Contents ▪ Obsah

ARTICLES ▪ STATE

DISCUSSIONS ▪ DISKUSIE

BOOK REVIEWS ▪ RECENZIE

MISCELLANEA ▪ MISCELANEÁ

Organon F 22 (2) 2015: 143-162

Inference to the Best Explanation and Disjunctive Explanations

JAEHO LEE

Chung-Ang University. Department of Philosophy
84 Heukseok-ro. Donjak-gu. Seoul 156-756. Korea
jaeho.jaeho@gmail.com

RECEIVED: 30-07-2014 ▪ ACCEPTED: 17-12-2014

ABSTRACT: In this paper I will examine Helen Beebee's argument that anti-Humeans are not in a better position to justify induction. I will first argue that her argument proves too much and that it can jeopardize the status of inference to the best explanation as a useful inductive principle. I will then propose a principle that should govern our use of disjunctive explanations in the context of inference to the best explanation and show that Beebee's use of disjunctive explanations violates this principle.

KEYWORDS: David Armstrong – Helen Beebee – Disjunctive explanation – Inference to the best explanation.

1. Explanationist approaches to the problem of induction and Armstrong's argument

By 'explanationist approaches to the problem of induction' I mean the attempts to justify inductive generalizations with inference to the best explanation (hereafter IBE). The overall strategy is as follows.

(1) All observed Fs are Gs.

(2) This observed regularity cries out for an explanation.

(3) The best explanation *or* a consequence of the best explanation of this evidence is the corresponding general regularity that all Fs are Gs.

(4) By IBE, we are justified to infer 'all Fs are Gs' from 'all observed Fs are Gs.'

Note that (3) is disjunctive, meaning that there are two ways for explanationists to go. Some explanationists think that the best explanation of 'all observed Fs are Gs' is that 'all Fs are Gs' (cf. Harman 1980). I will call them type-A explanationists. Other explanationists think that 'all Fs are Gs' is not the best explanation but a mere consequence of the best explanation (cf. Armstrong 1983; BonJour 1998; Foster 1983; Peacocke 2004). I will call them type-B explanationists.

David Armstrong is a type-B explanationist. He thinks that we can solve the problem of induction in the following way (see Armstrong 1983, 52-53):

(1) All observed Fs are Gs.

(2) This observed regularity cries out for an explanation.

(5) The best explanation of this evidence is N(F,G)

(6) By IBE, we are justified to infer N(F, G) from 'all observed Fs are Gs.'

(7) N(F, G) entails 'all Fs are Gs.'

(8) Therefore, we are justified to infer 'all Fs are Gs' from 'all observed Fs are Gs.'

According to Armstrong, N(F, G) is a necessitation relation and is wholly distinct from 'all Fs are Gs' even though the former entails the latter. Since Armstrong introduces a necessary connection between wholly distinct states of affairs (namely between N(F, G) and 'all Fs are Gs') his view is anti-Humean.

Armstrong's argument has two controversial steps. First, (6) is controversial because IBE itself is extremely controversial. Even though there are many philosophers who are suspicious of IBE, Armstrong does not provide a systemic justification of IBE.[1] So his argument is at best conditional; *if*

[1] For criticisms of IBE, see Salmon (2001), van Fraassen (1989). In fact, Armstrong (1983, 59) provides an argument for the rationality of IBE which I find neither systematic nor plausible.

IBE is justified then, unlike the Humean, the anti-Humean can solve the problem of induction. Second, and more importantly, (5) is controversial. Why should we think that N(F, G) is the best explanation of 'all observed Fs are Gs'? Suppose that 'all Fs are Gs' *can* explain 'all observed Fs are Gs'. Then, as type-A explanationists claim, N(F, G) *cannot* be the best explanation. Most philosophers agree that simplicity is an explanatory virtue. Since 'all Fs are Gs' is an ontologically simpler explanation than N(F, G), all other things being equal, we should prefer 'all Fs are Gs' to N(F, G).

Here Armstrong does provide an argument. His basic idea is that a conjunction cannot explain its conjunct(s) – call this 'Armstrong's principle'. We have a strong intuition that a state of affairs cannot explain itself and Armstrong's principle seems to be a natural consequence of this intuition. Once we accept Armstrong's principle, quite trivially we should conclude that 'all Fs are Gs' cannot explain 'all observed Fs are Gs' because 'all Fs are Gs' is logically equivalent to 'all observed Fs are Gs and all unobserved Fs are Gs' (see Armstrong 1983, 40; for a similar argument see Bird 2007, 86-90).

Most critics of Armstrong's argument have focused on Armstrong's principle and tried to prove that it is unjustifiable. Some Humeans argue that this principle simply begs the question against influential theories of explanation, such as the D-N model and unification account. Others, such as Rodger White, claim that Armstrong fails to recognize the important distinction between instance explanation and regularity explanation.[2] I examined these criticisms closely elsewhere and don't want to further discuss them in this paper (see Lee 2013a; 2013b). There is another criticism raised by Helen Beebee in her recent paper (cf. Beebee 2011). She does not focus on Armstrong's principle. Instead, she claims that N(F, G) is not the best *anti-Humean* explanation. So her argument does not lose its credibility even if Armstrong's principle turns out to be true.

In the rest of this paper, I will examine her argument. First, I will explain Beebee's argument in the next section. In section 3, I will show that her argument proves much more than she thinks, which suggests that there must be something wrong with her argument. In section 4, I will argue that Beebee's argument is based on a misconception regarding IBE. In

[2] For the former type of criticisms, see Lewis (1994, 478-479); Loewer (1996, 113). For White's criticism, see White (2005).

the last section, I will provide another argument against Beebee, which is closely related to the argument in section 4.

2. Beebee's argument

Beebee does not deny that N(F, G) can explain 'all Fs are Gs'. Instead she argues that N(F, G) is not the best anti-Humean explanation. She argues in Beebee (2011, 510) that the following (anti-Humean) explanation is at least as good as N(F, G).

(SF) F and G have been necessarily connected so far.[3]

Since (SF) does not entail that 'all Fs are Gs', if (SF) is at least as good as N(F, G), Armstrong's type B approach is hopeless even if Armstrong's principle is true.

One thing we should note is that (SF) is quite different from a time-limited necessitation relation, such as $N_{\text{until-2014}}$(F, G). Beebee acknowledges that N(F, G) is a better explanation than $N_{\text{until-2014}}$(F, G) because only the latter has a temporal parameter. Again, simplicity is an explanatory virtue and additional parameters decrease the degree of simplicity. For this reason, Beebee emphasizes that (SF) does not have temporal-parameter unlike $N_{\text{until-2014}}$(F, G). Based on this fact, she argues that there is no reason to think N(F, G) has the advantage of simplicity over (SF).

Another important issue is whether or not predictive power is an explanatory virtue. (SF) is a disjunctive hypothesis because it is *contextually equivalent* to ($N_{\text{until-2014}}$(F, G) or $N_{\text{until-2015}}$(F, G) or $N_{\text{until-2016}}$(F, G) or $N_{\text{until-2017}}$(F, G) or ... or N(F, G)).[4] Even though most disjuncts of this disjunction have predictive power, there is a disjunct which does *not* have predictive power, namely $N_{\text{until-2014}}$(F, G).[5] Just one disjunct is enough to re-

[3] 'SF' stands for 'so far'.

[4] Two comments on this claim are in need. First, the notion of 'contextual equivalence,' and hence that of 'disjunctive hypothesis' (or 'disjunctive explanation') needs to be clarified. I will address this issue in section 4. Second, obviously (SF) is not contextually equivalent to this disjunction. We need much more fine grained disjunction. I believe, however, that my readers will easily catch what I intend here.

[5] For the sake of argument, let me assume that we are at the last moment of 2014.

move the predictive power of the whole disjunctive hypothesis. Since N(F, G) has predictive power, if predictive power is an explanatory virtue, N(F, G) is a better explanation compared to (SF).

Beebee claims that in the current context predictive power is not an explanatory virtue. First, she argues that in our context we are talking about a *metaphysical* explanation rather than a *scientific* explanation. She says: "Prediction is not part of the point of metaphysics, in either a practical or a theoretical sense" (Beebee 2011, 517). Moreover, since we are talking about the problem of induction, Armstrong's opponent is the inductive skeptic. And "the inductive skeptic holds that, pending a good argument to the contrary, a hypothesis that makes predictions is *eo ipso* a hypothesis that we have no grounds for believing". In short, the simplicity criterion does not discriminate between (SF) and N(F, G), and the predictive power criterion is not applicable in our context. So there is no reason to think N(F, G) is a better explanation than (SF), which means that (5) in Armstrong's argument is not justified.

3. What does Beebee's argument, if good, show?

For the sake of argument, let me assume that Beebee's argument is a good one. Then, as Beebee claims, it shows that Armstrong's type-B approach is hopeless. It also shows many other things, however. First, it shows that the type-A approach is hopeless as well. Suppose that Armstrong's principle is not true. Now type-A explanationists will claim, ignoring Armstrong's argument, that 'all Fs are Gs' is the best explanation of 'all observed Fs are Gs', but compare this explanation with the following Beebee-style hypothesis.

(SF') All Fs have been Gs so far.

Since (SF') has no temporal parameter, the simplicity criterion does not discriminate between (SF') and 'all Fs are Gs'. Because we are talking about metaphysical explanations, the predictive power criterion is irrelevant. Therefore, there is no reason to think 'all Fs are Gs' is a better explanation compared to (SF').

At this point, one might claim that there is an important difference between these two hypotheses. The worry goes like this. (SF') is a disjunctive hypothesis because it is contextually equivalent to ('all Fs are Gs until 2013' or 'all Fs are Gs until 2014' or, ..., or 'all Fs are Gs'). Most disjuncts of this disjunctive explanation have temporal parameters and time-limited regularities are not genuine law-like regularities. Only law-like regularities have explanatory power, so (SF') is not considered to be a genuine explanation. I think that this worry is groundless. First, there is no reason to think that only law-like regularities have explanatory power. For example, we can explain why all objects I picked out of this barrel are green with the fact that all objects in this barrel are green.[6] However, 'all objects in this barrel are green' is not a law-like regularity. Second, even if we accept that only law-like regularities can have explanatory power, it does not make much difference. As far as I know, the most powerful Humean theory of law is the Mill-Ramsey-Lewis theory of law, according to which laws are axioms and high-level theorems of best axiomatic deductive system. Imagine a possible world where everything is exactly the same as our world until 2014, and then there is no regularity what so ever after 2014. The Humean must accept this possibility. And the best axiomatic system of this world will contain the following time-limited laws: Boyle's law$_{\text{until 2014}}$, Charles's law$_{\text{until 2014,}}$ and etc. Just as Armstrong must accept the *possibility* of time-limited necessitation relation, the Humean must accept the *possibility* of time-limited law.[7] Once we realize that even the Humean must accept this possibility, we can dodge this criticism by converting (SF') to the following (SF").

(SF") All Fs have been Gs so far as a nomological fact.

Here, 'nomological facts' means facts that hold as a consequence of law(s) of nature. Since the Humean cannot exclude the possibility of 'all Fs are Gs until 2014"s being a law, (SF") does not entail 'all Fs are Gs'.

In short, Beebee's argument, if good, undermines the explanationist approach to the problem of induction in general.[8] What makes things worse

[6] For a similar idea, see White (2005, 12).

[7] Beebee nicely explains why Armstrong must accept the possibility of time limited necessitation; see Beebee (2011, 511-513).

[8] I am not saying that this consequence raises a problem to Beebee. After all, Beebee seems to think that both Humeans and anti-Humeans cannot justify induction. My

is that her argument undermines scientific realism too. Let me assume, following most scientific realists, that IBE is the basic inductive principle for scientific inquiries. So, we justify the existence of such an unobservable entity as atoms with IBE in the following way.

(9) There are Brownian movements and other observable evidences.

(10) The best explanation of this empirical evidence is the existence of atoms.

(11) By IBE, we are justified to believe that there are atoms.

However, (10) is questionable, or so Beebee should think. Compare the explanation, which postulates the existence of atoms – call it '(EA)' – with the following hypothesis '(UM)' for 'unobservable mechanism').

(UM) There is some unobservable mechanism, which produces the observable consequences atoms are supposed to produce.

There is no reason to think (EA) is simpler than (UM). (EA) introduces one additional kind of unobservable entity, namely atoms. (UM) does not exclude the possibility of more than one additional unobservable entity; let's call this hypothesis (TWO). This does not mean that (UM) has commitment to (TWO). (UM) is a disjunctive hypothesis. Just as (SF) has $N_{\text{until-2014}}$(F, G) as its disjunct, (UM) has (TWO) as its disjunct. Just as it is not the case that (SF) is less simple compared to N(F, G) simply because $N_{\text{until-2014}}$(F, G) is (SF)'s disjunct, it is not the case that (UM) is less simple compared to (EA) simply because (TWO) is (UM)'s disjunct. Therefore, as far as the simplicity-criterion is concerned, there is no reason to discriminate between (UM) and (EA).

What about predictive power? Here we are talking about scientific explanation. So predictive power may be an explanatory virtue in our context but this fact does not make a difference. By hypothesis, (UM) has the exact same observable consequences as (EA). Therefore, predictive power criterion does not discriminate between them either.

The problem regarding (10) can be generalized. Whenever scientists introduce a new theoretical entity X to explain the empirical evidences, we

point is that Beebee's argument, if good, undermines IBE itself. This consequence is just the first step toward my point.

can make an alternative explanation that has the following disjunctive form: there is some unobservable mechanism, which has the same observable consequences as X. This alternative explanation will always block the sort of IBE scientists want to use. This means, if Beebee's argument is good, IBE is useless for the justification of scientific realism.

In fact, this line of argument can be further generalized. Imagine that someone claims that H is the best explanation of our evidence E, then there must be less simple hypotheses which have exactly the same empirical consequences as H. Make a disjunctive hypothesis which has H and those less simple hypotheses as its disjuncts, then argue that this disjunctive hypothesis is at least as good as H. In short, if Beebee's argument is good, then it undermines IBE itself. This means that Beebee's position is not an internally coherent one because she does not question the rationality of IBE.[9] There must be something wrong in Beebee's argument.

4. IBE and disjunctive explanations

One lesson we can learn from the discussion from the previous section seems to be this: when we make IBE, use of disjunctive explanations should be restricted. In this section, I will propose a principle that should govern our use of disjunctive explanations in the context of IBE and defend it.

Let me first define 'disjunctive explanation' and 'disjuncts of a disjunctive explanation.' An explanation (i.e. explanans) will be called a disjunctive explanation in this paper iff it is *contextually equivalent* to a disjunction. And by "disjuncts of a disjunctive explanation" I will mean those disjuncts of a disjunction to which the disjunctive explanation is contextually equivalent. A and B are contextually equivalent iff 'A iff B' is true in all worlds whose possibilities are considered seriously under the context of debate. So if all logical possibilities are seriously considered under the context, contextual equivalence becomes nothing but logical equivalence. Likewise, if all and only metaphysically (or physically) possible worlds are considered under the context, A and B are contextually equivalent iff 'A iff B' is metaphysically (or physically) necessary. So,

[9] One might think that my argument does not undermine the rationality of IBE but only shows that it is not very interesting. I think, however, that undermining *interesting* IBE is virtually the same as undermining IBE itself.

(12) There is some amount of water in this cup

and

(13) 'There is (exactly) one H_2O molecule in this cup' or 'there are (exactly) two H_2O molecules in this cup' or 'there is (exactly) three H_2O molecules in this cup', or .....

are contextually equivalent under the context in which only metaphysically possible worlds are seriously considered because 'water is H_2O' is metaphysically necessary.

Likewise,

(14) An object is moving

and

(15) 'An object is moving at the speed of 1 m/sec' or 'an object is moving at the speed of 2m/sec' or .... or 'an object is moving at the speed of 299,792,452 m/sec.'

are contextually equivalent under the context in which only physically possible worlds are seriously considered because the speed of light in a vacuum is 299,792,452 m/sec in our world and it is a law that no object can move faster than the speed of light in a vacuum.[10]

My notion of 'contextual equivalence' could be weaker than physical equivalence. Imagine that we are examining an explanatory hypothesis, H, and that under the current context there are only three potential truth-makers of H, namely T_1, T_2, T_3. If there are other physically possible potential truth-makers of H, which can be excluded under the current context, H is contextually equivalent to 'T_1 V T_2 V T_3' even though they are not physically equivalent.[11]

[10] Again, strictly speaking, (14) is not physically equivalent to (15) because (15) is not sufficiently fine grained.

[11] This does not mean that contextual equivalence is necessarily weaker than physical equivalence. As I pointed out, under some context contextual equivalence is nothing but logical equivalence which is much stronger than physical equivalence.

According to my definition, all explanations are quite trivially disjunctive explanations. Let A be our explanation. This is a disjunctive explanation because A is logically equivalent and hence contextually equivalent to (A&B or A&~B). This triviality, however, will turn out to be harmless for the reason I will explain later.

At this point, I must concede that my notion of 'disjunctive explanation' does not fit the ordinary usage of the term. Usually we don't think of N(F, G) as a disjunctive hypothesis simply because it is logically equivalent to a disjunction, say '(N(F, G)&P) ∨ (N(F, G)&~P)'. To the contrary, it is quite natural to think that N(F, G) is a *non*-disjunctive hypothesis because Armstrong thinks that N relation is a second-order universal and believes there is no such thing as disjunctive universals (see Armstron 1978, 19-22). For this reason, one might think that it would be better to replace 'disjunctive explanation' with 'multiply truth makeable explanation' and 'disjunct' with 'potential truth-maker' (I am pretty sure that Beebee would prefer these terms to my terms).[12] For those who are comfortable with truth making talk, I believe that this change of terminology is quite harmless. Main arguments in this paper which use 'disjunctive explanation' and 'disjunct' can be easily converted with minor adjustments to the arguments which use 'multiply truth-makeable explanations' and 'potential truth maker.' After all, it is quite obvious that most examples of 'disjunctive explanations' in this paper are multiply truth makeable explanations. For example, since both N(F, G) and $N_{\text{until-2014}}$(F, G) are potential truth maker of (SF), it is a multiply truth-makeable hypothesis.

Even though I accept that the use of 'disjunctive explanation' could produce some confusion among readers, I don't want to use 'multiply truth makeable explanation' because I think neither 'truth maker principle' nor 'truth making relation' are well-understood concepts. I simply don't want to be involved in truth-making talk. Unlike 'multiply truth makeable explanation,' my notion of 'disjunctive explanation' is quite clear as long as we remember the definition of this term.

Since every explanation is a disjunctive explanation, it *trivially* follows that sometimes we can use disjunctive explanations in the context of IBE. However, there are some non-trivial cases in which we can use disjunctive explanations in the context of IBE. Consider the following explanation.

[12] I thank Sung-il Han for this point.

> (Lethal-Dose) A man drank a cup of some kind of liquid. After some time, he died showing typical toxic symptoms of potassium cyanide (KCN). Why did the man die? My explanation: the liquid he drank contained more than the lethal dose of KCN.

The lethal dose for KCN is 200 – 300 mg. For the sake of simplicity, however, let me assume that the lethal dose is 300 mg. My explanation is disjunctive because it is (physically) equivalent to 'the liquid contains 300 mg of KCN' or 'it contains 301 mg of KCN' or 'it contains 302 mg of KCN' or, …'.[13] This disjunctive feature of my explanation is no problem. There is no reason to think that the inference I made in the example is a bad one. Here is another example.

> (Informer) Four bad guys (Adam, Bill, Curt, Dan) conspired to assassinate the president. However, the plot failed because the presidential guards knew the conspiracy. Why did the plot fail? My explanation: at least one of those four guys was a rat.

My explanation is a disjunctive explanation because it is (logically) equivalent to ('Adam was a rat' or 'Bill was a rat' or 'Curt was a rat' or 'Dan was a rat'). Again, this disjunctive feature of my explanation is not a problem. Obviously, we can use this explanation in the context of IBE.

So we need a principled way to restrict our use of disjunctive explanations in the context of IBE. The principle should not be too strict because it should not make (Lethal-Dose) or (Informer) disqualified for IBE-triggering explanations. It should also not be too lenient because, as we saw in the previous section, IBE itself can be undermined if we are allowed to use disjunctive explanations freely.

The principle I propose is very simple: A disjunctive explanation is justified as a complete IBE-triggering explanation only when it is a permissible disjunctive explanation which is better than any other permissible disjunctive explanation that is not its disjunct. When is a disjunctive explanation permissible? There are three principles of permissibility.

[13] I am assuming that in the context of (Lethal-Dose), we are considering only physically possible hypotheses.

(P1) A disjunctive explanation is permissible as a complete IBE-triggering explanation when none of its disjuncts is a genuine explanation.

(P2) A disjunctive explanation is permissible as a complete IBE-triggering explanation when its disjuncts are all significantly worse than it.

(P3) A disjunctive explanation is permissible as a complete IBE-triggering explanation when none of its disjuncts is explanatorily salient.

Some comments and explications are in need. First, "permissible" has a very weak sense here. It is not the case that if a disjunctive explanation satisfies at least one of P1 – P3, then we are justified to accept the disjunctive explanation. There might be many permissible disjunctive explanations and in that case we must choose the best out of them. So P1 – P3 should be read in the following way:

A disjunctive explanation is *preferable to its disjuncts* when...

Once we read "permissible" in this way, it is almost self-explanatory that P1 – P3 are justified. An explanation is always better than no explanation, so P1 is justified. A significantly better explanation is always better than a significantly worse explanation, so P2 is justified as well. If two explanations are virtually equal in the explanatory sense, to regard one of them as an IBE triggering explanation is not justified, so P3 is justified. Since P1 – P3 are all self-explanatory, the only question concerning them is whether or not P1 – P3 are exhaustive. Imagine that a disjunctive explanation satisfies none of P1 – P3. Then it would have an explanatorily salient disjunct which is at least as good as the disjunctive explanation. Can this explanation be an IBE-triggering explanation? I don't think so. Remember that IBE is inference to the *best* explanation. If the disjunct is at least as good as the disjunction, the disjunction is not the best explanation. To be sure, according to P3, we can regard some non-best explanations as IBE-triggering explanations,[14] but it is because this is inevitable. We have an independ-

[14] This is the reason why my principle contains "it is a permissible disjunctive explanation which is better than any other permissible disjunctive explanation which is not its

ently justifiable principle that if two explanations are equally good, we should not discriminate them in the context of IBE. Unless we find another principle, which can do a similar job, we have good reason to think P1-P3 is exhaustive.

Second, strictly speaking, P1 is redundant. If all its disjuncts are no explanation, they are all worse than the disjunctive explanation. So P1 is an instance of P2. P1 is also an instance of P3 because if all its disjuncts are no explanation, there is no disjunct that is explanatorily salient.

Third, why do we need to insert 'as a complete IBE-triggering explanation?' Consider the following example.

> (Short-Circuit) There was a fire last night. The investigators found a typical pattern of soot that is often caused by a short circuit in the fuse box. Why did the fire occur? My explanation: There was a short circuit in the fuse box.

(Short-Circuit) is a disjunctive explanation because it is logically equivalent to 'there was a short circuit in the fuse box and there was sufficient amount of oxygen in the air' or 'there was a short circuit in the fuse box and there was not sufficient amount of oxygen in the air', but this disjunctive explanation is not permissible if we remove 'as a complete IBE-triggering explanation' from P1 – P3. First, the first disjunct is a good explanation, so it does not satisfy P1. Second, the first disjunct is a *better* explanation than the disjunctive explanation itself because the former is relatively close to the complete explanation compared to the latter, so it does not satisfy P2. Lastly, the first disjunct is explanatorily salient because the second disjunct is no explanation, so it does not satisfy P3. These results are unacceptable because intuitively the IBE I used in this example seems to be a good one. If we insert 'as a complete IBE-triggering explanation' in P1 – P3, however, we can handle this problem. In fact, my explanation should not be allowed as a *complete* IBE-triggering explanation because it is at best a partial explanation. This does not mean that my explanation should not be allowed as IBE-triggering explanation. My (partial) explanation seems to come from a perfectly permissible *complete* IBE-triggering explanation.

disjunct" rather than a simpler expression "it is the best permissible disjunctive explanation."

