. A každé si určilo okruh upřednostněných otázek. Tyto okruhy jsou, ve shodě s uvedeným pořadím středisek, následující. 1. Realismus v teologii, boží jednání a prozřetelnost, vykoupení, boží vtělení a lidská tělesnost, povaha a možnost nesmrtnosti, boží zjevení, boží inspiraci náboženských textů, vztah biblických studií a současné analytické filosofie (více na

initiatives/analytic-theology). 2. Boží jednání, osobní identita a zmrtvýchvstání, náboženský pluralismus a racionalita mezináboženského dialogu, vztah lidské svobody a boží milosti a předvědění, vztah stvoření a evoluční biologie, vztah boží spravedlnosti a lidského násilí a utrpení, víra a justifikace náboženského přesvědčení (www.uibk.ac.at/analytic-theology). 3. Filosofické zkoumání hebrejské bible, Talmudu a Mišny (www.bibleandphilosophy.org).

V letech 2012 a 2013 se projektu účastní skupina českých teologů a analytických filosofů, přidružená k rakouskému středisku a vedená Petrem Dvořákem (Filosofický ústav AV ČR, Praha) a Tomášem Machulou (Teologická fakulta Jihočeské univerzity, České Budějovice). Skupina během těchto dvou let pořádá v dvouměsíčních odstupech tvůrčí setkání a přednášky, otevřené odborné veřejnosti a konané střídavě v Praze (Filosofický ústav AV ČR) a Českých Budějovicích (Teologická fakulta Jihočeské univerzity). Tematický okruh skupiny má dvě části. 1. Učení o Trojici ve vztahu k logickým a metafyzickým otázkám identity a univerzálií. 2. Epistemologie víry ve vztahu k logickým a epistemologickým otázkám náboženského jazyka, zázraků, pravděpodobnosti a jistoty. Členové týmu zatím diskutovali epistemologické paralely mezi tzv. tvrdým problémem vědomí ve filosofii mysli a problémem konzistence učení o Trojici, dále analytické a scholastické výklady tohoto učení a problematiku Trojice v současné neanalytické teologii. Přední americký medievalista Kent Emery Jr. z Univerzity Notre Dame u příležitosti svého pobytu v ČR vyložil Bonaventurův pohled na možnosti filosofického poznání nezávislého na zjevení. Příspěvek vyjde ve *Studia Neoaristotelica* pod názvem „Bonaventure (via Alexander) on Cognition *in statu isto*“. Na půli listopadu 2012 proběhlo několik přednášek britského analytika Richarda G. Swinburna. Jeho dílo, měřeno současně šířkou záběru a nároky na přesnost, je dosavadním vrcholem filosofie křesťanství. Swinburne je někdy dokonce považován za nejvýznamnějšího filosofa náboženství ve 20. století (Langtry 2009). V Praze (Filosofický ústav AV ČR, 15. 11. 2012) vyložil své pojetí Trojice, historicko-filosofické důvody pro zmrtvýchvstání Ježíše z Nazareta a své argumenty pro existenci Boha. V Českých Budějovicích (Teologická fakulta Jihočeské univerzity, 16. 11. 2012) se vyjádřil k argumentům ze zla proti existenci Boha. V první půli dubna 2013 Prahu navštíví americký analytický medievalista, metafyzik a filosof náboženství Brian Leftow. Ten, ačkoli je na Oxfordské univerzitě od roku 2002 Swinburnovým nástupcem na pozici profesora filosofie křesťanství, proti němu hájí nečasovost a metafyzickou nutnost Boha a pojetí Trojice vyhraněné proti pojetí Swinburna. Konečně 19. a 20. 9. 2013 se v Praze (v konferenční vile Lanna) uskuteční mezinárodní konference s názvem Analytická teologie: víra, poznání a Trojice. Na ní představí výsledky své práce členové českého týmu i mezinárodně uznávané osobnosti analytické filosofie náboženství a analytické teologie. Účast už přislí-

bil např. čelní americký epistemolog a filosof náboženství Jonathan L. Kvanvig. Konference vyústí do sborníku vybraných příspěvků.

Český tým doufá, že projekt *Analytická teologie* napomůže informovanějšímu povědomí o analytickém přístupu k náboženským otázkám, jemuž domácí filosofové i teologové zatím věnují pozornost spíše mizivou. Zároveň tým zve na workshopy a přednášky, průběžně avizované na stránce www.analytic-theology.cz.