Fourth, given my definition of 'disjunctive explanation,' virtually all explanations are disjunctive in infinitely many senses. A is equivalent to (A&B or A&~B) and it is equivalent to (A&C or A&~C), and so on. So, my principle should mean this: a disjunctive explanation is permissible as a complete IBE-triggering explanation when all of their possible "disjunctification" satisfies at least one of P1 – P3.

I said that my definition of 'disjunctive explanation' makes all explanation disjunctive in a trivial sense and that this result is harmless. Here is why. Suppose that a hypothesis, H, is a potential explanation of evidence E. Furthermore, suppose that A is an arbitrarily chosen explanatorily irrelevant factor. We can trivially "disjunctify" H using A because H is logically equivalent to (H&A or H&~A). Since A is explanatorily irrelevant to E, ~A is irrelevant to E as well. As familiar counterexamples against D-N model show, irrelevancy is fatal to explanation.[15] So, both H&A and H&~A are not explanations, which means that (H&A or H&~A) satisfies P1. In short, we don't have to worry about the possibility of trivial disjunctifications via explanatorily irrelevant factors.

Now let me explain why my principle is neither too strict nor too lenient. First, let me explain why it is not too strict. Consider (Lethal-Dose). It is a disjunctive explanation because it is contextually equivalent to "the liquid contains 300 mg of KCN' or 'it contains 301 mg of KCN' or 'it contains 302 mg of KCN' or, ... '. I think (Lethal-Dose) satisfies P1 although it could be slightly controversial. Even if it does not satisfy P1, there is no question that it satisfies P2 and P3, which are weaker than P1. Suppose that the liquid the man drank contained exactly 327 mg of KCN. Can we say that the man died because the liquid he drank contained exactly 327 mg? I don't think so. One reliable test for the existence of explanatory relation is to see whether there is a counterfactual dependence between the alleged explanans and the explanandum. This test is particularly reliable when it is not applied to laws and when there is no worry of backup cause situations, such as preemption, trumping, and overdetermination. Is it true that if the liquid he consumed had not contained exactly 327 mg of KCN, he would not have died? The closest possible world in which the liquid does not contain exactly 327 mg of KCN

[15] Probably the most famous counterexample of this category would be the hexed-salt example (originally) by H. Kyburg; see Salmon (1989, 50).

would be the world where it contains, say, 326 or 328 mg of KCN and in this world the man would have died because the lethal dose of KCN is 300 mg. This result can be generalized so that all disjuncts of (Lethal-Dose) turn out to be no explanations. Unlike its disjuncts, (Lethal-Dose) passes the counterfactual dependence test. It is true that if the liquid the man drank had not contained 300 mg or more of KCN, he would not have died. The closest possible world in which the liquid does not contain 300 mg or more of KCN would be the world where it contains 299 mg of KCN and in this world he would not have died. So, (Lethal-Dose) satisfies P1 and it is a permissible disjunctive explanation. I believe that the counterfactual dependence test I used is reliable in this case. It is not applied to laws and there is no worry of a backup cause situation. Even if it is not reliable in our context, I am sure that (Lethal-Dose) satisfies at least P2 and P3. First, compare these two explanations: “The man died because the liquid contained exactly 324 mg of KCN” vs. “The man died because the liquid contained more than lethal dose of KCN”. There is no question that the second explanation is much better than the first. Second, compare “The man died because the liquid contained exactly 324 mg of KCN” with “The man died because the liquid contained exactly 325 mg of KCN”. There is no reason to think one of them is explanatorily salient, which means that (Lethal-Dose) satisfies P3.

The following example of a red ball by Beebee (2011, 515) is similar to my (Lethal-Dose) example:

> (Red Ball) There are twenty balls in a bag, all of which (unknown to me) are different shades of red. You pull a ball from the bag, and you want to know why you pulled out a red ball. My answer: all the balls are red.

As Beebee emphasizes, (Red Ball) is a multiply truth makeable explanation and hence it is a disjunctive explanation in my sense because it is contextually equivalent to “all the balls are a shade$_1$ of red’ or ‘all the balls are a shade$_2$ of red’ or … or ‘one of the balls is a shade$_1$ of red and others are all a shade$_2$ of red’ or ‘one of the balls is a shade$_1$ of red and others are all a shade$_3$ of red’ or…’. However, as Beebee emphasizes, (Red Ball) is a permissible disjunctive explanation because it satisfies P1. For example, ‘all the balls are a shade$_1$ of red’ is not a potential explanation of why the ball I pulled is a red ball. Again, counterfactual dependence test is helpful here for it is not true that if it had not been the case that all the balls are shade$_1$ of red, then the ball I pulled would not have been a red ball. Probably in the closest world in which the antecedent is true, some balls would be different shades of red and in that world the ball I pulled would be a red ball.[16]

Next, consider (Informer). Unlike (Lethal-Dose), (Informer)'s disjuncts are good potential explanations. Suppose that Adam was the rat. Then it is true that the plot failed because Adam was a rat. In fact, this explanation seems to be better than (Informer) because it is more informative. Therefore, (Informer) does not satisfy P1 and it does not satisfy P2 either, but (Informer) satisfies P3. Compare 'Adam was the rat' with 'Bill was the rat.' Explanatorily speaking, they are perfectly symmetric. There is no reason to think one is better than the other, which means that there is no explanatorily salient disjunct here. So (Informer) is a permissible disjunctive explanation.

Now, let me explain why my principle is not too lenient. My principle does not have the consequences that Beebee's argument has. Consider first (UM). For the sake of argument, let me assume that we are considering seriously all metaphysical possibilities. (UM) is contextually equivalent to "There are atoms' or 'there are *shatoms,* which are different from atoms but produces the same observable consequences atoms are supposed to produce' or 'there is a Cartesian demon who produces Brownian movement and other alleged evidence for atoms' or…'. (UM) cannot satisfy P1. (EA), or 'There are atoms' is a good potential explanation. (UM) cannot satisfy P2 either. There is no reason to think (UM) is much better than (EA). To the contrary, (EA) seems to be better than (UM). (EA) is much more informative than (UM). Lastly, (UM) also cannot satisfy P3. (EA) is significantly salient in the explanatory sense, and that is why scientists believe in atoms.[17] Since (UM) satisfies none of P1 – P3, (UM) is not a permissible

[16] Again, even if (Red Ball) does not satisfy P1, quite obviously it does satisfy at least one of P2 and P3.

[17] One might think we cannot know that (EA) is salient. (UM) is a disjunctive explanation, which has in principle infinitely many disjuncts. We human beings cannot examine those infinitely many disjuncts. In fact, we cannot even know those infinitely many disjuncts! The idea behind this criticism is same as the idea behind the "argument from bad lot" by van Fraassen. For the argument see van Fraassen (1989, 142-143). For Psillos' criticism of this argument, see Psillos (1999, 220). All I want to say here is that

disjunctive explanation. Unlike (UM), there is no reason to think that (EA) is not a permissible disjunctive explanation. So, under the assumption that (EA) is the best permissible disjunctive explanation, we are justified to believe in atoms. My principle allows the defenders of IBE to support scientific realism.

It is quite clear by now that I can reject Beebee's argument with my principle. In short, Beebee's (SF) is not a permissible disjunctive explanation. (SF) is (probably metaphysically) equivalent to the following explanation: ($N_{\text{until-2014}}$(F, G) or $N_{\text{until-2015}}$(F, G) or $N_{\text{until-2016}}$(F, G) or $N_{\text{until-2017}}$(F, G) or ... or N(F, G)). Therefore, (SF) cannot satisfy P1. N(F, G) is a perfectly good potential explanation and even Beebee does not deny this. (SF) cannot satisfy P2 either. There is no reason to think (SF) is significantly better than N(F, G). To the contrary, N(F, G) seems to be better than (SF) because it is more informative. Lastly, and most importantly, (SF) also cannot satisfy P3. Compare N(F, G) with $N_{\text{until-2014}}$(F, G). Beebee herself concedes that the first explanation is better than the second one because it, unlike the second one, has no temporal parameters. The question is this: how *significantly* better is N(F, G) compared to $N_{\text{until-2014}}$(F, G)? Since saliency is a vague concept, if we want to assert that N(F, G) is the salient disjunct we need to show that N(F, G) is not just better but significantly better than $N_{\text{until-2014}}$(F, G). I believe that N(F, G) is significantly better than $N_{\text{until-2014}}$(F, G). $N_{\text{until-2014}}$(F, G) requires us to radically revise our conception concerning space and time. We don't think that a particular time or space has causal/explanatory power. We do think a particular *length* of time or space can have causal/explanatory power. For example, we can mention a particular half-life to explain radioactive decay, but to say that a particular *length* of time has causal/explanatory power is one thing and to say that such a particular time as 2014 has causal/explanatory power is another. If we accept $N_{\text{until-2014}}$(F, G), we should attribute some kind of causal/explanatory power to a particular time, namely 2014. I am not saying that this is unintelligible; what I am saying is that this is a radical revision of our belief system. Other things being equal, an explanation that does not require such a radical revision is much better

even if this worry is a genuine worry, this is a criticism of IBE itself. So I don't have to have an answer to this criticism. We (including Beebee) are assuming that IBE is justifiable.

than an explanation that does. So, there is good reason to think N(F, G) is significantly better than $N_{until-2014}$(F, G), which means that (SF) cannot satisfy P3. Since (SF) satisfies none of P1 – P3, it is not a permissible disjunctive explanation. Since there is no reason to think that N(F, G) is not permissible, under the assumption that N(F, G) is the best permissible disjunctive explanation, we are justified to believe in N(F, G).

5. Informativeness as an explanatory virtue

In this last section, I will argue that there is an additional reason to think Beebee's argument fails, which was in fact implicitly suggested in the previous section. Why should we restrict our use of disjunctive explanations in the context of IBE? My explanation in section 3 was that unrestricted use of disjunctive explanations tends to undermine IBE itself because it makes IBE uninteresting. Compare (EA) and (UM) once again. If we allow for the free use of disjunctive explanations in the context of IBE, all we can know is that there is some unobservable mechanism that produces our observable evidence. This knowledge is not particularly exciting. This is not exciting because it is not informative. In other words, it does not exclude many possibilities. So, a lesson we can learn from the discussion in section 3 is that IBE can be a useful inductive principle only if informativeness is an explanatory virtue. The principle I proposed in section 4 can be seen as one way to embody the idea that informativeness is an explanatory virtue.

In fact, the idea that informativeness is an explanatory virtue is a quite familiar one. It is controversial whether Molière's famous dormitive virtue explanation is a genuine explanation. In my opinion, it is a genuine explanation. However, even if it is a genuine explanation, it still does not seem to be a *good* explanation. Behind this intuition is the fact that this explanation is not very informative.

Once we admit that informativeness is an explanatory virtue. The idea that we can make an equally good explanation by disjunctively combining a good explanation with bad explanations seems to be unjustifiable. All we can get by this kind of "disjunctification" is some increase of probability. (This increase in probability should not be very impressive because it is achieved by *bad* explanations.) However, by this disjunctive combining, we lose all-important informativeness. An inductive inferential principle that does not produce informative conclusions is useless.

To prevent potential misunderstandings, I must emphasize at this point that I am not saying that *any* kind of informativeness can be regarded as an explanatory virtue. One can make a potential explanation more informative in a way that destroys the potential explanatory relation between explanandum and potential explanans. For example, one can make a potential explanation more informative by conjunctively combining it with an explanatorily irrelevant factor; but the increase of informativeness in this sense is no explanatory virtue. Beebee's own example is helpful here. She writes:

> You want to know why Liverpool has failed to score against much weaker teams so far this season. I tell you it's because Torres has been injured and so out of the team. That's an answer that suppresses adjustable parameters in something like the way that (SF) does, in that my answer leaves it open whether or not Torres will be back in the team next week, next month, next season, or never. But again, so what? (Beebee 2011, 516)

Beebee's answer (call it (Torres)) is less informative than 'Torres has been out of the team so far but he will be back next week' but, as Beebee claims, this more informative answer is no better than (Torres). It is because the more informative answer contains an explanatorily irrelevant conjunct, namely 'he will be back next week.' This explanatorily irrelevant factor undermines the alleged explanatory relation between the explanandum and the alleged explanans. Unlike this answer, N(F, G) does not contain any explanatorily irrelevant conjunct and is more informative than (SF). In short, there are two ways we can increase the degree of informativeness of potential explanations: One that destroys the explanatory relation itself and one that does not. (Lethal-Dose), (Red Ball), and (Torres) are all examples of the former, whereas N(F, G) and (EA) are examples of the latter. My claim is that the latter kind of increase in informativeness is always an explanatory virtue. A problem with Beebee is that she seems to conflate these two. Beebee's (Red Ball) and (Torres) does not show that informativeness is not an explanatory virtue.

References

ARMSTRONG, D.M. (1978): *A Theory of Universals: Volume 2: Universals and Scientific Realism*: Cambridge University Press.

ARMSTRONG, D.M. (1983): *What Is a Law of Nature?* Cambridge – New York: Cambridge University Press.

BEEBEE, H. (2011): Necessary Connections and the Problem of Induction. *Noûs* 45, No. 3, 504-527.

BIRD, A. (2007): *Natures' Metaphysics: Laws and Properties*: Oxford University Press.

BONJOUR, L. (1998): *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge – New York: Cambridge University Press.

FOSTER, J. (1983): Induction, Explanation and Natural Necessity. *Proceedings of the Aristotelian Society* 83, 87-101.

HARMAN, G. (1980): Reasoning and Explanatory Coherence. *American Philosophical Quarterly* 17, No. 2, 151-157.

LEE, J. (2013a): Explanationist Approach to the Problem of Induction and Humean Theories of Explanation. *Korean Journal for the Philosophy of Science* 16, No. 1, 57-80.

LEE, J. (2013b): Non-Instantial Regularity Explanation and the Explanationist Approach. *Philosophical Analysis* 14, No. 1, 1-30.

LEWIS, D. (1994): Humean Supervenience Debugged. *Mind* 103, No. 412, 473-490.

LOEWER, B. (1996): Humean Supervenience. *Philosophical Topics* 24, No. 1, 101-127.

PEACOCKE, C. (2004): *The Realm of Reason*. Oxford – New York: Clarendon Press, Oxford University Press.

PSILLOS, S. (1999): *Scientific Realism: How Science Tracks Truth*. London – New York: Routledge.

SALMON, W.C. (1989): Four Decades of Scientific Explanation. In: Kitcher, P. – Salmon, W.C. (eds.): *Scientific Explanation*. Minneapolis: University of Minnesota Press.

SALMON, W.C. (2001): Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation. In: Hon, G. – Rakover, S.S. (eds.): *Explanation: Theoretical Approaches and Applications*. Kluwer Academic Publishers.

VAN FRAASSEN, B.C. (1989): *Laws and Symmetry*. Oxford – New York: Oxford University Press.

WHITE, R. (2005): Explanation as a Guide to Induction. *Philosophers' Imprint* 5, No. 2, 1-29.

Organon F 22 (2) 2015: 163-177

Four Quine's Inconsistencies

GUSTAVO PICAZO

Department of Philosophy. University of Murcia
Edificio Luis Vives. Campus de Espinardo, Punto 12. Murcia 30100. Spain
http://webs.um.es/picazo/
picazo@um.es

RECEIVED: 05-02-2015 ▪ ACCEPTED: 23-03-2015

ABSTRACT: In this paper I argue that the idiosyncrasy of linguistic competence fosters semantic conceptions in which meanings are taken for granted, such as the one that Quine calls 'uncritical semantics' or 'the myth of the museum'. This is due to the degree of automaticity in the use of language which is needed for fluent conversation. Indeed, fluent conversation requires that we speakers instinctively associate each word or sentence with its meaning (or linguistic use), and instinctively resort to the conceptual repertoire of our language, without calling into question that the meaning of a particular word, or the conceptual repertoire of our language, could have been different than they are. This habit of taking meanings for granted, inherent to our linguistic ability, sometimes interferes with our semantic research, hampering it. In order to illustrate this problem, I pinpoint four places in Quine's work where, despite his acknowledged analytical rigour, and despite his congenital aversion to the habit of taking meanings for granted, he himself appears to slip into this habit, inadvertently.

KEYWORDS: Linguistic competence – meaning theory – myth of the museum – uncritical semantics.

0. Introduction

There are two ways in which naive views of meaning, such as the one that Quine calls 'uncritical semantics' or 'the myth of the museum', take meanings for granted.[1] In the first place, they take for granted the connection between each word or sentence and its meaning (or linguistic use), without dealing with the reasons why that word or sentence points at the particular meaning that it does, instead of at a different one; they do not focus, for instance, on the reasons why 'raining' points at the rain, instead of pointing at snow. In the second place, they take for granted the repertoire of meanings of the language, without dealing with the reasons why each meaning is constrained within the particular limits that it is, and no others; they do not focus, for instance, on the reasons why there is a concept for rain and another for snow, instead of there being one concept which encompasses both phenomena.

The main tenet of this paper is that the idiosyncrasy of linguistic competence fosters semantic conceptions in which meanings are taken for granted in these two ways. Indeed, fluent conversation would be impossible if we stopped at every step to question which word is suitable to express a certain meaning, or whether a meaning belongs to the conceptual repertoire of our language; it is true that such hesitations occur occasionally, but the rule is precisely the opposite: the usual situation is that in which we speakers speedily choose the words we need to express what we want to say – so to speak, 'without thinking'. Then, our very ability to do this induces us to forget that it is a purely contingent fact that the words and sentences of our language have the meanings they have and no other, and that it is a purely contingent fact that our language has the particular repository of concepts that it has and no other. This is how our linguistic competence pushes us to embrace uncritical semantics.

Wittgenstein and Quine have no doubt been among the major 20th century opponents of uncritical semantics. A good semantic theory, they taught us, is one which does not take meanings for granted, but addresses the origin of signification itself: Quine – following Dewey – placed the origin of signification in the speaker's behavioural dispositions;[2] Wittgenstein,

[1] Cf.: "Uncritical semantics is the myth of a museum in which the exhibits are meanings and the words are labels" (Quine 1968, §I, 186).

[2] Cf.: "Dewey was explicit on the point: 'Meaning ... is not a psychic existence; it is primarily a property of behavior' [in reference to J. Dewey, *Experience and Nature*, 1925] ... Semantics is vitiated by a pernicious mentalism as long as we regard a man's semantics as somehow determinate in his mind beyond what might be implicit in his dispositions to overt behavior" (Quine 1968, §I, 185-186).

from a more imprecise but less restrictive angle, placed it in use.[3] On the other hand, Quine's innate analytical rigour and his stature as a mathematical logician made him particularly unlikely to run into contradiction in expressing his thought. And yet I am going to pinpoint four places in Quine's work, belonging to books from four different decades, in which uncritical semantics pops up in his text in the form of inconsistencies, some of them quite evident. I think this should be taken as evidence of the background influence of uncritical semantics, and of the difficulty that even its most tenacious opponents have experienced in trying to get rid of it.

I do not know of publications by other authors in which any of the four inconsistencies that I am going to pinpoint here is clearly identified; I will mention the ones I know that come closest. Quine has already been accused of falling himself into the myth of the museum, but in a much broader context, different from the type of 'local' inconsistency (i.e., one circumscribed to a short fragment of text) with which I am going to deal here.[4] The very fact that these incoherences have gone unnoticed, despite how evident they look once we have put our finger on them, is yet another symptom of the background influence that uncritical semantics continues to have over the philosophy of language today.

[3] Cf.: "The meaning of a word is to be defined by the rules for its use ... Two words have the same meaning if they have the same rules for their use" (Wittgenstein 1979, I, §2). Aiming at a more complete perspective, I have myself pointed to the global process of interaction of the linguistic community with one another and with the environment, as the phenomenon from which meaning emerges: "Meanings are the result of a dynamic process of interaction of the cognitive-linguistic community, between its members and with the environment ... if the process is cut off or seriously disturbed, meaning fades away—just as water stops flowing by a river if we cut off the hydrological cycle which feeds it and keeps it alive" (Picazo 2014, 716).

[4] Cf.: "[O]nce [the] indeterminacy is taken seriously and applied to our own current language as well as to other languages, the manual-relative notions of denotation and signification are not acceptable, either. By employing them, Quine himself has become a victim of the 'myth of the museum'" (Field 1974, 207).

1. You shall not take meaning, synonymy, or analyticity for granted

Our first inconsistency is to be found in Quine's celebrated article 'Two Dogmas of Empiricism', which appeared in *The Philosophical Review* in 1951 and was included two years later in his collection *From a Logical Point of View* (of which a second revised edition was issued in 1961). In the quotes that follow, page numbers correspond to the 1961 edition, though the text quoted here is exactly the same as that of the original versions of 1951 and 1953.

In this paper, as is well known, Quine lays out an attack against the notion of meaning and a cluster of other intensional notions (such as synonymy and analyticity) which, he argues, can only be explained by a circular reference to one another:

> Once the theory of meaning is sharply separated from the theory of reference, it is a short step to recognizing as the primary business of the theory of meaning simply the synonymy of linguistic forms and the analyticity of statements; meanings themselves, as obscure intermediary entities, may well be abandoned. (Quine 1961, §1, 22)

> Analyticity at first seemed most naturally definable by appeal to a realm of meanings. On refinement, the appeal to meanings gave way to an appeal to synonymy or definition. But definition turned out to be a willo'-the-wisp, and synonymy turned out to be best understood only by dint of a prior appeal to analyticity itself. So we are back at the problem of analyticity. (Quine 1961, §4, 32)

> [F]or all its a priori reasonableness, a boundary between analytic and synthetic statements simply has not been drawn. That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith. (Quine 1961, §4, 37)

In sum, Quine seems to be saying: 'you shall not take meaning, synonymy, or analyticity for granted'.

However, just one page before the last of these quotations, an observation sneaks in that completely disregards that commandment:

> It is obvious that truth in general depends on both language and extralinguistic fact. The statement 'Brutus killed Caesar' would be false if the

> world had been different in certain ways, but it would also be false if the word 'killed' happened rather to have the sense of 'begat' (Quine 1961, §5, 36).

To what does Quine refer with 'the sense of "begat"', if not to its meaning? What is he talking about when says that '"killed" happened to have the sense of "begat"', if not the synonymy between those two expressions? And assuming that synonymy, how could then a statement such as '*a* killed *b* if and only if *a* begat *b*' not be analytic?[5, 6]

[5] This flaw went unnoticed by Grice and Strawson in their classic 1956 reply: "If Quine is to be consistent in his adherence to the extreme thesis, then it appears that he must maintain not only that the distinction we suppose ourselves to be marking by the use of the terms 'analytic' and 'synthetic' does not exist, but also that the distinction we suppose ourselves to be marking by the use of the expressions 'means the same as', 'does not mean the same as' does not exist either" (Grice – Strawson 1956, 145). They would not have said this if they had noticed Quine's use of the expression 'having the sense of', clearly equivalent to 'meaning the same as'.

[6] Following suggestions by the referees, I will spell out the contradiction detected in more detail. According to the quotations just given, Quine (1961) asserts: (1) synonymy can be understood only by a prior appeal to analyticity; (2) analyticity is most naturally definable by appeal to meanings; (3) meanings as obscure entities may be abandoned; (4) a boundary between analytic and synthetic statements has not been drawn; and (5) that there is such a distinction is a metaphysical article of faith. From these propositions, three things clearly emerge: (a) that there are no entities which can be called 'meanings'; (b) that meanings cannot be used to give sense to the notion of synonymy; and (c) that meanings cannot be used to draw a boundary between synthetic and analytic statements. However, when Quine says: (6) *if the word 'killed' happened to have the sense of 'begat'*, he is admitting that there is something which is the sense (i.e. the meaning) of a word, thereby contradicting (a); at the same time, by (6) Quine is admitting the possibility that a different word (the word 'killed') had the same meaning that the word 'begat' has, which contradicts (b), because admitting that two words have the same meaning amounts to admitting that they are synonymous; and lastly, by (6) Quine also contradicts (c), because once you have admitted a relation of synonymy between these two words, it is immediate to derive analytical statements thereof. Hence, despite having made an explicit resolution to renounce the notion of meaning, Quine is effectively re-introducing it by talking about the sense of the word 'begat' and the possibility that the word 'killed' happened to have that sense. His own linguistic competence has driven him to take meanings for granted, on that spot, inadvertently.

2. You shall not regard translation as a correspondence between ideas

We will find our second inconsistency in Chapter 2 ('Translation and Meaning') of Quine's book *Word and Object*, published in 1960 and reprinted since then uncountable times, without changes.[7] This chapter is devoted to the mental experiment of radical translation (the task of translating the language of a community with which there has been no previous contact).[8] The moral that Quine extracts from this imaginary situation is that it is wrong to equate translation with a correspondence between meanings (or ideas) of one language and those of the other:

> [T]wo men could be just alike in all their dispositions to verbal behaviour under all possible sensory stimulations, and yet the meanings or ideas expressed in their identically triggered and identically sounded utterances could diverge radically, for the two men, in a wide range of cases. (Quine 1960, §7, 26)

> The stimulus meaning of a sentence for a subject sums up his disposition to assent to or dissent from the sentence in response to present stimulation. (Quine 1960, §8, 34)

> [S]timulus meaning, by whatever name, may be properly looked upon still as the objective reality that the linguist has to probe when he undertakes radical translation. (Quine 1960, §9, 39)

A second commandment emerges from this: 'you shall not regard translation as a correspondence between ideas'.

Notwithstanding, just one page after the last of these quotations, Quine makes the disconcerting observation that:

> We do best to revise not the notion of stimulus meaning, but only what we represent the linguist as doing with stimulus meanings. The fact is

[7] The posthumous so-called 'new edition' (by Cambridge, Mass.: MIT Press, 2013) includes a foreword by Patricia Smith Churchland and a preface by Dagfinn Føllesdal, but no changes within Quine's text itself.

[8] Cf.: "What is relevant rather to our purposes is *radical* translation, i.e., translation of the language of a hitherto untouched people" (Quine 1960, §7, 28; italics are as in the original, unless otherwise stated).

> that he [the radical translator] translates not by identity of stimulus meanings, but by significant approximation of stimulus meanings.
>
> If he translates 'Gavagai' as 'Rabbit' despite the discrepancies in stimulus meaning imagined above, he does so because the stimulus meanings seem to coincide to an overwhelming degree and the discrepancies, so far as he finds them, seem best explained away or dismissed as effects of unidentified interferences ... In taking this rather high line, clearly he is much influenced by his natural expectation that any people in rabbit country would have *some* brief expression that could in the long run be translated simply as 'Rabbit'...
>
> In practice, of course, the natural expectation that the natives will have a brief expression for 'Rabbit' counts overwhelmingly. (Quine 1960, §9, p. 40)

According to this, then, we have to admit that the radical translator relies on his 'natural expectation' to find in the native language an expression which corresponds to the English sentence 'Rabbit', and we have to admit that such expectation influences his translation task 'overwhelmingly'. This amounts, in practice, to taking his own use of 'Rabbit' as the anchor point of the translation, and then looking for an expression of the native language which corresponds to it. However, the very supposition that the native language will have 'some brief expression' which coincides with the English sentence 'Rabbit', and the very modus operandi of focusing on an English sentence first, and then looking for a counterpart to it in the native language, completely deflate the alleged radicality of the scenario. It would be much more radical indeed if the natives did not have a single sentence for 'Rabbit' but various different ones, and none directly translatable into English—e.g. 'Big male rabbit', 'Gray baby rabbit', 'Rabbit affected by a tropical disease not translatable into English', etc.[9, 10]

[9] Erik Stenius identified part of this problem: "[H]ad not our linguist better try to learn the language from within, without taking it for granted that it can be translated into English? The natives may have a culture very different from ours, and even though they operate with the same kind of physical objects as we do, their concepts need not as a rule have exact counterparts in English" (Stenius 1969, §IV, 32). Indeed, the prototypical Sapir-Whorf case is that of Eskimos – i.e., inhabitants of 'snow country' – *not* having a brief expression for 'Snow', but different expressions for different kinds of snow (cf. Lyons 1981, §10.2, 306; and Kilarski 2014, §3 in relation to how many such Eskimo

3. You shall distinguish sentences from their interpretations with the utmost attention

Next we will look at Chapter 1 ('Meaning and Truth') of Quine's 1970 book *Philosophy of Logic.*[11] In this chapter Quine emphasises the need to distinguish between sentences (as strings of symbols devoid of content) and propositions (as entities postulated in order to encapsulate sentence meanings). Quine hastens to reject the existence of the latter,[12] so that the need to differentiate a sentence from its interpretation is for him even more pressing. Hence, the commandment in this case would be: 'you shall dis-

expressions there really are). (I thank José López Martí for drawing my attention to Sapir-Whorf cases in this connection.)

[10] As before, following suggestions by the referees, I will spell out the contradiction detected in more detail. According to the quotations just given, Quine (1960) asserts: (1) two men could be alike in their dispositions to verbal behaviour and yet the meanings or ideas expressed in their utterances could diverge radically; and (2) stimulus meaning is the objective reality that the linguist has to probe when he undertakes radical translation. From these two propositions a conclusion emerges: that translation cannot be regarded as the task of looking, for each expression *e* of the native language, for an expression of the foreign language which most closely matches the idea corresponding to *e*. Such would be the uncritical view of translation – the view of translation derived from the myth of the museum that Quine opposes. However, by stating: (3) *the radical translator is much influenced by his natural expectation that people in rabbit country would have some brief expression that could be translated as 'Rabbit'*, and (4) *in practice this expectation counts overwhelmingly*, Quine is effectively vindicating the uncritical view of translation that he had initially set out to oppose. Again, it seems that it is Quine's own linguistic competence which drives him to assume that the conceptual repertoire of the native language will match that of his own (at least with respect to this simple sentence), without realising that such an assumption is not only questionable – for the reasons I have explained in the previous footnote –, but completely alien to the conception of meaning he is trying to articulate.

[11] A second edition was published in 1986, though with no changes affecting the passages that we are going to quote here, nor their pagination.

[12] Cf.: "The notions of proposition and meaning will receive adverse treatment" (Quine 1970, Preface). "In inveighing against propositions in ensuing pages, I shall of course be inveighing against them always in the sense of sentence meanings" (Quine 1970, 2). "The uncritical acceptance of propositions as meanings of sentences is one manifestation of a widespread myth of meaning. It is as if there were a gallery of ideas, and each idea were tagged with the expression that means it; each proposition, in particular, with an appropriate sentence" (Quine 1970, 8).

tinguish sentences from their interpretations with the utmost attention'. And this chapter indeed contains various remarks to that effect:

> [S]ome writers ... are careless about the distinction between sentences and their meanings. (Quine 1970, 2)

> The quotation is a name of a sentence... . (Quine 1970, 12)

> [A]n eternal sentence that was true could become false because of some semantic change occurring in the continuing evolution of our own language. Here again we must view the discrepancy as a difference between two languages: English as of one date and English as of another. The string of sounds or characters in question is, and remains, an eternal sentence of earlier English, and a true one; it just happens to do double duty as a falsehood in another language, later English. (Quine 1970, 14)

Of course, if quoting a sentence is enough to name it, it must be that what is named is the mere string of symbols, given that sentences are often ambiguous, or indexical, so that the same string of symbols is used to signify different things. And if a sentence can change its truth value in consequence of the evolution of language, it must be that the sentence is again the mere string of symbols, and not the meaning conveyed.[13]

Notwithstanding, amidst these pages Quine makes a remark that sharply deviates from such a guideline:

> No sentence is true but reality makes it so. The sentence 'Snow is white' is true, as Tarski has taught us, if and only if real snow is really white. The same can be said of the sentence 'Der Schnee ist weiss'; language is not the point. In speaking of the truth of a given sentence there is only one indirection; we do better simply to say the sentence and so speak not about language but about the world. So long as we are speaking only of the truth of singly given sentences, the perfect theory

[13] For example, the sentence 'Snow is white' would cease to be true if, as a consequence of the evolution of English, the word 'snow' shifted its meaning to 'grass'. However, nobody would say that the proposition <Snow is white> (the meaning conveyed by the sentence 'Snow is white' in present English) had ceased to be true because of that. We would say – admitting talking of propositions – that the proposition <Snow is white> continues to be true, but the sentence 'Snow is white' no longer expresses it in English as of that later date.

> of truth is what Wilfrid Sellars has called the disappearance theory of truth. (Quine 1970, 10-11).

But how can it be that reality on its own (the whiteness of snow, in this case) makes true a mere sequence of symbols? And how can this have nothing to do with language? How can it be that language has nothing to do with the fact that 'Snow is white' comes true in virtue of the colour of snow? How, then, does the whiteness of snow 'connect', as it were, with that string of symbols? How does the whiteness of snow manage to bring about the fact that such a string of symbols – supposedly devoid of content – comes out true, instead of false, or undetermined?

The problem becomes worse if we look at language evolution: we have just read on page 12 that a true eternal sentence could become false in consequence of a semantic change in the diachronic evolution of language; however, 'Snow is white' is one such sentence, and yet Quine says on page 10 that it is true if and only if snow is white, language not being the point. So what is the story?

And if we take into account meaning differences derived from the context of utterance (indexicality), or meaning differences between different synchronic languages, even more difficulties arise:

> Having now recognized in a general way that what are true are sentences, we must turn to certain refinements. What are best seen as primarily true or false are not sentences but events of utterance. If a man utters the words 'It is raining' in the rain, or the words 'I am hungry' while hungry, his verbal performance counts as true. Obviously one utterance of a sentence may be true and another utterance of the same sentence be false. (Quine 1970, 13).

> Conceivably, by an extraordinary coincidence, one and the same string of sounds or characters could serve for '2 < 5' in one language and '2 > 5' in another. When we speak of '2 < 5' as an eternal sentence, then, we must understand that we are considering it exclusively as a sentence in our language, and claiming the truth only of those of its tokens that are utterances or inscriptions produced in our linguistic community. (Quine 1970, 14).

Combining these last observations with the idea that the world is by itself, independently of language, responsible for each sentence having a truth value or another, we arrive at the following conclusion: the world is supposedly equipped with a means to provide a truth value to each sentence, in every possible context of utterance, with respect to every language to which it belongs, and with respect to every stage in the evolution of such a language. The supposition that Quine is postulating such a portentous mechanism, of which he explains absolutely nothing, is absurd. It is much more charitable to interpret that what Quine is doing in that troublesome passage (the passage on pages 10-11 in which he mentions Tarski and Sellars) is to take the sentence 'Snow is white' as bounded to its ordinary English meaning. In other words, it is more charitable to interpret Quine in that passage as taking the sentence 'Snow is white', not as a mere string of symbols devoid of content, but as a communicative piece of English with a defined use inside the English semantic arsenal.[14, 15]

[14] Thomas (2011) examines that troublesome passage in relation to the apparent 'blunder' detected by Künne (2003). From his analysis Thomas concludes, with relief, that the blunder is only apparent: "(4): 'Snow is white' is true iff [real] snow is [really] white. As we have seen, the truth of (4) is consistent with there being no dependency between the truth of 'Snow is white' and snow's being white, and so it seems that Quine cannot appeal to (4) to account for the fact that 'Snow is white' is made true by snow's being white" (Thomas 2011, §1, 115); "[T]o make sense of the above quote from Quine ... one can point out that if (4) is invoked in an explanation of the truth of 'Snow is white' then it is implicated that the truth of 'Snow is white' depends on snow's being white" (Thomas 2011, §2, 118); "This removes a much-discussed problem for deflationism (and saves Quine from the suggestion that he has made an obvious blunder)" (Thomas 2011, §4, 122). However, Thomas fails to notice that whether the truth of (4) implicates a truthmaking dependency of 'Snow is white' on the whiteness of snow is not the only controversial issue here: the very truth of (4) is directly questionable. Indeed, as we have seen in the previous footnote, the sentence 'Snow is white' could be untrue despite snow being white, on condition of the meaning of that sentence being different than it actually is.

Künne too fails to notice this point: "The predicates '*x* is made true by *y*' and '*x* is true in virtue of *y*' signify asymmetrical relations, so we cannot preserve the point of the slogan 'No sentence is true but reality makes it so' by using a 'symmetrical' (commutative) connective even if we embellish the right-hand side of the bicondicional by a generous use of 'real(ly)'" (Künne 2003, §3.5.1, 152). However, Künne does not notice that if we regard sentences as strings of symbols devoid of content (which is the way in which Quine says we have to do it), then the problem is not whether biconditional (4) is insufficient to represent the asymmetry of the truthmaking relation: the problem is

4. On no account shall you suppose that a sentence by itself points to a particular meaning

Finally, we will turn to Quine's 1981 book *Theories and Things*, and in particular to its Chapter 5, 'Use and Its Place in Meaning'.[16] The incoherence that we are going to pinpoint here is closely related to the previous one, given that both of them are based on the same difficulty: the difficulty of contemplating words and sentences separately from our competence to use them correctly as speakers of the language to which they belong. Indeed, our linguistic competence drives us to do precisely the opposite: our degree of automaticity in sentence comprehension pushes us to take for granted the connection between each sentence and the content it conveys.

The chapter with which we are concerned now begins emphasising again the need to regard words and sentences as uninterpreted sequences of symbols:

> An expression, for me, is a string of phonemes – or, if we prefer to think in terms of writing, a string of letters and spaces. Some expressions are sentences. Some are words. Thus when I speak of a sentence,

that biconditional (4) is straightforwardly false. (I identified this problem in Picazo 2014, 724.)

[15] Again I proceed to spell out this contradiction in more detail. According to the quotations just given, Quine (1970) asserts: (1) we have to distinguish between sentences and their meanings; (2) an eternal sentence that was true could become false because of some semantic change in the evolution of our own language; and (3) one utterance of a sentence may be true and another utterance of the same sentence be false. From these propositions two conclusions emerge: (a) that a sentence by itself has no predetermined meaning; and (b) that it is only through the use of a sentence that a meaning becomes attached to it. But then it is impossible to accept Quine's further assertion that: (4) *the sentence 'Snow is white' is true if and only if snow is white*. Indeed, there is no way in which the whiteness of snow per se can manage to make true a string of symbols devoid of content. It is once again the idiosyncrasy of linguistic competence which makes Quine to slip on this spot, pushing him to take for granted that the sentence 'Snow is white' has a predetermined meaning – the meaning it has, the meaning he is trained to automatically attach to it – instead of viewing it as an empty sign, which was the way he had set out to do.

[16] The text of this chapter is made up from two previous papers of Quine, published in 1978 and 1979 (see Quine 1981, Ch. 5, 43, for more details).

> or of a word, I am again referring to the sheer string of phonemes and nothing more. I must stress this because there is a widespread usage to the contrary. The word or sentence is often thought of rather as a combination, somehow, of a string of phonemes and a meaning... This use ... cannot be allowed here, because our purpose is to isolate and clarify the notion of meaning.
>
> A meaning, still, is something that an expression, a string of phonemes, may *have*, as something external to it in the way in which a man may have an uncle or a bank account. It has it by virtue of how the string of phonemes is used by people...
>
> The point is that the notion of an expression must not be allowed to presuppose the notion of meaning. (Quine 1981, 44)

The commandment is clear, once more: 'on no account you shall suppose that a sentence by itself points to a particular meaning'.

However, just five pages later Quine introduces his distinction between occasional and non-occasional sentences, with the following words:

> [W]e must limit our attention for a while in yet another way: we must concentrate on occasion sentences. These, as opposed to standing sentences, are sentences whose truth values change from occasion to occasion, so that a fresh verdict has to be prompted each time. Typically they are sentences that contain indexical words, and that depend essentially on tenses of verbs. Examples are 'This is red' and 'There goes a rabbit'... . (Quine 1981, 49)

Looking at this definition we must wonder, once again: how can an uninterpreted string of phonemes have a truth value, and how can such a truth value change from one occasion to another? What does it mean that an uninterpreted string of phonemes contains indexical words, or that it depends 'essentially' on tenses of verbs? And what reason might there be for pointing to the strings of phonemes 'This is red' and 'There is a rabbit' as examples of occasion sentences, if not the fact that *the meaning they express* (the use they have in present English) exemplify the kind of occasion variability that Quine has in mind?

It is important to notice that occasionality cannot be a property of the sentence, because a sentence might be ambiguous between two different readings, one of which constitutes an occasional meaning and the other does not. One such example is the sentence 'The church survived communism', which might be predicated of a particular physical church (say, the church of a town belonging to a former communist country), and in that case it will be true or false depending on the church in question; but the same sentence can also be used as a historical observation about the Church as a whole, and in such a case it will behave as a standing sentence, whose truth value will not change with the context of utterance. To be precise, in formal writing the second use of 'Church' should be capitalised, but we will still have a unique sentence in oral language – i.e. a unique sequence of phonemes.

On the other hand, occasionality can neither be attributed to the very occasion of utterance. For by definition each occasion of utterance determines a particular meaning for the sentence uttered, so it would be absurd to count some occasions of utterance as occasional and others as not (there are no occasions 'more occasional' than others). The occasionality of which Quine is talking about is neither attributable to the sentence nor to the occasion of utterance.

The only thing that may or may not be occasional is the *sense* of the sentence, i.e. its meaning. Indeed, occasion meanings are those which behave as meaning-schemata, that is, as fragmentary meanings that need to be filled in by reference to the context of utterance; while standing meanings are those that can be understood independently of the context. The difference between the occasional 'The church survived communism' and the standing 'The Church survived communism' is that in order to understand what is meant by the former we need to know what particular church we are talking about, something which will depend on the context of utterance; while in order to understand what is meant by the latter, the context of utterance is irrelevant.[17, 18]

[17] Once again I will spell out the contradiction detected in more detail. According to the quotations just given, Quine (1980) asserts: (1) an expression is a string of phonemes, or of letters and spaces; (2) some expressions are sentences; (3) when speaking of sentences we refer to the sheer string of phonemes and nothing more; (4) sentences cannot be thought of as a combination of a string of phonemes and a meaning; (5) a meaning is something that an expression may have as something external to it, by virtue of how it is used by people; and (6) the notion of an expression must not presuppose the notion of meaning. From these propositions, again, it emerges: (a) that a sentence by itself has no predetermined meaning; and (b) that it is only through the use of a sentence that a meaning becomes attached to it. But then it is impossible to accept

References

FIELD, H. (1974): Quine and the Correspondence Theory. *The Philosophical Review* 83, No. 2, 200-228.

GRICE, H.P. – STRAWSON, P.F. (1956): In Defense of a Dogma. *The Philosophical Review* 65, No. 2, 141-158.

KILARSKI, M. (2014): Complexity in the History of Language Study. *Poznań Studies in Contemporary Linguistics* 50, No. 2, 157-168.

KÜNNE, W. (2003): *Conceptions of Truth*. Oxford: Clarendon Press.

LYONS, J. (1981): *Language and Linguistics: An Introduction*. Cambridge: Cambridge University Press.

PICAZO, G. (2014): Truths and Processes: A Critical Approach to Truthmaker Theory. *Philosophia* 42, No. 3, 713-739.

QUINE, W.V.O. (1951): Two Dogmas of Empiricism. *The Philosophical Review* 60, No. 1, 20-43.

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

QUINE, W.V.O. (1961): *From a Logical Point of View*. 2nd ed. Cambridge (Mass.): Harvard University Press.

QUINE, W.V.O. (1968): Ontological Relativity. *The Journal of Philosophy* 65, No. 7, 185-212.

QUINE, W.V.O. (1970): *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall.

QUINE, W.V.O. (1981): *Theories and Things*. Cambridge (Mass.): Harvard University Press.

STENIUS, E. (1969): Beginning with Ordinary Things. In: Davidson, D. – Hintikka, J. (eds.): *Words and Objections: Essays on the Work of W.V.Quine*. Dordrecht: Reidel, 27-52.

THOMAS, A. (2011): Deflationism and the Dependence of Truth on Reality. *Erkenntnis* 75, No. 1, 113-122.

WITTGENSTEIN, L. (1979): *Wittgenstein's Lectures: Cambridge, 1932–1935*. Ambrose, A. (ed.). Oxford: Blackwell.

Quine's further assertion that: (7) *occasions sentences, as opposed to standing sentences, are sentences whose truth values change from occasion to occasion—typically they are sentences that contain indexical words, such as 'This is red' and 'There goes a rabbit'* because as I have argued, these claims only make sense if we are talking about the meanings (or uses) of sentences, not if we are talking about sentences as empty signs. The oversight is due, again, to the strong – automatic – association between these sentences and their meanings, in Quine's mind.

[18] I am grateful to two anonymous referees of *Organon F* for their suggestions on an earlier draft of this paper.

Organon F 22 (2) 2015: 178-198

Kant and the Problem of Self-Identification[1]

LUCA FORGIONE

Dipartimento Scienze Umane. Università degli Studi della Basilicata
via N.Sauro 85. 85100 Potenza. Italia
luca.forgione@gmail.com

RECEIVED: 02-12-2014 ▪ ACCEPTED: 12-03-2015

ABSTRACT: Ever since Strawson's *The Bounds of Sense*, the transcendental apperception device has become a theoretical reference point to shed light on the criterionless self-ascription form of mental states, reformulating a contemporary theoretical place tackled for the first time in explicit terms by Wittgenstein's *Blue Book*. By investigating thoroughly some elements of the critical system the issue of the identification of the transcendental subject with reference to the *I think* will be singled out. In this respect, the debate presents at least two diametrically opposed attitudes: the first – exemplified in the works by Hacker, Becker, Sturma and McDowell – considers the features of the *I think* according to Wittgenstein's approach to the *I as subject* while the second, exemplified by Kitcher and Carl, criticizes the various commentators who turn to Wittgenstein in order to interpret Kant's *I think*. The hypothesis that I will attempt at articulating in this paper starts off not only from the transcendental apperception form, but also from the characterizations of empirical apperception. It may be assumed that Kant's reflection on the problem of self-identification lies right here, truly prefiguring some features of Wittgenstein's uses of *I*, albeit from different metaphysical assumptions and philosophical horizons.

KEYWORDS: Empirical apperception – Kant – self-identification – self-reference – transcendental apperception – Wittgenstein.

[1] Kantian English quotations are from the *Cambridge Edition of the Works of Immanuel Kant*. The citations include the page numbers in the 'Akademie' edition of Kant's works. For references to the first *Critique* (KrV), the page numbers are from the A (1781) and B (1787) editions.

1

In a well-known passage Wittgenstein (1958, 66-67) introduces his philosophico-linguistic analysis of the grammatical rule of the term *I*, where he distinguishes two types of uses, the *use as object* ('I have grown six inches') and the *use as subject* ('I have toothache'):

> One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error . . . On the other hand there is no question of recognizing a person when I say I have toothache. To ask "are you sure it's you who have pains?" would be nonsensical.

This passage should be considered as part of the philosophical framework articulated by Wittgenstein starting from the 1930s on the basis of some theses that might be regarded as the background for the analyses of the two uses of *I*.[2] While the *I used as object* performs a referential function relative to the body and to physical features in general, the *I used as subject* apparently regards mental states as well as processes and no subject identification is taken into account.[3]

Likewise, Strawson (1966, 165) argues that in the self-ascription of a mental state (e.g., 'I'm hungry'), a subject of experiences uses the term *I* employing no identification criteria:

> It would make no sense to think or say: *This* inner experience is occurring, but is it occurring to *me*? (This feeling is hunger; but is it I who am feeling it?)