Vlastimil Vohánka

v.vohanka@gmail.com

Literatura

- CRAIG, W. L. (1979): *The Kalam Cosmological Argument*. London: Macmillan & Co. – New York: Barnes & Noble.
- CRAIG, W. L. – MORELAND, J. P. (eds.) (2009): *The Blackwell Companion to Natural Theology*. Oxford: Wiley-Blackwell.
- CRISP, O. D. (2009): *A Reader in Contemporary Philosophical Theology*. New York: T & T Clark.
- CRISP, O. D. – REA, M. C. (eds.) (2009): *Analytic Theology: New Essays in the Philosophy of Theology*. New York: Oxford University Press.
- FLINT, T. P. – REA, M. C. (eds.) (2009): *The Oxford Handbook of Philosophical Theology*. Oxford: Oxford University Press.
- CHANDLER, J. – HARRISON, V. S. (eds.) (2012): *Probability in the Philosophy of Religion*. Oxford: Oxford University Press.
- LANGTRY, B. (2009): Richard Swinburne. In: Oppy, G. – Trakakis, N. (eds.): *The History of Western Philosophy of Religion*. Vol. 5. Durham: Acumen Publishing, Chap. 22.
- LESLIE, J. (1989): *Universes*. London – New York: Routledge.
- MACINTYRE, A. – FLEW, A. (eds.) (1955): *New Essays in Philosophical Theology*. London: SCM.
- MARMODORO, A. – HILL, J. (eds.) (2011): *The Metaphysics of the Incarnation*. New York: Oxford University Press.
- MARTIN, M. (ed.) (2007): *The Cambridge Companion to Atheism*. Cambridge: Cambridge University Press.
- MCCALL, T. – REA, M. C. (eds.) (2009): *Philosophical and Theological Essays on the Trinity*. New York: Oxford University Press.
- MORRIS, T. V. (1986): *The Logic of God Incarnate*. Ithaca – London: Cornell University Press.
- NAGASAWA, Y. (ed.) (2012): *Scientific Approaches to the Philosophy of Religion*. New York: Palgrave Macmillan.
- NAGASAWA, Y. – WIELENBERG, E. J. (eds.) (2008): *New Ways in Philosophy of Religion*. New York: Palgrave Macmillan.

- PLANTINGA, A. (1967): *God and Other Minds: A Study of the Rational Justification of Belief in God*. Ithaca – London: Cornell University Press.
- PLANTINGA, A. (1984): Advice to Christian Philosophers. *Faith and Philosophy* 1, 253-271.
- POUVET, R. (2011): On the Polish Roots of the Analytic Philosophy of Religion. *European Journal for Philosophy of Religion* 3, 1-20.
- QUINN, P. L. – TALIAFERRO, C. (eds.) (1999): *A Companion to Philosophy of Religion*. Oxford: Blackwell.
- REA, M. C. (ed.) (2009): *Oxford Readings in Philosophical Theology, Vol. 1: Trinity, Incarnation, and Atonement; Vol. 2: Providence, Scripture, and Resurrection*. New York: Oxford University Press.
- SMITH, Q. (2001): The Metaphilosophy of Naturalism. *Philo* 4, 195-215.
- SWINBURNE, R. (1977): *The Coherence of Theism*. Oxford: Clarendon Press.
- SWINBURNE, R. (1979): *The Existence of God*. Oxford: Clarendon Press.
- TALIAFERRO, C. – MEISTER, C. (eds.) (2009): *The Cambridge Companion to Christian Philosophical Theology*. Cambridge: Cambridge University Press.

////////////////////
////////////////////
//////////////////// **TEORIE VĚDY / THEORY OF SCIENCE**
časopis pro mezioborová zkoumání vědy / journal for interdisciplinary studies of science

TEORIE VĚDY se zaměřuje na zkoumání filosofických a metodologických základů vědeckého poznání, sleduje vzájemné vztahy vědy, techniky a společnosti, problematiku historického vývoje vědy a vědění a interdisciplinární vztahy mezi různými obory. TEORIE VĚDY je recenzovaný vědecký časopis zařazený do několika prestižních akademických databází (ERIH, CEJSH). Vychází čtvrtletně tiskem i online, publikuje odborné studie, komentované překlady, recenzní stati a zprávy z výzkumů v češtině a angličtině. TEORIE VĚDY je vydávána Kabinetem pro studium vědy, techniky a společnosti při Filosofickém ústavu Akademie věd ČR, v.v.i.

THEORY OF SCIENCE focuses on the inquiry into philosophical and methodological principles of scientific knowledge. It traces the interrelationship of science, technology, and society; the problems of the historical development of science and knowledge; and the interdisciplinary relations across and within Humanities, Social, Natural, and Life Sciences. THEORY OF SCIENCE is a peer-reviewed academic journal listed currently in several prestigious databases (ERIH, CEJSH). It publishes quarterly, in print and online, scholarly articles, annotated translations, reviews, and research reports in Czech and English. THEORY OF SCIENCE is published by the Centre for Science, Technology, and Society Studies at the Institute of Philosophy of the Academy of Sciences of the Czech Republic.

redakční rada / editorial board /// Jaroslav Anděl / Gerhard Banse / Geoffrey Batchen / Gil Eyal / Jiří Fiala / Sean Field / Adolf Filáček / Miloš Havelka / John Holmwood / Daniel Just / Hans-Herbert Koegler / William Outhwaite / Ladislav Tondl

adresa / address /// TEORIE VĚDY / THEORY OF SCIENCE /// Institute of Philosophy / Academy of Sciences of the Czech Republic / Jilská 1 / 110 00 Praha 1 / Czech Republic
tel. +420/222 220.107
fax +420/222 220 725
e-mail teorievedy@ilu.cas.cz
url <http://teorievedy.flu.cas.cz>

ISSN 1210-0250 (Print)
ISSN 1804-6347 (Online)
MK ČR E 18677

aktuální výzvy k tematickým sekcím jsou dostupné na stránkách <http://teorievedy.flu.cas.cz>

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