[2] For example: a) the irreducibility of the manifold 'games' that build up language, whose rules are to be made explicit in order to solve any sort of philosophical problem; b) anti-referentialism, which lies on the recognition of the manifold functions performed by language as well as on the necessity to avoid the erroneous search for the use of a sign on the basis of the object-sign relation; c) anti-mentalism, for which suggesting that thinking is a mental activity is misleading.

[3] From a Wittgensteinian angle, the *I used as subject* has no referential function: according to this thesis – supported by Geach (1957), Hacker (1972), and Anscombe (1975) – it is just our inclination to assume that a linguistic term has a meaning only if it stands for an object that induces us to believe that the *I used as subject* denotes the thinking subject, mind, soul, etc. Cf. Sluga (1996); Wright (1998).

More precisely, Strawson refers to *criterionless selfascription*. On the other hand – unlike Wittgenstein, as we shall see – the absence of an identification device does not entail that the use of *I* will not perform a referential function.

Some judgments bearing a first-person reference (e.g., 'I have pain') display what Shoemaker (1968, 565) defines *self-reference without identification* (which is linked to the feature of the essential indexical *I* singled out by Kaplan, Castañeda and Perry):

> My use of the word "I" as the subject of my statement is not due to my having identified as myself something of which I know, or believe, or wish to say, that the predicate of my statement applies to it.

In other words, in the self-ascriptions of mental properties, the self-reference underlying some self-conscious forms occurs without any inference from conceptual properties ascribable to the subject: there is no previous identification of something as its own self owing to properties that can be ascribed to that same something. Due to the absence of any identification component, some *singular judgments* involving the self-ascription of mental (and physical, as will be seen) properties are *immune to error through misidentification* relative to the first-person pronoun (*IEM*). The subject formulating this sort of judgments in given epistemic contexts cannot be mistaken as to whether it is he who is attributing a particular mental property to his own self.

In his turn, Evans goes beyond the terms of the matter as suggested by Wittgenstein and, to some extent, by Shoemaker. In some self-ascriptions, self-reference is direct and unmediated: as Evans notes, here we are dealing with *identification-free self-reference*. In particular, Evans (1982, 220) contends that a judgment of the kind 'I am F' is *identification-free* unless it corresponds to the inferential conclusion drawn from the two premises, i.e., 'a is F' (*predication component*) and 'I am a' (*identification component*), such judgment is based on the unmediated self-ascription of properties through introspective consciousness (as is the case with mental properties) or proprioception (as with physical properties). For example, according to our general capacity to perceive bodies, to our sense of proprioception, of balance, of heat and cold, and of pressure, the kind of information generated by each of these modes of perception seems to give rise to judgments that are immune to error through misidentification:

> None of the following utterances appears to make sense when the first component expresses knowledge gained in the appropriate way: ‘Someone’s legs are crossed, but is it my legs that are crossed?’ (Evans 1982, 220)

Peacocke’s (1999; 2008) strategy in its turn consists in tracing *IEM* proprieties back to more fundamental characterizations. More precisely, Peacocke (1999, 274) distinguishes between a representationally dependent and a representationally independent use of the first-person concept in order to define what he terms *delta account*. The point at issue here is primarily epistemological as it concerns the philosophical branch of *self-knowledge* as well as the possibility of forming beliefs relative to the self-ascription of mental and physical properties. While the representationally dependent use of the first-person concept is based on the fact that the subject is represented in the content of the judgment, in the representationally independent use of the first-person concept the very occurrence of a particular experience (e.g., visual, its content being, in Peacocke’s example, ‘I see the phone is on the table’) determines the reason why the subject is justified in making a judgment about herself, without the thinking subject being represented in the judgment itself:

> the explanation is just the occurrence of the experience itself to its subject. Nor does any thought or representation of herself as the subject of the experience enter her reasons for her judgement. (Peacocke 1999, 273)

Mutatis mutandis, Recanati (2007; 2009) employs a similar strategy through a philosophical analysis on the distinction between *de re* and *de se* thoughts, a distinction gained by Chilsholm, Lewis and Perry. The author distinguishes two types of *self-ascriptions*, two types of *de se* thoughts – *implicit de se* and *explicit de se* – according to the presence or absence of the representational reference of the subject in terms of judgment content. Although very problematically, Peacocke relates the *IEM* phenomenon to the representationally independent use of the concept *I*, and in the same way Recanati finds a link between *IEM* and implicit *de se* thoughts. As will be seen, Peacocke appeals to the representationally independent use of *I* to explain the origin of the transcendental subject in Kant.

2

The question of the identification of the subject in the transcendental apparatus can be developed through two theoretical dimensions. Firstly, it is necessary to introduce a metaphysical reflection on the transcendental subject in order to detect the characterizations assigned to transcendental apperception and contained in several passages of the *Transcendental Deduction* and *Paralogisms* sections: from the point of view of spontaneity of understanding, *I* manifests itself neither as it is nor as phenomenon. Secondly, within the transcendental constraints characterizing the designation of the *I* of the *I think*, I will discuss the empirical dimension and the epistemic conditions in and under which the subject reveals himself in the temporal sphere of receptivity.

As is well known, transcendental constraints represent the conditions of possibility of experience and knowledge, and in the final analysis they are based on transcendental apperception, i.e., on self-consciousness, which, via the *I think*, Kant regards as the highest point of transcendental philosophy. Apperception is the foundation of representational synthesis in order for knowledge to occur, and the *I think* must be able to accompany every representation: regarded as an analytical unity of apperception, the representation *I* produced by apperception is a feature of every representation as the *I think* must be able to accompany each representation; regarded, on the other hand, as a synthetical unity of apperception, the *I* produced by apperception is a feature of representations synthesized horizontally, which calls in question the categories' claim to have objective validity and to be predicates of objects in general, so that the judgments can be formed wherever knowledge arises. Although it presents the thinking subject as substantial, simple, identical in time and separate from body, the *I* of the *I think* is not the concept of an object but refers to 'something in general (transcendental subject)' in which thoughts inhere as its own predicates.

Pure apperception is original consciousness and can be expressed by *sum*. Here, as emphasized by Capozzi (2007, 288), an ontological question arises: *sum* is nothing but activity – which has nothing receptive about it – as it will not mingle with any element of the sensible dimension; hence, it is a thinking activity to the extent that *sum* and *cogito* are *on a par:* in the first act of knowledge, "I am a thinking thing is a tautology." The ontological question is specified in the assertion that "the subject bound to the first act of knowledge, to former apperception, is the first subject as well as the first *Wesen* being thought: with the first act of knowledge, the subject is the being itself."

From this metaphysical perspective emerge a few characterizations of the transcendental subject which may explain the lack of identification in representational synthesis. On the one hand, the *I think/I am* is a formal condition of all thinking: "the *I think* must be able to accompany all representations" (*KrV* B 132). On the other hand, this subject/being is something in general, unidentifiable from an epistemic point of view; it is an intellectual self-existence awareness summarized by the *I am* or *I think* representations which accompany every other representation and, as such, don't present any propriety. In point of fact, due to the absence of intuition, it is not possible to determine whether that something is existent as a persistent substance in order to make knowledge:

> The consciousness of myself in the representation I is no intuition at all, but a merely intellectual representation of the self-activity of a thinking subject. (*KrV* B 278)

> In the synthetic original unity of apperception, I am conscious of myself not as I appear to myself, nor as I am in myself, but only that I am. This representation is a thinking, not an intuiting. (*KrV* B 157).

What is being assumed on the basis of the representation *I* is just an existent devoid of any propriety. The subject is able to know that he exists as a thinking activity, but he is not able to know what he is: the subject's being is inaccessible from an epistemic point of view, and what is given is nothing but thoughts regarded as his predicates, which do not enable us to grasp the thinking subject's nature. In a famous passage Kant states:

> Through this I, or He, or It (the thing), which thinks, nothing further is represented than a transcendental subject of thoughts = x, which is recognized only through the thoughts that are its predicates, and about which, in abstraction, we can never have even the least concept. (*KrV* A 346/B 404)

Accordingly, there emerge a few peculiarities of the self-referential apparatus involved in transcendental apperception: "the subject of inherence is designated only transcendentally through the I that is appended to thoughts, without noting the least property of it, or cognizing or knowing anything at all about it" (*KrV* A 355). It follows that the act of reference performed by the subject to refer to her own self entails no mediation of knowing, namely it involves no identification by means of properties ascribable to the subject herself.

With the notion of *transcendental designation*, Kant anticipates some of the *self-reference without identification* features (cf. Howell 2000; and Brook 2001). The condition of possibility of all judgments relies on the act *I think* and, at this level, the intellectual representation *I* designates only transcendentally, no conceptual mediation being involved: it is a simple representation bearing no content and solely referring to something in general, namely to a transcendental subject: "its properties [of subject] are entirely abstracted from if it is designated merely through the expression 'I', wholly empty of content (which I can apply to every thinking subject)" (*KrV* A 355). An empty or bare form (cf. *KrV* A443/B471), *I* designates but does not represent (cf. *KrV* A 381; Kant 1786, 542-543). The difference is important, the *I* designates the transcendental subject without representing it, i.e. without any content mediation and therefore without any prior instance of identification because the *I* is not a conceptual representation, articulable in conceptual marks, nor an intuitional one, which presupposes a relation to the sensible spatio-temporal forms, but is a 'simple' or 'empty' representation.

3

The analysis of the form of the *I* is intertwined with several epistemic and metaphysical questions. In general, it should be highlighted that the absence of an identification component does not imply that the *I* doesn't perform a referential function, nor that it necessarily involves a specific metaphysical thesis on the nature of the self-conscious subject. As a matter of fact, the *I-thoughts* self-reference features have been supported by both a materialist conception regarding the self-conscious subject as a bodily object – for example, by Strawson and Evans – and a different metaphysical framework, as in Wittgenstein's eliminativist thesis or in Kant's exclusion thesis.[4]

[4] At face value, Kant suggests a metaphysical thesis of exclusion according to which the *I* of the *I think* as intellectual representation produces no knowledge as to the nature

The point at issue here is the possible contiguity between the analysis of the form of *I think* and the contemporary reflections on the question of self-identification: although it is possible to find different elements of affinity, the theoretical contexts are still deeply distant.

With his transcendental designation, Kant does not certainly seem anachronistic with respect to the considerations raised by the contemporary debate. On the contrary, at least as far as the genesis of the Cartesian illusion on the thinking subject's immaterial nature is concerned – one of the issues addressed in the *Blue Book* – Kant and Wittgenstein seem to share the same philosophical concerns and both focus, although not exclusively, on the type of reference involved by the *I,* of course through very different philosophical paths, and, as already said, with what at first may appear to be antipodal metaphysical assumptions.

Wittgenstein (1958, 43) starts from the analysis of language and the use of the *I as subject* to dissolve any question on the nature of the ego in an anti-metaphysical key. Philosophical inquiry must investigate only the *grammars* of the mentalistic terms used and no metaphysical distinction between the mental and the physical should follow from the distinction between propositions describing facts of the world and propositions describing psychological experiences. It is necessary to analyze the uses and related grammars of terms such as *thinking*, *meaning*, *wishing* because the investigation "rids us of the temptation to look for a peculiar act of thinking, independent of the act of expressing our thoughts, and stowed away in some peculiar medium". Thinking is using signs according to rules and philosophical difficulties may arise only from the misleading use of language

of the thinking subject, and refers only to something which is no object: the transcendental subject. At the same time, the empty form of the referential apparatus in transcendental apperception has been appraised in intrinsically different ways, from Heinrich and Guyer's *Substantial Ownership Reading* to the *Formal Ownership Reading* upheld by Allison and Ameriks, and, more recently, by Bermúdez (1994). Further, more than once an elusive reading suggesting that the *I* of the *I think* has no reference has been argued, i.e., the so-termed *No-Ownership Reading*; the close affinity between Wittgenstein and Kant alleged by some commentators lies within this framework – cf. Becker (1984), Sturma (1985), Powell (1990), McDowell (1994). In this paper, contiguity between Wittgenstein and Kant on this issue is refused, since, as already pointed out, the *I* of the *I think* has a referential function which refers to the transcendental subject.

which leads us to look for something that might correspond to a noun. This may be the case in the use of the *I as subject*.

The referential thesis according to which the use of a sign is based on its relation with the object – strongly criticized when taken as the sole basis to explain the semantics of the language, along with the proper consideration that some uses of the *I* do not denote physical properties – leads to false Cartesian metaphysical conclusions:

> We feel then that in the cases in which "I" is used as subject, we don't use it because we recognize a particular person by his bodily characteristics; and this creates the illusion that we use this word to refer to something bodiless, which however, has its seat in our body. In fact this seem to be the real ego, the one of which it was said, "Cogito ergo sum". (Wittgenstein 1958, 69)

In no way is the question of the absence of identification in the use of the *I* lacking in Kant, as already seen with the transcendental designation of the *I* of the *I think*. Given that there is no empirical intuition, the *I* of the *I think* cannot be based on public employment through the identifying mediation of properties attributable to the transcendental subject. However, and in contrast to Wittgenstein, in the first place Kant moves from a metaphysical reflection in the sense of transcendental idealism concerning the conditions of possibility of experience and knowledge and from the transcendental assertion that the *I think* is the center of such conditions. Philosophical inquiry can only analyze the formal constraints of knowledge. In revealing the genesis of the illusion of a Cartesian immaterial ego, mainly addressed in the analysis of paralogisms, Kant argues that nothing about the metaphysical order and the ontological nature of the transcendental subject can be elicited from the conscious form of unity of apperception and from the representational order of the *I think*: the *I* of the *I think* is not a concept of an object, but an empty representation, 'the concept of a mere something'.

Needless to say, it was Strawson himself who insisted on the characterizations of the form of the *I think*: based on some arguments recalling Wittgenstein's theses in more than one way, Strawson (1966, 166) claims that Kant has revealed the source of the Cartesian error from a purely internal referential use of the *I*, which severs all ties with the ordinary and empirical criteria of self-identity. At the same time, Strawson criticizes Kant for not considering explicitly, as a condition of possibility of experience, the requirement that the subject is recognizable as an object of intuition, a thesis further developed by Evans (1982) and Cassam (1997), and challenged, in its turn, by some Kantian commentators (cf. *infra*).

Although Peacocke does not explicitly mention Strawson, he nonetheless uses his same type of argument against Kant, stating that the German philosopher would have mistaken an epistemological phenomenon for a metaphysical one. The author refers, among others, to the above-mentioned *KrV* B 404 passage to specify the notion of transcendental subject: this is not an empirically determinable object through the application of the categories and it can only be known through the thoughts regarded as its own predicates. This metaphysical conclusion of exclusion can only stem from a failure to recognize the representationally *independent use* of the *I* of the *I think*, which is the form of any judgment; yet, as such, it neither presents the subject in the content of the judgment itself, nor can determine or identify it *a fortiori* (see Peacocke 1999, 284).

Now, for those who harbor sympathies for the hypothesis according to which Kant asserts a metaphysical thesis of exclusion (cf. *supra*, footnote 4), Peacocke's argument might be reversed in its turn by contending that Kant may have held a metaphysical position on the exclusion of the transcendental subject from the metaphysical order of reality to generate the epistemic phenomenon of the representationally independent use of *I*, not vice versa. For those wishing to recall the *Formal Ownership Reading* and the above-mentioned Kantian arguments contained in the analysis of paralogisms, it is not possible to elicit any metaphysical conclusion about the transcendental subject from the formal and representational order of the *I think*. Indeed, when Kant focuses on the logic exposition of apperception, he describes the transcendental subject's function through a completely abstract locution such as *Das Denken* (see *KrV* B 428-9). At this transcendental level for the simple thought the thinking thing is the being itself and shows no proprieties at all, to the extent that Kant leaves even the pronoun indeterminate ('I', 'he', 'it'), and yet he points out that it is a something in general, i.e., the transcendental subject.

4

Therefore – at last we are thus confronted with the assessments marking the distance between Kant's approach and contemporary reflections – considerations on the *I* of the *I think* rest at a very different level of investigation than the contemporary approaches. The transcendental unity of apperception is the foundation of representational synthesis, through which an objective determination of representations arises for possible cognition: each empirical manifold given in the intuitions of sensibility is determined by the functions of the power of judgment based on the application of the categories of understanding that bring it back to consciousness. In this sense, every manifold bears a necessary relation with the *I think* that is the foundation of the necessary unity of the objectively valid connection of all representations expressed by the judgment. In this picture, the *I think* resides in a metaphysical frame which necessarily involves any thinking activity since it does identify with such activity. At least at this level of investigation, and with respect to the passages considered, this represents the highest level of abstraction in the transcendental reflection.

At the same time, in transcendental self-consciousness, the self-attributions of any thought (and also of a transcendental category: I think substance, cause, etc.) are not based on identification component relative to the representation *I* underlying the determination of those thoughts: as already remarked, although the representation *I* designates the transcendental subject, it cannot be determined to identify the thinking entity as empirical object (cf. *KrV* A 346/B 404).

However, if the act of spontaneity expressed by the *I think* is necessarily involved in the making of any judgment, the lack of identification component entailed in the transcendental designation appears to be totally empty of meaning: the Kantian reflections on the *I think* cannot articulate the different types of singular judgments expressing self-ascriptions of mental and physical properties as these regard form and condition of possibility of any kind of judgment, regardless of the particular use of *I* (as *subject* or as *object,* in Wittgensteinian terms) involved in the singular judgments produced. In other words, and more concretely, the transcendental designation mechanism of the *I think* cannot account for the presence or absence of subjective identification component relative to the first person in judgments such as

'I have grown six inches' or 'I have toothache' since that is the condition of possibility of both.

In this regard, Longuenesse (2012) distinguishes two different uses of the *I as subject*. The first use is relative to Strawson and Evans, as well as Cassam (cf. Longuenesse 2006), who point to the consciousness of the self as a spatio-temporally located object and to the channels from which the subject draws information about himself so as to produce possible *IEM* judgments relative to the first person. The second *I as subject* use, instead, relates to Kant's position as well as to the subject's awareness of mental unity. Further, Longuenesse rejects Evans's criticism against Kant – as a matter of fact, a criticism already made by Strawson (1966) and, years later, also emphasized by McDowell (1994) – whereby the *I think* presents a purely formal characterization which is not sufficient to account for the self-referential capacity of the self-conscious subject.

Indeed, it seems difficult to compare two levels of investigation that are so different, *mutatis mutandis* highlighted in a different theoretical framework by Perry (1986) and Recanati (1997): plainly, it is assumed that – starting from the *I think* – Kant articulates a metaphysical reflection also on the conditions of possibility of the identification mechanisms by applying the intellectual forms to the sensitive ones. Such reflection, as already said, is concerned, in contemporary terms, also with the *self-knowledge* domain, establishing that the transcendental subject cannot be the object of neither knowledge nor identification because it is not a phenomenon manifesting itself in time and space. For this reason, Kant's *I as subject* is an empty form. However, perhaps less plainly, this thesis is an epistemological conclusion gained from a metaphysical reflection on the transcendental subject's features, and rests on a different theoretical level than Wittgensteinian reflections (cf. Carl 1997). It is therefore necessary to move to a different level that is empirical apperception articulated by Kant from the sensitive dimension of receptivity.

5

Like Wittgenstein, Kant (1798, 135) also introduces the 'I as subject' and 'I as object' on the basis of the distinction between transcendental and empirical apperception but, as mentioned, their theoretical uses are completely different. Such distinction determines that the subject can (re)present herself in two ways: through the *I* that thinks and through the *I* that intuits itself. In another passage Kant (AA 28, 224; cit. in Carl 1997, 156) states that "the I can be taken in a twofold manner: I as human being and I as intelligence. I in the first sense means: I am an object of the inner and the outer sense. I in the second sense means that I am the object of the inner sense only." Obviously, this does not imply that there are 'two's I', on the contrary, the "I as a thinking being am one and the same subject with myself as a sensing being" (Kant 1798, 142).

From the angle of transcendental dimension, and abstracting from any modality of intuition, while *I* is a pure representation, transcendental apperception does not render the thinking subject neither as noumenon nor as phenomenon: "I think myself only as I do every object in general from whose kind of intuition I abstract" (*KrV* B 429).

From the angle of the empirical dimension, considering the *I think* as an empirical proposition equivalent to the *I exist thinking*, there is no logical function any longer but only the determination of the subject at the level of existence, the object of intuition that necessarily involves inner sense: the *I as object* of the perception is revealed by empirical apperception as a *phenomenon* that unfolds through the form of time. On the side of receptivity, at first glance the consciousness of the self as object of perception appears variable:

> The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances and is customarily called inner sense or empirical apperception. (*KrV* A 107)

Nonetheless, if the Kantian *I* as subject of thinking, i.e. transcendental apperception, is the condition of possibility of all judgments, only on empirical level of investigation some characterizations of the identification mechanisms of the egological dimension may be added to one another: empirical apperception involves the pure forms of sensibility that articulate, in the specific terms of transcendentalism, the Wittgensteinian uses of the *I as subject* and *I as object* in judgments expressing self-ascriptions of mental and physical properties. The argument can be articulated as follows:

1) *In primis*, also as regards empirical apperception, Kant rejects the possibility to move from the inner perception of something existing as thinking (what Kant calls 'eine unbestimmte empirische Anschauung') to the determination of this very something as existing substance in time and space, the forms of inner and outer sense through which all phenomena are given; differently, this would be no thought but matter. Indeed, the consciousness of the self as contemplated by empirical apperception is the inner perception of something that is not the object of outer sense (cf. Kant 1783, 334).

2) Even so, according to the arguments drawn from *Refutation of Idealism*, it is necessary to introduce the external sense on another plane: for the subject to determine its existence in time, it is necessary to assume the existence of objects perceived by outer sense, starting with the subject's very body.[5] Thus, following Capozzi (2007), in the consciousness of the self lying on the empirical determination of inner perception – i.e., on the subject's capability to perceive himself, principally in the paradigmatic instance of the psychological mechanism of attention as something that thinks while apprehending representations in the psychological flow towards the outside – *I* reveals itself as an intuition and phenomenises, *de facto* obtaining indirectly a persistence which exceeds the afore-mentioned variable nature of empirical apperception.

3) This does not imply that the subject phenomenizing in empirical apperception, and which requires outer sense in order to be determined in time as inner phenomenon, can account for an instance of identification. Just as the *I as subject* in transcendental apperception poses no question of identification – it is neither a phenomenon nor a noumenon – the *I as object* of empirical apperception cannot account for any instance of identification: it is a phenomenon that unfolds in time only. What is more, the *temporal* nature of experience, in the absence of space, allows us to count numbers but not to identify objects.

[5] Taking the cue from Allison (2004, 298), who maintains that in the framework of *Refutation of idealism* "one's body functions as the enduring object, with reference to which one's existence is determined in time", several scholars, such as Cassam (1993) and Hanna (2000), have questioned the notion of 'embodied subject' in Kant's reflection. The historical-critical reconstruction by Capozzi (2007) clarifies some aspects of the issue.

6

According to Kant, it is possible to refer to *Erkenntnis*, which is always *discursive,* only as regards the product of the application of conceptual forms to forms of sensibility, and only with reference to both pure forms of sensibility: time and space. If in the first pure intuition all possible representations are revealed, in space there appears only a specific sub-class of representations, i.e., those referring to sheer *external* objects. Thanks to space itself – the form of outer sense – the objects are represented as something real and different from the subject. Only through a spatio-temporal collocation the object is knowable in the strict sense, and the relative representations become *Erkenntnis*.

As to the empirical apperception features, the activity of thinking (*I exist thinking*) only manifests itself in time, not space, as has been said: it follows that the representation *I* cannot turn into knowledge, since, paradoxically, this would require an intuition in space, i.e., in the form of sensibility wherein the only representations refer to what is represented as something different from the subject. For this reason, the subject cannot determine itself as object within inner sense (see *KrV* A 22-3/B 37).

On the basis of the instances of distinction of time and of the re-identification of space – and, obviously, of the application of the conceptual dimension – not only is it possible to count numbers but also objects, i.e., it is possible to know a world of objects, different from the subject, which bear properties and are endowed with identification conditions. On the other hand, time, without space, represents phenomena as belonging to the thinking subject's internal sphere. For this reason, from the empirical dimension, the *I think* "expresses an indeterminate empirical intuition, i.e., a perception" (*KrV* B 423 n.). Such is an *empirical* intuition because the empirical-existential proposition *I think/I exist* lies on a sensation which indeed belongs to sensibility and reveals itself only in time. Additionally, it is *indeterminate* due to the lack of space, i.e., the form in and through which objects manifest themselves and can be determined. In this context, it is clear that, in some self-attributions, the *I* of empirical apperception presents a Wittgensteinian use of the *I as subject*.

First of all the question concerns the judgment, the *Prolegomena* (Kant 1783, 298) contains the famous distinction between *judgments of perception* and *judgments of experience*: the former have subjective validity only, while *judgments of experience* involve the principles of understanding that make *empirical judgments* objectively valid. But some judgments of perception can never become judgments of experience due to their being solely based on *subjective sensations* or *feelings* (to readapt Kant's examples: 'I'm hot in this room', 'I am disgusted by wormwood'): while a dual reference to the subject's experience and consciousness is required in order for representations to fall within the knowledgeable (cf. *KrV* A 320/B 376; Kant 1800, 33), the internal or subjective sensation (cf. Kant 1798, 156), thus equated to the *Gefühl* (cf. Kant 1790, 206), lacking an even potential reference to an object of reality – such as intuitions (in an immediate way) or concepts (in a mediate way) – is a representation exclusively connected with the subject and, for that reason, manifests itself only in time.[6]

The distinction between judgment of experience and judgments of perception has to be intended as a counterpart to the *Transcendental Deduction* (see Allison 2004, 179), and two accounts of judgment have been individuated (cf. Allison 2004; Longuenesse 1998). The first considers the act of judgment as the unification of distinct representation in a concept that is correlated with a unity in the consciousness. Since "a concept is never immediately related to an object, but is always related to some other representation of it (whether that be an intuition or itself already a concept)", the judgment is the mediate cognition of an object, the representation of a representation of it (see *KrV* A68/B93). Therefore, the judgment includes two concepts related both to each other and to object judged about (cf. *KrV* A68-69/B93-94).

The second account focuses on the objectivity: in *KrV* (B §18-19) Kant points out the distinction between an objective unity of self-consciousness, which presupposes the use of categories, and subjective unity, which is only the product of the reproductive imagination. For this reason he criticizes the logicians' definition of judgment as the "the representation of a relation between two concepts", since they don't specify what this relation amounts to and the rule of the copula: "That is the aim of the copula *is* in them: to

[6] In another passage Kant (1783, 334 n.) reaffirms that *I*, as a representation of apperception, is no concept: "it is nothing more than a feeling of an existence without the least concept, and is only a representation of that to which all thinking stands in relation (*relatione accidentis*)."

distinguish the objective unity of given representations from the subjective" (*KrV* B 142).

From a Kantian point of view, and with regard to the forms of sensibility and understanding, a judgment such as 'I'm hot' doesn't involve an identification component since it consists of perceptions revealed in time only, whose nexus cannot involve any pure concept of understanding, and therefore the objective unity of apperception, because there is no object directed to the outer sense to be determined and, therefore, identified.

The heart of the matter changes in the self-ascriptions of properties manifesting themselves in time and space which thus refer to a spatio-temporally located subject/body. If the thinking being, as a human being, is at the same time an object of outer sense (cf. *KrV* B 415), from a Kantian point of view, a judgment such as 'I have a dirty hand' is a *judgment of experience* in that it involves an outer object – the body – and the principles of pure understanding, which are based on the categories applied to the formal conditions of a possible intuition: the nexus of the representations is necessarily produced by the presence of the object which affects sensibility and is objectively valid through the intervention of the intellectual dimension thus determining the object.

Obviously, this concerns the conditions making a judgment of experience possible based on the relations between forms of thoughts and forms of sensibility considered as rules which, according to the Copernican revolution, represent the universal laws without which nature in general, as object of sense, could not be thought (cf. Kant 1790, 183). It goes without saying that the transcendental account cannot tell whether the judgment 'I have a dirty hand' is true or false, it provides the conditions of possibility required to produce an assertive *Erkenntnissatz*, which can be confirmed or denied, with regard to the attribution of the predicate, since the hand of the person having made such judgment may not be dirty, as well as to the subject's identification. If the judgment is produced on the basis of the reflected image of a tangle of hands in a mirror, the dirty hand might not belong to the person who has produced the judgment but, rather, to someone else. In this context, a *judgment of experience* expressing the self-attribution of a subject/body physical property involves a subject's identification component as, in point of fact, the subject can be mistaken as to whether he is the one who is attributing himself this particular property.

Therefore, only with the objective unity of apperception it is possible to make a truth-evaluable judgment (i.e. a judgment of experience) that contains a identification component and consequently makes *Erkenntnis* possible; instead a judgment that doesn't involve the objective unity of apperception (i.e. a judgment of perception) is not truth-evaluable, so doesn't include an identification component.

7

In conclusion, following Kitcher (2000, 34) it is true that Kant and Wittgenstein start from what at first glance might appear to be different theses so that 'it is an interpretive and philosophical mistake to try to force an alliance between what are, in fact, deeply opposed camps'. At the same time, following Carl (1997, 149) it is true that Kant's distinction of the 'I as subject' and the 'I as object' is not concerned with Wittgenstein's project of distinguishing different kinds of predicates to be ascribed to oneself, but is concerned with an epistemological perspective focusing on the distinction between spontaneity and receptivity, seen as conditions of all possible knowledge, in order to give an account which incorporates the *I as subject* of apperception into the foundation of the formal conditions of knowledge.

However, it is also true that, even seen via different philosophical horizons, they both come to similar conclusions: the *I used as subject* concerns the self-attributions of mental properties and involves no instance of identification, while the *I used as object* concerns the self-attributions of physical properties and the subject's identification will be provided.

As already seen, the question is further articulated in the transcendental and empirical dimension and can be connected to Carl's (1997, 157) distinction between two classes of self-ascriptions, the kind of self-ascriptions that considers the 'I as passive' related to receptivity and the other sort of self-ascriptions that regards the 'I as active' related to spontaneity:

a) Since it is necessarily involved in the making of any judgment, the *I* as *subject* of thinking, which means pure apperception, is the condition of possibility of the Wittgensteinian uses of the *I as subject* and *I as object* in judgments expressing self-ascriptions of mental and physical properties: as already remarked, although the representation *I* designates the transcendental subject, it cannot be determined to identify the thinking entity and thus always lacks the identification component.

b) For the *I* as passive of empirical apperception, there are two possibilities: b.1) if the subject reveals itself only in time, the self-attributions concern only mental properties and so there is no question of identification, the *I* is used as subject in judgment of perception; b.2) if the subject reveals itself in time and space, the question of identification arises since there is an *explicit self-attribution* of body physical property relative to *I* which is used as object in a judgment of experience. This Kantian reading is also obtained through a reflection on the sources of epistemic sensibility and mediation of the forms of time and space and, also for this reason, it must be included in a very different theoretical framework from Wittgenstein's.[7]

References

ALLISON, H. (2004): *Kant's Transcendental Idealism*. Revised and Enlarged Edition. New Haven – London: Yale University Press.

ANSCOMBE, G.E.M. (1975): The First Person. In: Guttenplan, S.D. (ed.): *Mind and Language*. Oxford: Clarendon Press.

BECKER, W. (1984): *Selbstbewusstsein und Erfahrung: zu Kants transzendentaler Deduktion und ihrer argumentativen Rekonstruktion*. Freiburg: K. Alber.

BERMÚDEZ, J.L. (1994): The Unity of Apperception in the Critique of Pure Reason. *European Journal of Philosophy* 2, 213-240.

BROOK, A. (2001): Kant, Self-awareness and Self-reference. In: Brook, A. – DeVidi, R.C. (eds.): *Self-reference and Self-awareness*. Philadelphia: John Benjamins.

CARL, W. (1997): Apperception and Spontaneity. *International Journal of Philosophical Studies* 5, No. 2, 147-163.

CAPOZZI, M. (2007): L'io e la conoscenza di sé in Kant. In Canone, E. (ed.): *Per una storia del concetto di mente*. Firenze: Olschki.

CASSAM, Q. (1993): Inner Sense, Body Sense, and Kant's Refutation of Idealism. *European Journal of Philosophy* 1, 111-127.

CASSAM, Q. (1997): *Self and World*. Oxford: Oxford University Press.

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

GEACH, P. (1957): *Mental Acts. Their Content and Their Objects*. London: Routledge.

HACKER, P.M.S. (1972): *Insight and Illusion*. Oxford: Clarendon Press.

HANNA, R. (2000): The Inner and the Outer: Kant's Refutation Reconstructed. *Ratio* 2, 146-174.

[7] I would like to give particular thanks to Pasquale Frascolla for his valuable suggestions.

HOWELL, R. (2000): Kant, the I Think, and Self-Awareness. In: Cicovacki, P. (ed.): *Kant's Legacy*: *Essays in Honor of Lewis White Beck*. Rochester: University of Rochester Press.

KANT, I. (1900 ff): *Gesammelte Schriften* Hrsg.: Bd. 1–22 Preussische Akademie der Wissenschaften, Bd. 23 Deutsche Akademie der Wissenschaften zu Berlin, ab Bd. 24 Akademie der Wissenschaften zu Göttingen. Berlin.

KANT, I. (1781-87/1997): *Critique of Pure Reason* (*KrV*). Cambridge: Cambridge University Press.

KANT, I. (1783/1997): *Prolegomena to any Future Metaphysics*. Cambridge: Cambridge University Press.

KANT, I. (1790/2000): *Critique of the Power of Judgment*. Cambridge: Cambridge University Press.

KANT, I. (1786/2004): *Metaphysical Foundations of Natural Science*. Cambridge: Cambridge University Press.

KANT, I. (1798/2006): *Anthropology from a Pragmatic Point of View*. Cambridge: Cambridge University Press.

KANT, I. (1800): *The Jäsche Logic*. In: Kant, I. (1992): *Lectures on Logic*. Cambridge: Cambridge University Press.

KITCHER, P. (2000): On Interpreting Kant's Thinker as Wittgenstein's 'I'. *Philosophy and Phenomenological Research* 61, 33-63.

KITCHER, P. (2011): *Kant's Thinker*. New York: Oxford University Press.

LONGUENESSE, B. (1998): *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Princeton: Princeton University Press.

LONGUENESSE, B. (2006): Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian theme. *Philosophical Topics* 34, 283-309.

LONGUENESSE, B. (2012): Two Uses of 'I' as Subject? In: Prosser, S – Recanati, F. (eds.): *Immunity to Error through Misidentification*. Cambridge: Cambridge University Press.

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

PEACOCKE, C. (2008): *Truly Understood*. Oxford: Oxford University Press.

PEACOCKE, C. (1999): *Being Known*. New York: Oxford University Press.

PERRY, J. (1986): Perception, Action, and the Structure of Believing. In: Grandy, R. – Warner, R. (eds.): *Philosophical Grounds of Rationality*. Oxford: Oxford University Press.

POWELL, C.T. (1990): *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press.

RECANATI, F. (2009): De Re and De Se. *Dialectica* 63, 249-269.

RECANATI, F. (2007): *Perspectival Thought. A Plea for (Moderate) Relativism*. New York: Oxford University Press.

SHOEMAKER, S. (1968): Self-Reference and Self-Awareness. *Journal of Philosophy* 65, 555-567.

SLUGA, H. (1996): Whose House is That? Wittgenstein on the Self. In: Sluga, H. – Stern, D.G. (eds.): *The Cambridge Companion to Wittgenstein*. Cambridge: Cambridge University Press.

STRAWSON, P.F. (1966): *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason*. London: Methuen.

STURMA, D. (1985): *Kant über Selbstbewusstsein*. New York: Georg Olms Verlag.

WITTGENSTEIN, L. (1958): *The Blue and the Brown Books*. Oxford: Blackwell.

WRIGHT, C. (1998): Self-Knowledge. The Wittgensteinian Legacy. In: Wright, C. – Smith, B. – Macdonald, C. (eds.): *Knowing Our Own Minds*. Oxford: Oxford University Press.

Organon F 22 (2) 2015: 199-210

How Choice Blindness Vindicates Wholeheartedness

ASGER KIRKEBY-HINRUP

Department of Philosophy. University of Lund
Helgonavägen 3. 221 00 Lund. Sweden
asger.kirkeby-hinrup@fil.lu.se

RECEIVED: 12-06-2014 ▪ ACCEPTED: 17-02-2015

ABSTRACT: Recently the account of free will proposed by Harry Frankfurt has come under attack. It has been argued that Frankfurt's notion of wholeheartedness is in conflict with prevalent intuitions about free will and should be abandoned. I will argue that empirical data from choice blindness experiments can vindicate Frankfurt's notion of wholeheartedness. The choice blindness phenomenon exposes that individuals fail to track their own decisions and readily take ownership of, and confabulate reasons for, decisions they did not make. Traditionally this has been taken to be problem for the notion of free will. I argue that Frankfurt's account does not face this problem. Instead, choice blindness can be fruitfully applied to it, and vice versa. Frankfurt's notion of wholeheartedness, I suggest, delineates the range of the choice blindness effect. This makes wholeheartedness a useful meta-theoretical concept for choice blindness research. I conclude that, *pace* the recent criticism, wholeheartedness is a useful notion and should not be abandoned.

KEYWORDS: Choice blindness – decisions – free will – Harry Frankfurt – wholeheartedness.

1. Introduction

Elsewhere (see Kirkeby-Hinrup 2014) I have objected to Harry Frankfurt's account of free will (see Frankfurt 1971; 1988). The objection is that by grounding free will in the notion of wholeheartedness, Frankfurt allegedly renders the notion of free will sparse, practical and occasional. It becomes sparse because wholehearted identification with a choice rarely obtains in most everyday deliberative situations. It becomes practical because free will essentially depends on particular actions in a context. It becomes occasional because on this view free will only obtains on specific occasions, rather than being something a person can have across situations. I argue that a conception of free will that is sparse, practical and occasional is in conflict with our intuitions: Intuitively we consider free will as possessed by individuals across time. Therefore Frankfurt's notion of wholeheartedness is inadequate. In this article, I will take another perspective and instead of arguing from intuition I take my starting-point in experimental data. I will argue that experiments on choice blindness (cf. Hall – Johansson – Strandberg 2012; Hall – Johansson – Tärning – Sikström – Deutgen 2010; Hall et al. 2013; Johansson – Hall – Sikström – Olsson 2005; Johansson – Hall – Sikström – Tärning – Lind 2006) provides indirect support for Frankfurt's notion of wholeheartedness. In the next section, I will provide a brief introduction to the account of free will developed by Harry Frankfurt. Section three is an introduction to the phenomenon of choice blindness. In section four, I apply the notion of wholeheartedness to the choice blindness phenomenon. This will show that, not only are the two compatible, but from a theoretical point of view they complement each other. In the conclusion, I clarify why my previous critique of the notion of wholeheartedness is less serious than it seemed. Then I explain why the phenomenon of choice blindness is not a threat to accounts of free will grounded in the notion of wholeheartedness.

2. Wholeheartedness and free will

According to Frankfurt, free will is tied to higher-order desires (cf. Frankfurt 1971; 1988). Higher-order desires are desires about (lower-order) desires. Higher-order desires are desires about those desires the individual wishes to be effective. Frankfurt's theory addresses the distinction between simply desiring something and wanting to so desire it. According to Frankfurt, when an individual identifies herself with one of her higher-order desires, such identification will be followed by a higher-order volition. The effect of the higher-order volition is that the individual wishes her particular desire (the target of her higher-order desire) to be her *effective* desire. An effective desire is a desire that moves her all the way to action. For instance, I may occasionally have a desire for excessive amounts of ice cream. But, for reasons pertaining to health, I rarely wish to act on this desire. Thus, my higher-order desires do not endorse the desire for ice-cream. Consequently, I do not wish this desire to move me to action. Conversely, I have a desire to work out and stay in shape. While this desire may be weak, it is endorsed by my higher-order desires. I wish this desire to be effective, i.e., I wish that I act upon this particular desire. According to Frankfurt, the will is free when the individual acts in accordance with her higher-order volitions.

One objection that is usually leveraged against philosophical accounts invoking higher-order notions is the threat of regress. For Frankfurt's account, the threat of regress pertains to higher-order desires. Is it not possible that a second-order desire could be in conflict with a desire on a level above itself (e.g. a third-order desire)? Why should we not say that if this is the case, then following the higher-order volition, spawned by the third-order desire, is what is necessary for free will? Similarly should we then not say that the fourth-order desires may trump the third-order desires, and so on *ad infinitum*? The problem is that we can always posit a desire of a higher-order. This objection trades on the intuition that the higher the order of a desire, the closer it is to what an individual *really wants*. If such regress is allowed, Frankfurt's account can never get off the ground. The problem is that there is no support for a claim that any particular desire is the relevant one, whose corresponding higher-order volition should ground free will.

To solve the regress problem an account is needed of how an individual comes to identify herself in the right manner with a particular higher-order desire. Frankfurt answers by introducing his notion of wholeheartedness. Frankfurt says:

> When a person identifies himself *decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders. [...] The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked. (Frankfurt 1971, 16 italics from original)

What Frankfurt claims is that a wholehearted identification with a higher-order desire solves the problem. When a higher-order desire resounds throughout the system any question of further higher-order desires is unnecessary. Once the individual has wholeheartedly identified herself with a given desire, she cannot help but to want this desire to be effective. So only when a wholehearted identification with a particular desire occurs is a relevant higher-order volition formed.

To summarize, there are two requisites for free will on Frankfurt's account. The first requisite is that an individual identifies wholeheartedly with some desire, thus forming a higher-order volition. The second requisite is that the eventual decision and action of the individual are in accordance with the higher-order volition. According to Frankfurt, only when both requisites are satisfied, the individual has free will.

3. Choice blindness

Choice blindness experiments expose that people have difficulties keeping track of their decisions. In choice blindness paradigms the individual makes a choice. After making the choice the individual is asked why she preferred the chosen option over the alternative(s). However, in the experimental manipulations the individual is presented with an alternative she in fact did not choose, *as if* she had in fact chosen it. The choice blindness effect is that subjects rarely detect this manipulation. Instead, subjects will confabulate reasons for preferring the option they did not in fact choose. For instance, in one of the early choice blindness studies (see Johansson et al. 2005), subjects are presented with two pictures of individuals of the opposite sex. The subjects are instructed to point to the picture of the individual they find most attractive. After the choice is made (and the subject has pointed), the pictures are placed face down on the table. In the baseline condition, the chosen picture is then picked back up and shown to the subject, while the other picture is still face down on the table. The subject is then asked to describe why she preferred the individual she pointed to, and proceeds to provide an introspective account of her reasons. The experimental condition proceeds in exactly the same way as the baseline condition with the exception that the picture that is picked back up is actually the picture the individual did *not* choose.[1] In the experimental condition the subject will normally confabulate reasons for preferring the individual in the picture she is shown.

Since this early study, the choice blindness effect has been demonstrated in a wide range of domains. The various domains in which choice blindness has been demonstrated include the political, moral, aesthetic, gustatory, and olfactory (cf. Hall et al. 2012; Hall et al. 2010; Hall et al. 2013; Johansson et al. 2005; Johansson et al. 2006). Thus, it appears we are blind to the outcomes of a wide range of decisions we normally, and intuitively, take to be important to us. The fact that the choice blindness effect has been demonstrated in different sensory modalities, and across important social domains, such as politics and morals, underscores the pervasiveness of the phenomenon.

Now, it can be argued that choice blindness presents a problem for free will. This is the argument: Given that the choice blindness paradigm has shown that in several domains we are blind to the outcome of our decisions and furthermore that the reasons we provide for those decisions are generated *post hoc*, this may very well be the case with every decision we make. On this interpretation of the outcome of the choice blindness experiments, choice blindness threatens traditional notions of free will because reasons and decisions become epiphenomenal, and thus are outside of conscious control. If the outcome of our choices can be replaced without us noticing, and we readily take ownership of decisions we did not make, this might undermine popular conceptions of free will. Specifically it might undermine the claim that the reasons and deliberations we consciously experience before making a choice have an impact on the choices we end up making. The fact that we experience that our deliberations and decisions matter and are efficacious in our ordinary lives merely may be an illusion.[2] This suggestion in itself is nothing new (see, e.g. Dennett 1984; 2004; Wegner –

[1] Due to some sleight of hand and the setup of the scenario it appears to the subject as if the experimenter is picking up the correct picture.

[2] This experience is neatly demonstrated within the choice blindness paradigms as well. When individuals who have just participated in a choice blindness experiment are suggested that perhaps their choices had been switched (which they indeed had), many deny this as impossible and something they would certainly have noticed. This overestimation of own introspective competence has been called *Choice Blindness Blindness* (see Johansson et al. 2005; 2006). Given the choice blindness blindness phenomenon such a naïve faith in one's own capacity to keep track of decisions is clearly mistaken.

Wheatley 1999; Wegner 2002; 2003), but the choice blindness phenomenon *prima facie* provides such a view with more traction. Many accounts (e.g. those of Dennett and Wegner) argue that the experience of free will does not correspond to (let alone entail) any feature of human decision making that corresponds to the experience. These accounts suggest that the notion of free will is either in need of substantial revision (in order to remove the traditional connotations to anything substantively free), or must be abandoned completely.[3] The kind of empirical data that forms the basis of such accounts is especially problematic for theories of free will that rely heavily on explicit cognitive processes such as deliberation and introspection.

One option for the proponents of free will is to propose alternative interpretations of the empirical data. In brief, they may suggest that the fact the individuals cannot remember their reasons, entails neither that they did not have any, nor that the ones they did have were inefficacious with respect to their decision. While this defense may be tenable I will not pursue it in the present context. Rather, the objective here is to show that accounts of free will which do not place an emphasis on conscious deliberation, such as Frankfurt's, can sidestep objections based on such empirical data. It will be argued that such accounts can claim that the individual becomes aware of which of her desires are endorsed by higher-order desires by an automatic process. Only the result of the process will be known to the subject. Arguably it is not even necessary that the individual explicitly experiences 'wholeheartedness' consciously. An (unconscious) occurrence of a wholehearted identification would be sufficient.[4] Furthermore, whether

[3] How such revision would influence the view of humans as *free agents* is a separate question. Interestingly, the experience of agency face problems that are similar to the ones choice blindness pose for the experience of free will (see, e.g., Bayne – Levy 2006; Moore – Wegner – Haggard 2009).

[4] Note that this does not entail that it is futile for the individual to attempt to determine which of her desires she wants to be effective. Wholeheartedness reasonable should ordinarily reveal itself upon inspection. It might be tempting to think that the individual must be conscious of the wholehearted identification in order for this to matter for the decision. However, there are two reasons to hesitate in making this claim. The first reason is that the way Frankfurt explicates his theory, whether the wholehearted identification is conscious, does not seem to matter to whether the individual can exhibit free will. The second is that invoking consciousness of whole-

this process is fully determined is inconsequential to Frankfurt's account of free will. Moreover, there is no requirement that the process must be infallible. That the process may occasionally falter and the individual mistakenly believe that she wants a given desire to be effective is unimportant. This would merely be an instance of those occasions where the individual does not exhibit free will. Because such occasions are allowed on Frankfurt's account, they do not pose a problem.

4. Wholeheartedness and choice blindness

One question about choice blindness that looms large in the background is what kinds of choices can be successfully manipulated. Are there choices the experimenters cannot make subjects believe themselves to have made, and subsequently will not attempt to justify? Surely, intuition suggests, it is impossible to switch a bride at the altar thus tricking the subject into marrying the wrong individual (this intuition is shared by the choice blindness experimenters; see, e.g., Hall et al. 2010). Even more absurd is the idea that the tricked individual would subsequently take ownership of, and justify, the decision to marry the un-intended spouse. If we agree that this is absurd, as I think we should, it follows that there are limits to the choice blindness effect. Simply put, we cannot be tricked to believe we made all choices we are presented with as our own. Can we say anything about these limits?

To investigate the limits of choice blindness is to investigate what characterizes the choices that are immune to the choice blindness manipulation. From Frankfurt's perspective, making his position our own, we can answer this question by deploying the notions of wholeheartedness and higher-order volitions. Because higher-order volitions are about those desires the individual wants to be effective, it seems reasonable to expect that they also track the desires they are about in order to determine if those desires actually *are* effective. How might one argue for the view that such tracking occurs?

heartedness as a necessary aspect, pulls the theory in the direction of the cognitivist accounts of free will, and makes it susceptible to objections from the empirical data that haunt these.

One possibility is the following. Because wholeheartedness, according to Frankfurt, resounds throughout the system, it appears that it cannot be fleeting. To avoid allegations that the notion of wholeheartedness is not robust enough, it seems important for the theory to show how wholeheartedness may persist over time. Now consider the core idea Frankfurt espouses: a higher-order volition is generated when the individual wholeheartedly *identifies* herself with a decision. This idea suggests that there is a link between wholeheartedness and personal identity, in the specific sense of the identity over time provided by the characteristics through which an individual defines herself as herself. I maintain that this link provides the basis and motivation for wholeheartedness to persist over time. For instance, a sincere vegan might take central parts of her identity to consist in particular views on food and related issues.

One might object here that some individuals appear to constantly redefine themselves, while sincerely professing that "this is the new me". If this is the case, the objection goes, the kind of personal identity referred to above cannot provide the basis for wholeheartedness to persist over time. However, even in such extreme cases, wholehearted commitment to decisions first, does not change on the timescale involved in choice blindness experiments, and second typically occurs in the form of an explicit decision to change one's mind as opposed to how people change their opinion in the change blindness cases. Furthermore, there is nothing in the notion of wholeheartedness that entails that it can never change. People do change over time. There are beliefs and activities I wholeheartedly endorsed at the age of sixteen I no longer care about. At the very least, the link between wholeheartedness and personal identity provides sufficient reason to believe that individuals can be wholeheartedly committed to the same views or decisions for the entire time a choice blindness experiment runs.

Having made the case that wholeheartedness is geared to persist over time we may now address the question of whether wholeheartedness tracks the outcome of decisions. One way to show that this is true would be to perform choice blindness experiments on subjects who strongly identify with particular views. The idea is that with respect to those views such a subject would be immune to the choice blindness manipulation exactly because wholeheartedness is tied to personal identity. For instance, it seems likely that a professional politician located on the far left of the political spectrum would immediately object if she were presented with a choice where she supposedly had come out in favor of tougher immigration laws. While it is true that the choice blindness experiments pertaining to political views (i.e. Hall et al., 2013) showed that people on the street were susceptible to the choice blindness effect in this domain, it is unlikely that professional politicians are susceptible to the same extent. This is unlikely exactly because these views normally form a significant part of the how politicians conceive of themselves.

To sum up, it seems that the link between wholeheartedness and personal identity supports the idea that in some cases subjects track their desires and as a consequence the outcome of their decisions. These cases are such that the subject wholeheartedly endorses a particular view or course of action. However, so far choice blindness experiments have not been carried out to test this. To strengthen the case, we turn to another piece of evidence from everyday life that indicates that people tracking their desires and also the outcomes of their decisions.

In our everyday observations of ourselves we often realize when we fail to do what we actually wanted to do. How might this be if we were not keeping track of what we wanted to do? One might counter that such realizations are based on retroactively constructed decisions. However this does not appear to make sense. For instance, taking an evolutionary perspective, what selective pressure would result in a system that retroactively confabulated decisions it failed to follow through on? What would the benefit be? Conversely, it makes sense to track the outcomes of decisions, e.g. for purposes of error detection, learning, and behavioral optimization.

That people occasionally fail to do what they set their mind to is central to Frankfurt's account. The distinction between occasions where we succeed and where we fail to do what we wanted to delineates the occasions where we act on a desire that is not endorsed by a higher-order volition and the occasions where we succeed in acting on the desires we want to be effective. Imagine that I have decided that I really want to work out. To my dismay, I find myself having bought excessive amounts of ice-cream instead. The dismay is an indication that tracking did occur and that another desire was realized than the one that the higher-order desire concerned. I am dismayed because I failed to make the desire my higher-order volition was about (working out) my effective desire. The desire that carried me all the way to action (eating ice-cream) was not one endorsed by a higher-order desire; it was not one of my concerns.

In this way wholeheartedness demarcates the limit for the choice blindness effect. When I wholeheartedly identify myself with a desire, the higher-order volition tracks or locks on to the desire, and indirectly monitors the outcome of my action. This thwarts the choice blindness effect. We are now in a position to explain the choice blindness effect as occurring whenever decisions are not wholehearted. Because higher-order volitions track their desires, when the outcome concerns another desire, choice blindness occurs. Furthermore, only wholeheartedness generates higher-order volitions, which means that when wholeheartedness is absent there is no tracking going-on and no desire that constitutes a concern for the subject. Consequently, the manipulation of the subject's choice is successful (i.e. not detected) and choice blindness occurs.

5. Conclusion

One might think that intuitively, free will is relevant whenever we deliberate. However, on Frankfurt's account, only the decisions endorsed by a resounding wholeheartedness can be instances of free will. Since the experience of a resounding wholehearted commitment to a given desire or decision is not something that occurs regularly, this means that the majority of everyday decisions are made without free will. While the intuition mentioned above may be one that many agree with, this is not enough to reject Frankfurt's account. Indeed, the fact that the notion of wholeheartedness meshes very well with – and can be fruitfully applied to – empirical data from choice blindness, suggests the intuition might be wrong. If the empirical data show that there is no free will in many trivial choices and across domains, then this suggests the intuition may be wrong. If the data indicate that the intuition may be wrong, then the intuition seems ill suited as the foundation for criticism of Frankfurt's account. On the other hand, seeing the notion of wholeheartedness in light of the phenomenon of choice blindness vindicates the idea that free will might indeed be sparse, practical and occasional (*pace* Kirkeby-Hinrup 2014). Conversely, seeing choice blindness in light of the notion of wholeheartedness suggests that the former not necessarily is a threat to free will. More specifically, it is not a threat to free will of the kind proposed by Harry Frankfurt. Moreover, while it is likely to be difficult to operationalize empirically, wholeheartedness provides a useful meta-theoretical concept to delineate the limits of choice blindness.

References

Bayne, T. – Levy, N. (2006): The Feeling of Doing: Deconstructing the Phenomenology of Agency. In: Sebanz, N. – Prinz, W. (eds.): *Disorders of Volition*. Cambridge (Mass.): Cambridge University Press.

Dennett, D.C. (1984): *Elbow Room: The Varieties of Free Will Worth Wanting*. The MIT Press.

Dennett, D.C. (2004): *Freedom Evolves*. Penguin.

Frankfurt, H. (1971): Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68, No. 1, 5-20.

Frankfurt, H. (1988): Identification and Wholeheartedness. In: *The Importance of What We Care About*. Cambridge University Press, 159-176.

Hall, L. – Johansson, P. – Strandberg, T. (2012): Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-transforming Survey. *PloS one* 7, No. 9, e45457.

Hall, L. – Johansson, P. – Tärning, B. – Sikström, S. – Deutgen, T. (2010): Magic at the Marketplace: Choice Blindness for the Taste of Jam and the Smell of Tea. *Cognition* 117, No. 1, 54-61.

Hall, L. – Strandberg, T. – Pärnamets, P. – Lind, A. – Tärning, B. – Johansson, P. (2013): How the Polls Can Be Both Spot on and Dead Wrong: Using Choice Blindness to Shift Political Attitudes and Voter Intentions. *PloS one* 8, No. 4, e60554.

Johansson, P. – Hall, L. – Sikström, S. – Olsson, A. (2005): Failure to Detect Mismatches between Intention and Outcome in a Simple Decision Task. *Science* 310, No. 5745, 116-119.

Johansson, P. – Hall, L. – Sikström, S. – Tärning, B. – Lind, A. (2006): How Something Can Be Said about Telling More than We Can Know: On Choice Blindness and Introspection. *Consciousness and Cognition* 15, No. 4, 673-692.

Kirkeby-Hinrup, A. (2014): How to Get Free Will from Positive Reinforcement. *SATS, Northern European Journal of Philosophy* 15, No. 1, 20-38.

Moore, J.W. – Wegner, D.M. – Haggard, P. (2009): Modulating the Sense of Agency with External Cues. *Consciousness and Cognition* 18, No. 4, 1056-1064, available at: http://dx.doi.org/10.1016/j.concog.2009.05.004

Wegner, D.M. – Wheatley, T. (1999): Apparent Mental Causation: Sources of the Experience of Will. *American Psychologist* 54, No. 7, 480.

Wegner, D.M. (2002): *The Illusion of Conscious Will*. The MIT Press.

Wegner, D.M. (2003): The Mind's Best Trick: How We Experience Conscious Will. *Trends in Cognitive Sciences* 7, No. 2, 65-69, doi: 10.1016/s1364-6613(03)00002-0.

Organon F 22 (2) 2015: 211-228

K analýze deontických modalít v Transparentnej intenzionálnej logike[1]

DANIELA GLAVANIČOVÁ

Filozofická fakulta. Univerzita Komenského v Bratislave
Gondova 2. 814 99 Bratislava. Slovenská republika
dada.baudelaire@gmail.com

ZASLANÝ: 12-10-2014 ▪ AKCEPTOVANÝ: 04-03-2015

ABSTRACT: The aim of this paper is to outline a suitable analysis of certain *deontic modalities*. To avoid confusion as much as possible, I specify the subject-matter of my analysis explicitly. Subsequently, the paper argues that Transparent Intensional Logic (TIL) is an appropriate framework for developing deontic logic. The main contribution of the paper consists in a proposal concerning the analysis of deontic modalities in TIL as well as in offering a semantically based distinction between implicit and explicit deontic modalities. Finally, I introduce some definitions along with some inferential rules and show (using Ross' paradox) how it is possible to deal with the paradoxes of deontic logic in terms of my analysis.

KEYWORDS: Analysis – deontic modalities – descriptive versus prescriptive – implicit versus explicit – paradox – Transparent Intensional Logic.

1. Svet príkazov, zákazov a dovolení

Rodíme sa do sveta príkazov, zákazov a dovolení: Pokúšame sa riadiť právnymi, etickými či náboženskými normami, pravidlami rodiny či príkaz-

[1] Ďakujem Mariánovi Zouharovi za množstvo užitočných rád, námietok a pripomienok k všetkým predchádzajúcim verziám tejto state. Ďakujem aj M. Duží, F. Gahérovi, M. Kostercovi, P. Maternovi, J. Raclavskému, V. Svobodovi a anonymným recenzentom za cenné pripomienky, námietky či diskusiu.

mi nadriadených v práci. Časť prirodzeného jazyka, ktorá sa týka príkazov, zákazov a dovolení, budem nazývať termínom *deontický diskurz*. Časťou tohto diskurzu sú deontické slová a ich predteoretické významy, tzv. *deontické modality*. Deontické slová sa ďalej vyskytujú v deontických vetách a tým zase zodpovedajú nejaké predteoretické významy. Mojím hlavným cieľom je *analýza* deontických modalít, čím sa otvorí cesta k analýze deontických viet (a ich predteoretických významov) a k určeniu zodpovedajúcich *axióm* a *pravidiel odvodzovania*. Splnením týchto úloh dostaneme sémantický model, ktorý však nebude nikdy totožný s tým, čo modeluje (inak by nešlo o model; pozri Bielik – Kosterec – Zouhar 2014, 112).

2. Predmet analýzy

Aby sa predišlo zbytočným nedorozumeniam, začnem dôsledným vymedzením predmetu analýzy. Budem analyzovať deontické slová *prikázané*, *zakázané* a *dovolené*. Ilustračný príklad nám ukáže, že sa z nich môžu utvárať vety dvoma zásadne odlišnými spôsobmi. Majme kláštor, v ktorom sa musí mlčať. Kláštorný poriadok by mohol obsahovať niektorú z týchto dvoch viet:

(1) Je prikázané, aby mnísi mlčali.

(2) Mnísi majú prikázané mlčať.

V prvom prípade sa slovné spojenie *je prikázané, aby* vzťahuje na vetu, teda funguje ako vetný operátor. Ak opustíme syntaktickú úroveň, tvrdí sa tu, že je prikázané, aby nastal určitý stav vecí (taký, v ktorom mnísi mlčia). V druhom prípade môžeme vyčleniť slovné spojenie *mnísi majú*, deskriptívne deontické slovo *prikázané* a slovo *mlčať*. Ak opustíme syntaktickú úroveň, tvrdí sa tu, že je prikázaná určitá činnosť – mlčanie.

Vety (1) a (2) majú dôležitú spoločnú črtu: obidve *tvrdia*, že je niečo prikázané. Náš kláštorný poriadok by takéto vety mohol napriek tomu obsahovať, pričom by sa opieral o nejaký zamlčaný príkaz, napríklad: Ak chceš byť naším mníchom, rob všetko, čo je podľa kláštorného poriadku prikázané, a nerob nič, čo je zakázané!

Treba si uvedomiť, že takéto deontické výroky môžu byť pravdivé či nepravdivé. Ak to čitateľovi nepripadá ako samozrejmý jazykový fakt, náš ilustračný príklad môže túto tézu zdôvodniť. Obmedzme náš svet príkazov, zákazov a dovolení na svet spomínaného kláštora. Nejaký muž, nazvime ho Pavel, chce stráviť zvyšok života v tichosti. Pricestuje do nášho kláštora a pre istotu položí správcovi (nazvime ho Richard) otázku *Je prikázané, aby mnísi mlčali?* Ak ho Richard nechce oklamať, odpovie mu kladne. Analogicky, keď pricestujeme do cudzej krajiny, môžeme sa jej obyvateľov legitímne pýtať, či je prikázané jazdiť autom v pravom cestnom pruhu prípadne či je zakázané piť alkohol na verejnosti.

Mohlo by sa samozrejme argumentovať, že by sme mohli vety (1) a (2) *interpretovať* aj preskriptívne. Povrchová štruktúra týchto viet však nijako nenaznačuje, že by malo ísť o príkazy (tieto vety nekončia výkričníkmi, ale bodkami).[2] Navyše, predstavme si, že by Pavel našiel vetu (1) či (2) napísanú len tak na chodníku či namaľovanú na stene domu. Zrejme by ju nechápal ako príkaz, a už vôbec nie ako niečo, čím by sa mal riadiť. Preskriptívna interpretácia takýchto viet je dodaná takpovediac „zvonku".

Príkazy samé sú nepochybne *preskriptívne* – priamo rozkazujú, a preto im bežne nezvykneme pripisovať pravdivostné hodnoty. Spomínaný kláštorný poriadok by mohol obsahovať príkazy a zákazy, nie ich deskripcie. V našom prípade by šlo o vetu:

(3) Mnísi, mlčte!

Veta (3) patrí k tretiemu druhu viet, ktoré by mohli byť plnohodnotným predmetom skúmania v deontickej logike.

Všimnime si, že deontické vety (1) – (3) sú jednoduché v tom zmysle, že neobsahujú nič, čomu by mali v analýze zodpovedať výrokovologické spojky či kvantifikátory. Jednoduchú deontickú vetu možno vždy *mutatis mutandis* preformulovať tak, aby jej modifikácia patrila do ľubovoľného z troch uvedených druhov deontických viet. Dôvody v prospech konkrétnej voľby sú preto predovšetkým praktické.

Predmetom mojej analýzy budú vety prvého druhu, pretože ich skúmanie považujem za najmenej problematické. Analýzu viet typu (3) problematizuje *Jørgensenova dilema*: Rozkazy nemôžu nadobúdať pravdivostnú hodnotu a keďže je tradičná definícia vyplývania založená na pojme pravdivostnej hodnoty, nemožno skúmať úsudky, ktoré obsahujú rozkazy. To je však

[2] K rozlíšeniu povrchovej štruktúry a logickej formy pozri napríklad Zouhar (2009, 20-25). Nevylučujem však možnosť argumentácie v prospech tézy, že tieto vety možno interpretovať ako preskriptívne na úrovni ich logickej formy.

v rozpore s tým, že sa niektoré z nich zdajú byť platné. Táto dilema však očividne neproblematizuje skúmanie deontických viet typu (1) a (2). Analýza viet typu (2) je navyše problematickejšia ako analýza viet typu (1), pretože kým stavy vecí môžeme syntakticky reprezentovať pomocou viet, ktoré môžu byť pravdivé či nepravdivé, činnosti môžeme reprezentovať iba pomocou výrazov, ktoré takúto vlastnosť nemajú, čo je nevýhodou pri skúmaní vyplývania.[3]

Termíny *byť prikázaný* (*zakázaný*, *dovolený*) budem používať troma odlišnými spôsobmi. Vzhľadom na predteoretický spôsob používania termínov (*byť*) *prikázaný*, *zakázaný* či *dovolený* sú prikázané, zakázané či dovolené stavy vecí, pretože sa deontické vetné operátory viažu na vety a vety sa predteoreticky vzťahujú na stavy vecí. Vzhľadom na syntaktický spôsob používania týchto termínov sú prikázané^S (zakázané^S, dovolené^S) vety, pretože ide o vetné operátory. A napokon, mojím cieľom je zistiť, čo je (resp. čo by prijateľne mohlo byť) prikázané^T, zakázané^T a dovolené^T vzhľadom na teoretický spôsob používania týchto termínov. Hľadaná entita by mala byť vhodnou explikáciou stavov vecí.

3. Teoretický rámec analýzy

Jednoduché modely majú mnoho predností a na hrubú analýzu môžu byť vhodnejšie ako tie zložitejšie. Vety prirodzeného jazyka sú však často veľmi zložité a preto by nebolo praktické vopred drasticky okresať možnosti navrhovanej analýzy. Prijmem preto dostatočne komplexný systém Transparentnej intenzionálnej logiky (TIL).[4]

Na zdôvodnenie tohto rozhodnutia možno uviesť mnoho dôvodov. TIL má totiž veľmi bohatý technický aparát, ktorý obsahuje napríklad premenné pre časové okamihy, rôzne kvantifikátory či nástroje na analýzu anafory. Ako to súvisí s deontickou logikou? Premenné pre časové okamihy umož-

[3] Za prvý pokus o analýzu viet typu (2) možno považovať klasickú stať Wright (1951), k tejto myšlienke sa však vrátila aj moderná deontická logika, predovšetkým tzv. *Deontic action logic* (DAL), pozri napríklad Kulicki – Trypuz (2012). V TIL sa venoval analýze tohto druhu Kuchyňka (2012).

[4] TIL je parciálny typovaný lambda kalkul; čitateľ sa s ním môže zoznámiť v dielach Tichý (1988), Raclavský (2009), Duží – Jespersen – Materna (2010), Duží – Materna (2012) ako aj v článkoch uvedených autorov.

ňujú, aby sme v analýze rešpektovali časovú následnosť, čo môže byť v prípade deontických viet zásadné. Rôzne gramatické časy i slová reprezentujúce časovú následnosť sú obvyklou súčasťou deontických viet. Kvantifikované výrazy či anaforické odkazy sú v deontických vetách taktiež bežné – je preto vhodné mať technický aparát, ktorý by ich dokázal analyzovať. Tieto dôvody sú, prirodzene, veľmi všeobecné. Slúžia iba na zdôvodnenie tézy, že TIL je vhodným rámcom pre budovanie deontickej logiky, pričom nepopieram, že v mnohých prípadoch nám môžu poslúžiť aj jednoduchšie rámce.

Na účely práce teraz zavediem niekoľko základných termínov – čitateľ znalý TIL môže zvyšok tejto kapitoly preskočiť. Konštrukcia, základný stavebný kameň pojmovej výbavy TIL, sa chápe ako abstraktná štruktúrovaná procedúra, pričom konštrukcie možno priradiť výrazom prirodzeného jazyka ako ich význam. Výrazy označujú denotáty (funkcie, konštrukcie), prípadne neoznačujú nič. Ak konštrukcia *C* pri ohodnotení *v* nič nekonštruuje, povieme, že *C* je *v*-nevlastná; ak pri ohodnotení *v* konštruuje objekt *X*, povieme, že *C v*-konštruuje *X*.

Ontológia TIL je usporiadaná do rozvetvenej hierarchie typov, ktorá sa buduje nad určitou bázou. Na účely analýzy prirodzeného jazyka volíme štandardne bázu (ο, ι, τ, ω): ο (súbor explikátov pravdivostných hodnôt), ι (súbor explikátov indivíduí), τ (súbor explikátov časových okamihov – reálnych čísel) a ω (súbor explikátov možných svetov). Objektmi sú v TIL extenzie, intenzie a hyperintenzie. *Extenzie* sú entity ako čísla, indivíduá či množiny. *Intenzie* sú funkcie definované na svetamihoch (svetamih je dvojica možný svet – časový okamih). Intenziami sú napríklad propozície či vlastnosti indivíduí. *Hyperintenziami* sú už spomínané konštrukcie.

V TIL rozlišujeme šesť druhov konštrukcií: trivializáciu, premenné, vykonanie, dvojité vykonanie, uzáver a kompozíciu. Na účely tejto state naznačím, čo je to trivializácia, kompozícia, premenná a uzáver. ^0X je konštrukcia, ktorá sa nazýva *trivializácia*. Konštruuje objekt X bez akejkoľvek zmeny. $[X\ Y_1 \ldots Y_n]$ je konštrukcia nazývaná *kompozícia*. Kompozícia spočíva v aplikácii funkcie na argumenty, čím sa získajú hodnoty danej funkcie pre dané argumenty. *Premenná* je konštrukcia, ktorá konštruuje objekty v závislosti od ohodnotenia. *Uzáver* je konštrukcia, ktorá konštruuje funkciu abstrakciou od hodnôt jej argumentov.

4. Prvé priblíženie k analýze deontických modalít

Teraz je už načase, aby sme sa priblížili k analýze deontických modalít. Pomôžeme si konkrétnou vetou

(4) Je prikázané, aby Pavel mlčal.

Podľa vety (4) je prikázané, aby nastal taký stav vecí, v ktorom Pavel mlčí. Je tu zároveň prikázaná^{S} určitá empirická veta. V TIL rozlišujeme medzi denotátom vety (t. j. propozíciou) a významom vety (t. j. konštrukciou propozície). Majme konštrukciu propozície *E*. Môžeme ju v ďalších konštrukciách buď použiť, alebo sa o nej zmieniť (použijeme ^{0}E), pričom konštrukcia ^{0}E konštruuje konštrukciu *E*, ktorá zase konštruuje propozíciu. To, že sa uvedené deontické slovné spojenie viaže na vetu, nám v systéme TIL teda otvára dve možnosti: (i) prikázané^{T}, zakázané^{T} či dovolené^{T} budú propozície[5] a (ii) prikázané^{T}, zakázané^{T}, či dovolené^{T} budú konštrukcie propozícií.

5. Môžeme explikovať to, čo sa prikazuje, zakazuje, či dovoľuje, ako propozície?

Hneď, ako si uvedomíme, že skúmame deontické *vetné* operátory, propozície sa stávajú intuitívne prijateľnou možnosťou, ako analyzovať to, na čo sa tieto operátory viažu. Propozície sú v TIL objektmi typu $((o\tau)\omega)$, v bežne používanej skrátenej notácii $o_{\tau\omega}$. Sú to funkcie zo svetamihov do pravdivostných hodnôt.

Deontické operátory, ktoré sa viažu na propozície, sú potom objektmi typu $(oo_{\tau\omega})_{\tau\omega}$. Takto chápané deontické operátory sú *vlastnosťami propozícií*. Vzhľadom na možné svety a časové okamihy vyčleňujú množiny propozícií, ktoré sú v daných svetoch a časoch prikázané^{T}, zakázané^{T} či dovolené^{T}. V tejto stati budeme pre jednoduchosť používať iba deontické operátory troch druhov: *O* (z anglického *obligatory*) je funkcia označená slovným spojením *je prikázané, aby*, *F* (z anglického *forbidden*) je funkcia označená výrazom *je zakázané, aby* a *P* (z anglického *permitted*) je funkcia označená výrazom *je dovolené, aby*.

[5] Návrh prijať možnosť (i) možno nájsť v Duží – Jespersen – Materna (2010, 27).

Na syntaktickej úrovni môžeme členiť vety na tie, ktoré sú prikázané^S (resp. zakázané^S, dovolené^S), a tie, ktoré nie sú prikázané^S (zakázané^S, dovolené^S). Je vôbec možné takéto členenie? Predstavme si opäť jednoduchý svet, v ktorom existuje jediný normatívny systém: kláštorný poriadok z nášho príkladu. Keby sme nazreli do kláštorného poriadku, mohli by sme vyčleniť všetky vety, ktoré obsahujú hlavný vetný operátor *je prikázané, aby*; keby sme z nich toto slovné spojenie odobrali, získame množinu viet, o ktorých možno povedať, že sú prikázané^S (samozrejme len pri zjednodušujúcom predpoklade homogenity predmetu analýzy). Propozície, ktoré sú denotátmi týchto viet, potom tvoria množinu propozícií, ktorých vlastnosťou je to, že sú prikázané^T.

Výhodou tejto koncepcie je to, že propozície umožňujú simulovať odvodzovanie, ktoré bežne robíme a mohli by sme bez obáv povedať: *de minimis non curat propositio*. Ak v určitej vete vymeníme poradie disjunktov či konjunktov, ak zmeníme spojky (zachovávajúc pravdivostné podmienky), či ak pridáme k vete nejakú tautológiu, bude stále označovať tú istú propozíciu.

Predstavme si, že by kláštorný poriadok prikazoval mníchom, aby mlčali a čítali. Keby sa Pavel spýtal správcu, či je prikázané, aby mnísi čítali a mlčali, ako by mal správca správne odpovedať? Nesporne existuje silná jazyková intuícia v prospech kladnej odpovede. Takúto intuíciu analýza pomocou propozícií plne rešpektuje a poskytuje jej sémantickú oporu.

Analogickým spôsobom by bolo možné vo svetle navrhovanej analýzy zdôvodniť mnoho úsudkov, ktoré by sme bežne považovali za platné, čo je nepochybne dobrým argumentom v prospech jej prijateľnosti. Vychádzajúc z takejto sémantickej analýzy by (s prijatím určitých neproblematických definícií, axióm a pravidiel odvodzovania) mohla vzniknúť sľubná deontická logika. Mali by sme teda túto analýzu prijať? Odložme si toto rozhodnutie na neskôr a pozrime sa najprv na druhú alternatívu.

6. Môžeme explikovať to, čo sa prikazuje, zakazuje, či dovoľuje, ako konštrukcie propozícií?

Opäť najprv predstavím návrh a potom prejdem k úvahám o jeho prijateľnosti. Konštrukcie sú objektmi typu $*_n$; nech O^* je funkcia, ktorá je denotátom vetného operátora *je prikázané, aby*, F^* funkcia, ktorá je denotátom operátora *je zakázané, aby* a P^*, funkcia, ktorá je denotátom operátora *je do-*

volené, aby. Pôjde o funkcie zo svetamihov do množín konštrukcií propozícií, teda o objekty typu $(o^*{}_n)_{\tau\omega}$. Do takýchto množín budú patriť konštrukcie propozícií, ktoré vyjadrujú vety explicitne uvedené v určitých normatívnych systémoch, nariadenia vytesané do mramorových tabúľ či (v širšom zmysle) všetky explicitne formulované nariadenia.

Je takáto analýza prijateľná? Z hľadiska jazykových intuícií nepochybne je. Umožňuje nám totiž zachovať štruktúru tvrdení o príkazoch, zákazoch a dovoleniach presne tak, ako boli formulované, takže otázka, či je takáto analýza adekvátna alebo dosť jemná, ani neprichádza do úvahy. Otázkou však zostáva, či nie je *zbytočne* reštriktívna. Ukážem, že to tak nie je. Zvážme nasledujúce vety:

(5) Mnísi mlčia.

(6) Mnísi mlčia a prší alebo neprší.

(7) Mnísi mlčia, prší alebo neprší, každý starý mládenec je neženatý muž a 2+2=4.

Problémom návrhu predstaveného v prechádzajúcej kapitole je to, že vety (5) – (7) označujú tú istú propozíciu (majú rovnaké pravdivostné podmienky), no je tu silná jazyková intuícia, že ak prikážeme (5), (6) a (7), vzniknú tri odlišné príkazy (a tvrdenia o príkazoch).

Osvetlime si túto intuíciu pomocou nášho ilustračného príkladu. Ako som už písala, keby sa Pavel opýtal správcu, či je prikázané, aby mnísi mlčali, pravdovravný správca Richard by mu musel odpovedať kladne. Čo keby sa Pavel následne opýtal, či je prikázané aj to, aby mnísi mlčali a aby pršalo alebo nepršalo? Ako by mohol správca adekvátne zareagovať na túto otázku? Zrejme by (v ľahkých pochybnostiach o Pavlovom mentálnom zdraví) pokrútil neveriacky hlavou nad čudnou otázkou a odvetil, že to veru nie, kláštorný poriadok sa predsa nijak nepokúša regulovať pravdivosť tautológií. Odmietavú odpoveď by sme mohli očakávať aj pri vete (7). Zdá sa teda, že by nešlo o rovnaké príkazy (a tvrdenia o príkazoch), pretože keby šlo, správca by mal odpovedať jednotne vo všetkých troch prípadoch.

Ak budeme predpokladať, že sú prikázané^T, zakázané^T či dovolené^T konštrukcie propozícií, vyššie naznačená intuícia ostane zachovaná. Opäť budeme môcť na syntaktickej úrovni vymedziť prikázané^S (zakázané^S, dovolené^S) vety a na sémantickej úrovni prikázané^T (zakázané^T, dovolené^T) významy týchto viet.

Táto analýza je však očividne reštriktívna, pokiaľ ide o vyplývanie. Ak totiž máme určitú prikázanú^S vetu, v množine prikázaných^T konštrukcií budeme mať jej *presný* význam. Ak by kláštorný poriadok obsahoval vetu *Je prikázané, aby mnísi mlčali a aby správcovia nepili alkohol*, analýza pomocou propozícií nám umožňuje prirodzeným spôsobom odvodiť, že je prikázané, aby mnísi mlčali; že je prikázané, aby správcovia nepili alkohol; že je prikázané, aby správcovia nepili alkohol a aby mnísi mlčali atď. Analýza pomocou konštrukcií propozícií nedovoľuje ani len výmenu konjunktov, pretože by, striktne vzaté, išlo už o inú vetu s iným významom.

Ak chceme zistiť, čo logicky vyplýva z určitých tvrdení o príkazoch, zákazoch či dovoleniach (a nejakých empirických viet), bolo by užitočné, keby boli prikázané^T, zakázané^T či dovolené^T propozície. Ak však chceme rozlišovať medzi príkazmi viet (5), (6) a (7), prišlo by nám vhod, keby boli prikázané^T, zakázané^T či dovolené^T konštrukcie propozícií. Oba návrhy sa teda zdajú byť v určitých ohľadoch prijateľné a majú nezanedbateľné výhody, pričom najvýhodnejšie by bolo, keby sa mohli navzájom doplňovať. Sme odsúdení na Sofiinu voľbu alebo môžeme pracovať s oboma v medziach jedného systému?

7. Výsledná analýza deontických modalít

TIL umožňuje zachovať oba návrhy v medziach jednej logickej analýzy. Prirodzene, nebude to zadarmo. Cenou, ktorú bude treba zaplatiť, je systematická dvojznačnosť deontických vetných operátorov.

Analogická situácia nastala v TIL v prípade postojov. *Implicitné postoje* (postoje subjektov k propozíciám) viedli k predpokladu logicky dokonalého subjektu (avšak výborne simulovali vyplývanie, ktoré s väčšou či menšou úspešnosťou bežne robíme) a *explicitné postoje* (postoje subjektov ku konštrukciám propozícií) predpokladali subjekty neschopné odvodzovať (avšak výborne zachytávali to, čo si explicitne uvedomujeme).

Situácia, v ktorej sme sa ocitli s deskriptívnymi deontickými slovnými spojeniami, je úplne analogická – preto prijmem zaužívanú terminológiu a budem hovoriť o *implicitných* a *explicitných* deontických *funkciách.*

Chcela by som podotknúť, že návrh pracovať s explicitne prikázanými^T *konštrukciami* propozícií má nezanedbateľné výhody oproti konkurenčnému sentencializmu, ktorý poznáme najmä z epistemickej logiky. Ako ukázal Levesque (1984), tento prístup je vo všeobecnosti príliš reštriktívny a navyše zahŕňa syntaktické entity do sémantických štruktúr. Ak sa vrátime k nášmu príkladu z predchádzajúcej kapitoly, sentencialisti by povedali, že explicitne prikázané^{T}, zakázané^{T} či dovolené^{T} sú vety samé a takto by triviálne zachovali intuíciu, že ak prikážeme (5), (6) a (7), vzniknú tri odlišné príkazy (veď predsa ide o tri rôzne vety). Aj evidentne rovnaké tvrdenia o príkazoch vyjadrené v rôznych jazykoch by sa však potom museli považovať za odlišné.[6] Moja analýza takýmito neduhmi netrpí.

O, P a F budú implicitné a O^{*}, P^{*} a F^{*} explicitné deontické funkcie. Najprv vymedzíme pravdivostné podmienky pre O a O^{*}. Nech ${}^{0}T$ konštruuje pravdivostnú hodnotu pravda a ${}^{0}F$ pravdivostnú hodnotu nepravda. Budeme písať $\alpha : \beta$ vtedy a len vtedy, keď (vtt) α v-konštruuje ten istý objekt ako β;[7] potom pre ľubovoľnú valuáciu platí:

$$
\begin{aligned}
&{}^{0}T : [{}^{0}O_{wt}\ [\lambda w\lambda t\ [C]]] \text{ vtt } \lambda w\lambda t\ [C] \in O_{wt}\\
&{}^{0}F : [{}^{0}O_{wt}\ [\lambda w\lambda t\ [C]]] \text{ vtt } \lambda w\lambda t\ [C] \notin O_{wt}\\
&{}^{0}T : [{}^{0}O^{*}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]] \text{ vtt } {}^{0}[\lambda w\lambda t\ [C]] \in O^{*}_{wt}\\
&{}^{0}F : [{}^{0}O^{*}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]] \text{ vtt } {}^{0}[\lambda w\lambda t\ [C]] \notin O^{*}_{wt}
\end{aligned}
$$

Ako vidíme, pravdivostné podmienky O a O^{*} sú vymedzené tak, že konštrukcia $[{}^{0}O_{wt}\ [\lambda w\lambda t\ [C]]]$ v-konštruuje pravdivostnú hodnotu pravda vtedy a len vtedy, keď propozícia $\lambda w\lambda t\ [C]$ patrí do množiny prikázaných^{T} propozícií vo svete w a čase t, a pravdivostnú hodnotu nepravda, ak do tejto množiny vo w a t nepatrí. Konštrukcia $[{}^{0}O^{*}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]]$ v-konštruuje pravdivostnú hodnotu pravda vtedy a len vtedy, keď konštrukcia ${}^{0}[\lambda w\lambda t\ [C]]$ patrí do množiny prikázaných^{T} konštrukcií vo svete w a čase t, a hodnotu ne-

[6] Predstavme si, že by sa Pavel pýtal správcu na platnosť určitých príkazov v inom jazyku, ako je jazyk kláštorného poriadku. Keby bol správca sentencialista, musel by Pavlovi na každú otázku, či je niečo prikázané, zakázané, či dovolené, odpovedať negatívne. Po chvíľke pýtania by si Pavel musel myslieť, že v kláštore vládne číra anarchia. Obhajca sentencializmu by sa mohol vyrovnať s námietkou pomocou vzájomnej preložiteľnosti viet rôznych jazykov, čo je však, ako vieme, problém zdôvodniť bez predpokladu niečoho, čo majú tieto vety spoločné, t. j. významu, pozri Tichý (1988, 5-9).

[7] Tu využívam Tichého pojem zhody, pozri napríklad Tichý (1982, 64-65). Naznačeným spôsobom by sme mohli uviesť pravdivostné podmienky aj pre zvyšné formuly navrhovaného systému, na to však nemám dosť priestoru.

pravda, ak do tejto množiny vo w a t nepatrí. Analýza potom bude vyzerať nasledovne:

$$\begin{tikzcd}[column sep=large]
\begin{array}{c}\text{vetný operátor}\\ \textit{je prikázané, aby}\end{array} \arrow[r, "\textit{vyjadruje}"'] \arrow[rr, dashed, bend right, "\textit{označuje}"'] & \begin{array}{c}\text{význam}\\ \text{(i) } {}^{0}O\\ \text{(ii) } {}^{0}O^{*}\end{array} \arrow[r, "\textit{konštruuje}"'] & \begin{array}{c}\text{denotát}\\ \text{(i) } (oo_{\tau\omega})_{\tau\omega}\\ \text{(ii) } (o^{*}{}_{n})_{\tau\omega}\end{array}
\end{tikzcd}$$

F a F^* možno neproblematicky definovať pomocou operátorov O a O^*. Ak je totiž zakázané C, možno ekvivalentne tvrdiť, že je prikázaná negácia C. Nech $\lambda w\lambda t[C]$ je konštrukcia propozície. Potom bude pre ľubovoľnú valuáciu platiť:

$$\lambda w\lambda t\ [{}^{0}F_{wt}\ [\lambda w\lambda t\ [C]]] \overset{\text{def}}{=} \lambda w\lambda t\ [{}^{0}O_{wt}\ [\lambda w\lambda t\ [\neg C]]]$$

$$\lambda w\lambda t\ [{}^{0}F^{*}{}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]] \overset{\text{def}}{=} \lambda w\lambda t\ [{}^{0}O^{*}{}_{wt}\ {}^{0}[\lambda w\lambda t\ [\neg C]]]$$

Situácia sa trochu komplikuje v prípade operátorov P a P^*. Aj P a P^* by sme mohli elegantne definovať pomocou operátorov O a O^*.[8] Formula s P^* by potom však hovorila iba toľko, že nie je explicitne zakázané C. Takýto slabý pojem dovolenia by bol „explicitný“ iba v tom zmysle, že by bol definovaný pomocou explicitnej funkcie O^*. Bohužiaľ, tento pojem by v žiadnom prípade nemohol slúžiť ako adekvátna explikácia *explicitne dovoleného* a je otázne, či by bol vôbec na niečo užitočný. Operátor P^* preto nedefinujeme pomocou O^*, ale budeme postupovať analogicky ako pri O^*:

$${}^{0}T: [{}^{0}P^{*}{}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]] \text{ vtt } {}^{0}[\lambda w\lambda t\ [C]] \in P^{*}{}_{wt}$$

$${}^{0}F: [{}^{0}P^{*}{}_{wt}\ {}^{0}[\lambda w\lambda t\ [C]]] \text{ vtt } {}^{0}[\lambda w\lambda t\ [C]] \notin P^{*}{}_{wt}$$

Posledným problémom ostáva operátor P. Kedy je niečo implicitne dovolené? V duchu predošlej analýzy by sme mohli povedať, že propozícia konštruovaná konštrukciou $\lambda w\lambda t\ [C]$ patrí do množiny implicitne dovolených propozícií vtedy a len vtedy, keď táto konštrukcia patrí do množiny explicitne dovolených konštrukcií, no takýto pojem implicitného dovolenia

[8] Proti definovaniu dovolení pomocou príkazov argumentuje aj Svoboda (2013) v kapitole venovanej pojmu dovolenia.

by bol príliš reštriktívny a nezodpovedal by tomu, ako bežne rozumieme termínu *implicitný*. Je totiž zrejmé, že niečo je implicitne dovolené aj vtedy, ak to nie je implicitne zakázané. Pravdivostné podmienky pre operátor P teda budú nasledovné:

${}^0T$: $[{}^0P_{wt}\ [\lambda w\lambda t\ [C]]]$ vtt ${}^0[\lambda w\lambda t\ [C]] \in P^{*}_{wt}$ alebo $\lambda w\lambda t\ [\neg C] \notin O_{wt}$
${}^0F$: $[{}^0P_{wt}\ [\lambda w\lambda t\ [C]]]$ vtt ${}^0[\lambda w\lambda t\ [C]] \notin P^{*}_{wt}$ a $\lambda w\lambda t\ [\neg C] \in O_{wt}$

8. Axiómy a pravidlá odvodzovania

Nech $=_i$ označuje reláciu procedurálneho izomorfizmu,[9] nech $\vDash$ označuje vzťah vyplývania medzi konštrukciami,[10] nech konštrukcie c, c' a d konštruujú propozície a nech sú konštrukcie c_{wt} a d_{wt} konštrukciami c a d po vykonaní intenzionálneho zostupu, t. j. po aplikovaní na svet w a čas t. Potom bude pre ľubovoľnú valuáciu platiť:

(PL1) Všetky axiómy a pravidlá odvodzovania predikátovej logiky prvého rádu.
(R1) $[{}^0O^{*}_{wt}\ {}^0c] \vDash [{}^0O_{wt}\ c]$
(R2) (i) $\lambda w\lambda t\ [{}^0O^{*}_{wt}\ {}^0\mathrm{c}]$, (ii) $[{}^0{=_i}\ {}^0c\ {}^0c'] \vDash \lambda w\lambda t\ [{}^0O^{*}_{wt}\ {}^0c']$
(R3) (i) $[{}^0O_{wt}\ [\lambda w\lambda t\ [c_{wt} \rightarrow d_{wt}]]]$, (ii) $[{}^0O_{wt}\ c] \vDash [{}^0O_{wt}\ d]$
(R4) $[\forall^{\omega}w\forall^{\tau}t\ c_{wt}] \vDash [{}^0O_{wt}\ c]$

Pravidlá (R1) a (R2) sú neproblematické. (R1) hovorí, že ak je explicitne prikázaná$^{\mathrm{T}}$ určitá konštrukcia propozície, je implicitne prikázaná$^{\mathrm{T}}$ propozícia, ktorú konštruuje, a (R2) hovorí, že ak je explicitne prikázaná$^{\mathrm{T}}$ konštrukcia c a konštrukcia c' je s ňou procedurálne izomorfná, je explicitne prikázaná$^{\mathrm{T}}$ aj konštrukcia c'. Hranicu obmedzenosti vyplývania explicitného z explicitného by bolo možné posunúť prijatím viac či menej prísnej definície procedurálneho izomorfizmu. Pravidlá (R3) a (R4) sú analógiami pravi-

[9] Na účely práce stačí povedať, že ide o reláciu medzi konštrukciami, ktorá je „liberálnejšia“ ako identita a reštriktívnejšia ako ekvivalencia. Presné vymedzenie procedurálneho izomorfizmu je technickým problémom TIL, ktorým sa na tomto mieste nemusíme zaoberať.

[10] Viac k tomu pozri v Raclavský (2009, 160), resp. Raclavský (2012, 248).

diel (a axióm) Štandardnej deontickej logiky (SDL) – pozri McNamara (2006, 207-208). (R3) hovorí, že ak je implicitne prikázaná$^{\mathrm{T}}$ implikácia aj jej antecedent, je implicitne prikázaný$^{\mathrm{T}}$ aj jej konzekvent. (R4) hovorí, že ak propozícia c nadobúda pravdivostnú hodnotu pravda vo všetkých možných svetoch a časoch, tak je c implicitne prikázaná$^{\mathrm{T}}$.

V SDL sa objavuje aj pravidlo, podľa ktorého z toho, že je niečo prikázané, vyplýva, že nie je prikázaný opak. Toto pravidlo však považujem za problematické, a to predovšetkým vtedy, ak nepracujeme s deontickými operátormi obmedzenými na konkrétne normatívne systémy. Rôzne normatívne systémy (či autority) si môžu protirečiť – a dokonca si fakticky často protirečia – bolo by preto absurdné, aby logika požadovala opak. Tento problém by bolo možné prirodzene odstrániť tým, že by sme pracovali s deontickými operátormi relativizovanými na normatívne systémy. Predpokladať konzistentnosť jednotlivých normatívnych systémov by bol síce idealizujúci, no v žiadnom prípade nie absurdný predpoklad.

Uvedené axiómy a pravidlá odvodzovania považujem za vhodný a neproblematický *základ* pre deontickú logiku budovanú v medziach TIL.[11]

9. Russellov test

> Logickú teóriu možno testovať skúšaním jej schopnosti vyrovnať sa s ťažkosťami. Keď uvažujeme o logike, je užitočné zaťažiť myseľ toľkými ťažkosťami, koľkými sa len dá, lebo slúžia tomu istému účelu ako experimenty vo fyzikálnej vede. (Russell 2005, 68)

Nazvime tento Russellom naznačený spôsob testovania logických teórií *Russellov test* a skúsme mu podrobiť navrhovanú teóriu na príklade *Rossovho*

[11] Vytvorenie ucelenej deontickej TIL je náročný projekt, ktorý ďaleko presahuje rozsah tento state. Takáto teória by sa mala nejakým spôsobom vyrovnať s analýzou deontických viet zvyšných dvoch typov (spomeňme si na druhú kapitolu tejto state), mala by byť schopná skúmať logické vzťahy medzi nimi, mala by mať dobre definovanú syntax a sémantiku a mala by obsahovať určitý zoznam axióm a pravidiel odvodzovania. Pozorný čitateľ si navyše iste všimol, že situáciu tu oproti SDL komplikujú operátory P a P^{*} – neboli definované pomocou O a O^{*}, preto je potrebné pre ne zaviesť špecifické pravidlá. Naznačím aspoň tri najzákladnejšie. Čo sa týka P^{*}, zavedenie analógií k (R1) a (R2) pre P^{*} by bolo úplne neproblematické. Čo sa týka P, základné pravidlo by bolo nasledovné: $\neg[{}^{0}O_{wt}\ [\lambda w \lambda t\ [\neg C]]] \vDash [{}^{0}P_{wt}\ [\lambda w \lambda t\ [C]]]$.

paradoxu, ktorý je jedným z dobre známych paradoxov deontickej logiky. Riešenie, ktoré prestavím, však možno aplikovať na viaceré z paradoxov deontickej logiky.[12] Rossov paradox možno uviesť pomocou nasledujúceho intuitívne neplatného úsudku:

(8) Je prikázané, aby Pavel mlčal.

(9) Ak je prikázané, aby Pavel mlčal, je prikázané, aby Pavel mlčal alebo zabil Richarda.

(10) Teda: Je prikázané, aby Pavel mlčal alebo zabil Richarda.

V samom odvodení záveru z premís nie je problém, ide o obyčajnú aplikáciu pravidla modus ponens. Pochybná je premisa (9), ktorá je v niektorých deontických logikách dokázateľná. Vzhľadom na navrhnutú analýzu sa však musíme spýtať, či máme analyzovať deontické vetné operátory v premisách ako explicitné, alebo ako implicitné deontické funkcie. Ak ako explicitné funkcie, zachovala by sa paradoxnosť premisy (9) – no nevedeli by sme ju dokázať, a teda by nám nič nebránilo zbaviť sa celého argumentu odmietnutím tejto premisy. Ak však premisa (8) obsahuje explicitnú deontickú funkciu, vieme odvodiť premisu (9) s implicitnou deontickou funkciou a následne záver – tiež s implicitnou deontickou funkciou. V tomto prípade však záhada neexistuje. Disjunkcia v závere nám nedáva možnosť zvoliť si, ktorý disjunkt splníme – záväzné je pre nás najmä to, čo bolo prikázané^{T} explicitne. Môžeme sa, samozrejme, riadiť aj implicitným nariadením, ktoré z nejakého explicitného nariadenia vyplýva, no nesmieme tým zároveň porušiť dané explicitné nariadenie. Dokážme, že z „explicitnej“ premisy (8) naozaj vyplýva „implicitný“ záver. Doteraz neuvedené typy: $\rightarrow,\vee,\wedge/(ooo)$; $\forall^{\omega}/(o(o\omega))$; $\forall^{\tau}/(o(o\tau))$; Mlčať$/(o\iota)_{\tau\omega}$; *Zabiť*$/(o\iota\iota)_{\tau\omega}$; *Pavel*, *Richard*$/\iota$; $w \rightarrow_{v} \omega$; $t \rightarrow_{v} \tau$.

[12] Napríklad *paradox A. Priora*, či *paradox milosrdného samaritána*; pozri Åqvist (2002, 179-186, 197-205). Pod paradoxmi A. Priora sa obvykle rozumejú deontické analógie k paradoxom striktnej i materiálnej implikácie, hoci Prior (1954, 64-65) hovoril iba o analógiách k paradoxom striktnej implikácie. Týmto paradoxom sa pre nedostatok priestoru venovať nebudem, no v závere práce naznačím možné zovšeobecnenie predstaveného riešenia. Treba však upozorniť, že nejde o „univerzálny liek“ na všetky paradoxy deontickej logiky. Na riešenie paradoxov Chisholmovho typu by bolo treba rozlíšiť primárny a sekundárny výskyt O a O^{*} či zaviesť dyadický jazyk, na čo však v tejto stati nemám dostatok priestoru.

Lemma: $\forall^{\omega}w\forall^{\tau}t[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \rightarrow$
$[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee [{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]$

Pre ľubovoľnú valuáciu zachovávajú nasledujúce kroky pravdivosť:

1. $\exists^{\omega}w\exists^{\tau}t[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \wedge [\neg[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee$
   $[{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]]$ NP
2. $[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \wedge [\neg[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee$
   $[{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]$ O∃, 1
3. $[{}^{0}Mlčať_{wt}\,{}^{0}Pavel]$ O∧, 2
4. $\neg[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee [{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]$ O∧, 2
5. $[\neg[{}^{0}Mlčať_{wt}\,{}^{0}Pavel]] \wedge [\neg[{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]$ Neg∨, 4
6. $\neg[{}^{0}Mlčať_{wt}\,{}^{0}Pavel]$ O∧, 5

Spor v riadkoch 3 a 6. Lemma teda platí.

Dôkaz. Pre ľubovoľnú valuáciu zachovávajú nasledujúce kroky pravdivosť:

1. $[{}^{0}O^{*}_{wt}\,{}^{0}[\lambda w\lambda t\,[{}^{0}Mlčať_{wt}\,{}^{0}Pavel]]]$ PP
2. $[{}^{0}O_{wt}\,[\lambda w\lambda t\,[{}^{0}Mlčať_{wt}\,{}^{0}Pavel]]]$ R1,1
3. $\forall^{\omega}w\forall^{\tau}t[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \rightarrow$
   $[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee [{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]$ Lemma
4. $[{}^{0}O_{wt}\,[\lambda w\lambda t\,[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \rightarrow$
   $[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee [{}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]]]$ R4,3
5. $[{}^{0}O_{wt}\,[\lambda w\lambda t\,[[{}^{0}Mlčať_{wt}\,{}^{0}Pavel] \vee$
   $[\ {}^{0}Zabiť_{wt}\,{}^{0}Pavel\,{}^{0}Richard]]]]$ R3,2,4

Ako vidíme, z bodu 1, ktorý je „explicitnou“ analýzou (8), odvodili sme pomocou už dokázanej lemmy bod 5, ktorý je „implicitnou“ analýzou záveru. V tejto podobe je úsudok platný, no nezostalo na ňom nič paradoxné či neintuitívne. Pavel síce môže splniť implicitný príkaz opísaný v závere zabitím Richarda, no nesplní tým príkaz, o ktorom v prvom rade šlo – ten, ktorý je opísaný v premise. Ako som už naznačila, toto riešenie možno aplikovať na viacero paradoxov deontickej logiky. Deontické vety, ktoré opisujú explicitne zadané príkazy či zákazy, analyzujeme pomocou explicitných deontických funkcií. Následne sa ukáže, že nevieme odvodiť explicitnú analýzu záveru, ale iba implicitnú analýzu. Taký výsledok však nie je v ničom paradoxný.

10. Semi-explicitné a semi-implicitné deontické modality

Na záver zavediem drobné, no zaujímavé rozšírenie navrhovanej analýzy a rozlíšim navyše semi-implicitné a semi-explicitné deontické modality. Semi-implicitné deontické modality (možno ich značiť implicitné^{½}, resp. O^{½}, P^{½} a F^{½}) získame z implicitných deontických modalít veľmi jednoducho – bude pre ne platiť všetko to, čo pre implicitné modality, okrem zavedených pravidiel. Čitateľ sa iste pýta, na čo je to dobré. Odpoveď je jednoduchá: Implicitné^{½} deontické modality neumožňujú odvodzovanie slabších dôsledkov, no umožňujú v analýze viet zanedbať drobné formulačné rozdiely, ktoré nemajú vplyv na pravdivostné podmienky daných viet i pridávanie logických a matematických právd. Semi-explicitné deontické modality vzniknú tak, že nebudeme pracovať s veľmi prísnou reláciou procedurálneho izomorfizmu, ale zavedieme nejakú liberálnejšiu reláciu medzi konštrukciami, pričom ponechávam otvorené, ktorú konkrétnu reláciu si zvoliť.

11. Záver

Iste by nebolo prehnané povedať, že deontické modálne logiky zvrchovane vládnu v oblasti výskumu deontických modalít. Na tejto vláde sa podieľa aj počítačová veda, ktorá umožňuje jednoduché spracovanie dát, odvodzovanie či zisťovanie konzistentnosti. Výsledkov je veľa, no často chýba ich filozofické zdôvodnenie a vyriešenie problémov, ktoré spočívajú v samých základoch rôznych systémov deontických logík.

V tejto stati som navrhla základy analýzy deontických modalít v TIL ako alternatívu k hlavnému prúdu deontickej logiky. Analýza je naznačená iba v hrubých rysoch a treba upozorniť, že nejde o ucelený axiomatický systém, čo je daň za bohatosť systému TIL. Tento nedostatok však nepovažujem za principiálny.

V stati som zaviedla rozlíšenie medzi implicitnými a explicitnými deontickými funkciami, ktoré je analogické rozlíšeniu medzi implicitnými a explicitnými postojmi. Implicitné deontické funkcie sú funkciami zo svetamihov do množín propozícií a sú otvorené logickému vyplývaniu. Ak sme zvedaví, čo všetko vyplýva z určitých deskriptívnych deontických viet, mali by sme pracovať s implicitnými deontickými funkciami. Musíme sa však mať na pozore, pretože ak chceme plniť určité nariadenia, je pre nás záväzná v prvom rade ich explicitná formulácia: Ak dostaneme určitý príkaz a chceme ho splniť, nemôžeme ľubovoľne odvodzovať a nakoniec sa rozhodnúť namiesto pôvodného príkazu splniť jeho slabší dôsledok. Explicitné deontické funkcie sú funkciami zo svetamihov do množín konštrukcií propozícií, a keďže konštrukcie propozícií sú oveľa jemnejšie objekty ako propozície samé, možnosti odvodzovania sú tu značne obmedzené.

Navrhla som tiež niekoľko pravidiel odvodzovania, ktoré umožňujú overovať platnosť deontických argumentov. Navrhnutú analýzu som podrobila Russellovmu testu na príklade Rossovho paradoxu. Rozsah tejto state mi nedovoľuje venovať sa viacerým paradoxom, no podobným spôsobom možno rozlúsknuť viacero paradoxov deontickej logiky. Rozlišovanie implicitných a explicitných deontických funkcií totiž poskytuje účinnú zbraň proti paradoxom: Ak by určitý paradoxný argument obsahoval iba explicitné deontické modality, nebol by platný – ak by obsahoval explicitné deontické modality v premisách a implicitnú modalitu v závere, bol by síce platný, no vytratila by sa jeho paradoxnosť. Takéto riešenie nemožno aplikovať na deontické modálne logiky bez problémov – keďže nepracujú s konštrukciami, je problematické odlíšiť implicitné a explicitné deontické modality spôsobom, ktorý by nebol *ad hoc*.

Literatúra

ÅQVIST, L. (2002): Deontic Logic. In: Gabbay, D. M. – Guenthner, F. (eds.): *Handbook of Philosophical Logic.* Vol. 8. Dordrecht: Kluwer Academic Publisher.

BIELIK, L. – KOSTEREC, M. – ZOUHAR, M. (2014): Model metódy (1): Metóda a problém. *Filozofia* 69, č. 2, 105-118.

DUŽÍ, M. – MATERNA, P. (2001): Propositional Attitudes Revised. In: Childers, T. (ed.): *The LOGICA Yearbook* 2000. Praha: Filosofia, 163-173.

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

DUŽÍ, M. – MATERNA, P. (2012): *TIL jako procedurální logika (Průvodce zvídavého čtenáře Transparentní intensionální logikou).* Bratislava: aleph.

FORRESTER, J. W. (1984): Gentle Murder, or the Adverbial Samaritan. *The Journal of Philosophy* 81, No. 4, 193-197.

KUCHYŇKA, P. (2012): *Pravidla, jazyk a logika* (dizertačná práca). Brno: Masarykova univerzita, Filozofická fakulta.

KULICKI, P. – TRYPUZ, R. (2012): How to Build a Deontic Action Logic. In: Peliš, M. – Punčochář, V. (eds.): *The Logica Yearbook* 2011, 107-120.

LEVESQUE, H. J. (1984): A Logic of Implicit and Explicit Belief. In: *Proceedings of the National Conference on Artificial Intelligence*. Cambridge: AAAI Press/MIT Press, 198-202.

MCNAMARA, P. (2006): Deontic Logic. In: Gabbay, D. M. – Woods, J. (eds.): *Handbook of the History of Logic. Vol. 7: Logic and the Modalities in the Twentieth Century*. Amsterdam: Elsevier, 197-288.

PRIOR, A. N. (1954): The Paradoxes of Derived Obligation. *Mind* 63, 64-65.

RACLAVSKÝ, J. (2009): *Jména a deskripce: logicko-sémantická zkoumání*. Olomouc: Nakladatelství Olomouc.

RACLAVSKÝ, J. (2012): Je Tichého logika logikou? (O vztahu logické analýzy a dedukce). *Filosofický časopis* 60, č. 2, 245-254.

RUSSELL, B. (2005): *Jazyk a poznanie*. Bratislava: Kalligram.

SVOBODA, V. (2013): *Logika pro Pány, Otroky a Kibice. Filosofický průvodce světem deontické logiky*. Praha: Filosofia.

TICHÝ, P. (1978): Questions, Answers, and Logic. *American Philosophical Quarterly* 15, 275-284.

TICHÝ, P. (1982): Foundations of Partial Type Theory. *Reports on Mathematical Logic* 14, 59-72.

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

VAN ECK, J. A. (1982): A System of Temporally Relative Modal and Deontic Predicate Logic and Its Philosophical Applications. *Logique et Analyse* 25, No. 100, 339-381.

VON WRIGHT, G. H. (1951): Deontic Logic. *Mind* 60, No. 237, 1-15.

ZOUHAR, M. (2009): *Teória kvantifikácie a extenzionálna sémantika prirodzeného jazyka*. Bratislava: Filozofický ústav SAV.

Organon F 22 (2) 2015: 229-249

Suárezova teorie vzniku *species sensibilis* a kognitivního aktu vnějších smyslů v kontextu středověké a renesanční filosofie[1]

DANIEL HEIDER

Katedra filosofie a religionistiky. Teologická fakulta
Jihočeská univerzita v Českých Budějovicích. Kněžská 8. 37001 České Budějovice
Filosofický ústav AV ČR. Jilská 1
110 00 Praha 1. Česká republika
Daniel.Heider@seznam.cz

ZASLÁN: 22-09-2014 ▪ AKCEPTOVÁN: 08-12-2014

ABSTRACT: The study presents F. Suárez's theory of the principles of sensation in the context of medieval (Averroes, John of Jandun) and renaissance philosophy (Nifo, Cajetan). It proceeds in five steps. First, it considers Suárez's ontology of sensory cognitive act. Second, it treats Suárez's theory of the formation of sensible species. Third, it presents Suárez's ontology of sensible species. Fourth, it exposes Suárez's theory of the efficient causes of the sensory cognitive act. In conclusion, the author states that Suárez's theory, compared to the doctrines of Aquinas and Thomists, constitutes the significant historical shift from cognitive passivism to cognitive activism mirroring the *Zeitgeist* of the Renaissance philosophy without abandoning the basic tenets of the traditional Aristotelian – scholastic philosophy.

KEYWORDS: Activity – agent sense – efficient cause – passivity – sensible species – sensory cognitive act – Suárez.

[1] Tato studie je výsledkem badatelské činnosti podporované Grantovou agenturou České republiky v rámci grantu GA ČR 14-37038G „Mezi renesancí a barokem: Filosofie a vědění v českých zemích a jejich širší evropský kontext".

1. Úvod

Jednu z nejtypičtějších otázek kognitivní psychologie v renesančním peripatetismu představuje téma účinné příčiny kognitivního aktu majícího intencionální obsah. Vyjdeme-li z tradičního předpokladu, že při produkci aktu smyslového poznání je zapotřebí existence dvou hlavních faktorů, totiž objektu, resp. objektem emitované formy, a kognitivní mohutnosti, pak lze základní odpovědi na tuto otázku charakterizovat schématy „kognitivního pasivismu“ a „kognitivního aktivismu“. Kognitivní akt je vytvářen buď pouze tzv. intencionálním vtištěným obrazem[2] (*species impressa intentionalis*), přičemž kognitivní potence je čistě receptivní, anebo je k jeho produkci podstatným způsobem zapotřebí také aktivně působící smyslové mohutnosti, resp. duše, z níž tato mohutnost „vyrůstá“. Každému z těchto dvou přístupů k aktu smyslového poznání (*sensatio*) v tradici středověkého a renesančního aristotelismu odpovídá specifické pojetí vtištěné *species sensibilis*. Zatímco receptivistický postoj inklinuje k imateriálnímu pojetí této *species*, kognitivní dynamismus akcentuje její nehotovou, tj. materiální povahu.

Existence obou těchto přístupů, jak ukazuje diskuse v současné analytické filosofii (srov. Fish 2010, 1-3), odpovídá dvěma základním požadavkům vznášeným na každou (úspěšnou) filosofickou teorii percepce. Prvním kritériem „úspěšnosti“ filosofické teorie percepce je „test“ epistemologický. Každá pravdivá (*veridical*) vizuální zkušenost – odlišná od iluzí a halucinací – nám přináší pravdivou empirickou informaci o extramentálním světě. Tato informace je nezbytná pro naši životní orientaci. Úkolem filosofické teorie smyslového vnímání je tuto skutečnost vysvětlit. Druhá „zkouška“ filosofické teorie aspirující podat plausibilní výklad smyslové percepce, který přesahuje kompetenci čistě vědeckého přístupu, se týká fenomenologického rozměru naší smyslové zkušenosti. O smyslové zkušenosti jako takové platí, že je paradigmatickým příkladem zkušenosti *vědomé*. Každá filosofická teorie proto musí vysvětlit také tento rys naší smyslové percepce.

Není obtížné nahlédnout, že tyto dva požadavky se často nacházejí ve vzájemném napětí. Obecně platí, že čím více filosofická teorie percepce zdůrazňuje fenomenologický rozměr smyslové percepce, o to větší má problém se splněním kritéria epistemologického. Naopak, nakolik vyhovuje „testu“

[2] Vzhledem k neobrázkovému charakteru *species intentionalis* v Suárezově teorii ponecháme v textu výraz *species*.

epistemologickému, o to větší má problém s požadavkem fenomenologickým. Ve vztahu k uvedeným přístupům peripatetické tradice k námětu příčin perceptivního aktu tato tenze znamená následující: Zatímco explanační deficit ve vztahu k epistemologickému požadavku je potenciálním nebezpečím pro teorie „kognitivního aktivismu", hrozí nesplnění kritéria fenomenologického spíše pozicím „kognitivního receptivismu".

Přestože značná část témat a principů debat v současné analytické filosofii vychází z „Fragestellung" souvisejícím s karteziánským obratem ve filosofii mysli – máme na mysli především předpoklad existence mentálního stavu, který je společný fenomenologicky nerozlišitelným (pravdivým) percepcím, iluzím a halucinacím[3] –, existuje mezi pozdněscholastickými a současnými diskusemi analytických filosofů jistá paralela ohledně společného kladení problému. Omezíme-li se na téma příčin formace perceptivního aktu, které souvisí především s výše zmíněnou *vtištěnou* intencionální *species*, a necháme-li tak stranou dílčí otázky související spíše s termínem (zakončením) tohoto aktu – jde o otázky týkající se především tzv. *vyjádřené* intencionální *species* (*species expressa*), které zůstávají stranou naší pozornosti –, pak tematický „protějšek" scholastických debat lze spatřovat v kontextu úvah analytických filosofů o tzv. kauzální teorii percepce (*causal theory of perception*). Důležitý námět pak představuje v kontextu této teorie otázka vztahu kauzálního působení předmětu percepce a produkce účinku, tj. vizuální zkušenosti. Tento problém, jak ukazuje Fish (2010, 118-119), je analytickými filosofy formulován jako problém vztahu dvou tezí kauzální teorie percepce. Zatímco první je tezí o příčině (*the causal thesis*), druhá je tvrzením o účinku (*the effect thesis*). Podle první teze platí, že má-li vnímající subjekt *S* vidět veřejně dostupný objekt *O* (případ *pravdivé* percepce), musí *O* kauzálně působit na *S*. Podle druhé platí, že *O* musí v *S* vytvořit takový (mentální) stav, který je zaznamenatelný větou začínající formulací „*O* se ukazuje *S* jako ...". Necháme-li stranou (kartezianismem založenou) skutečnost, že tento účinek neboli mentální stav vyvolaný v *S* objektem *O* vychází ze zmíněného principu o mentálním stavu, který je společný percepci, iluzi a halucinaci,[4]

[3] Fish tento princip nazývá „The Common Factor Principle" (srov. Fish 2010, 3-5).

[4] Pro náš účel stačí poznamenat, že v rámci peripatetické filosofie platí, že vnější smysly jsou z hlediska své přirozenosti nasměrovány ke svým vlastním (formálním) předmětům, v případě vidění k barvě. Je-li poznávající subjekt a jeho smyslový orgán dobře disponován, a neexistuje-li v prostředí žádná překážka, nemůže dojít, co se reprezentace zá-

liší se různí stoupenci kauzální teorie právě v analýze povahy vztahu první a druhé teze. Základní nástroj aplikovaný v této diskusi pak je rozlišení dvou úrovní výkladu (srov. Dennett 1969, 93-94). Jde o rozlišení sub-personální a personální roviny výkladu. Zatímco podle prvního výkladu se o člověku hovoří jako o mentálním (vědomém) činiteli, podle druhého se člověk pojímá jako fyzikální systém, což v případě vizuální percepce neznamená nic jiného, než fyzikální (biologický) popis percepce vycházející z faktu odrazu světla vnímanými předměty, přes „putování" světla k sítnici, na níž vytváří patřičné obrazy, dále přes přenos stimulů obsažených v sítnicových buňkách optickými nervy do mozku, až k (bezproblémovému) vyvolání vizuální zkušenosti v mozku. Necháme-li stranou odlišnosti v traktování problematiky od pojetí peripateticko-scholastického, pak jedním ze společných námětů obou tradic je otázka, nakolik stav věcí obsažený v první tezi (sub-personální výklad percepce) má (nutně) za důsledek mentální událost obsaženou ve formulaci druhého pilíře kauzální teorie percepce (viz Fish 2010, 120-121).

V kontextu peripatetické filosofie nacházíme analogickou diskusi. Tuto diskusi lze formulovat v termínech (kognitivního) „přechodu" z nevědomé části světa do části vědomé (duše). Podle jednoho možného přístupu je třeba intencionální *species* natolik „spiritualizovat", a tak ji *de facto* popisovat pomocí „slovníku" personální roviny, že v případě produkce kognitivního aktu bude stačit pouhá (paradoxně: „sub-personální") recepce této *species*. Podle druhého modelu se naopak bude *species sensibilis* chápat spíše materiálně – na způsob sub-personálního výkladu –, což bude mít za důsledek požadavek o nutné doplnění výkladem personálním, který je představován faktorem aktivní duše (mysli).

Obrátíme-li pozornost ke středověké a renesanční filosofii, pak mezi představitele kognitivního aktivismu byli peripatetiky 16. století povětšinou řazeni Aristotelés, Tomáš Akvinský a někteří z jeho žáků.[5] Kognitivní aktivismus, naopak, byl ve středověké filosofii tradičně spojován se jménem sv.

kladních rysů smyslově vnímatelného předmětu týče, k nesprávné reprezentaci. Kromě toho, i když analytičtí filosofové v diskusích o smyslové percepci uvažují také komplementární případ pravé percepce, totiž případ halucinace, nejedná se z úhlu pohledu scholastické a peripatetické filosofie o instanci poznání vnějších, nýbrž pouze vnitřních smyslů, jakými je např. fantazie. Téma vnitřních smyslů však v této studii zůstává stranou naší pozornosti.

[5] K některým textům opravňujícím k takovému čtení těchto autorů, viz druhá část tohoto příspěvku.

Augustina,[6] v renesanční filosofii pak převážně s postavou latinského averroisty Jana z Jandunu (ca. 1285-1328). Ten v řadě svých textů[7] zformuloval teorii tzv. činného smyslu (*sensus agens*).[8] Přestože Jandunova teorie nenašla v renesanční filosofii mnoho stoupenců, představovala oficiální averroistickou nauku, čímž se *eo ipso* stala teorií, s níž je třeba se vyrovnat (srov. Mahoney 1971, 120). Tato teorie je zmiňována nejen v textech některých italských renesančních aristoteliků (Agostino Nifo[9]), ale také v *Commentaria una cum quaestionibus in libros Aristotelis De anima* španělského jezuity Františka Suáreze (1548–1617), u kterého představuje významný referenční bod v jeho diskusi o principech smyslové percepce.

V naší studii budeme postupovat v pěti krocích. 1) Předložíme Suárezovu argumentaci ve prospěch teze, podle níž jakýkoli (i rozumový) kognitivní akt, který jezuita identifikuje s tzv. vyjádřenou smyslovou species (*species sensibilis expressa*), vytváří ontologickou samostatnou „jednotku“, jmenovitě akcident kvality.[10] 2) Vyložíme Suárezovu teorii vzniku vtištěné *species sensibilis* vnějších smyslů. 3) Zaměříme se na téma Suárezovy ontologie této vtištěné *species sensibilis.* 4) Předložíme Suárezovu teorii formace aktu smyslové

[6] V této souvislosti často citovaná pasáž se nachází ve spise *De musica*, liber 6, caput 5, n. 10: „... *videtur mihi anima cum sentit in corpore, non ab illo aliquid pati, sed in ejus passionibus attentius agere, et has actiones sive faciles propter convenientiam, sive difficiles propter inconvenientiam, non eam latere: et hoc totum est quod sentire dicitur*“; viz Augustinus (2006a, 1169).

[7] K těmto textům (jejich transkripci) a dalším relevantním textům ze středověké filosofie k problematice činného rozumu viz Pattin (1988).

[8] K vynikajícímu úvodu do teorie činného smyslu Jana z Jandunu viz MacClintock (1956, 10-50).

[9] K Nifově koncepci viz níže.

[10] Zatímco vtištěná smyslová *species* představuje princip, jímž je poznávací mohutnost uváděna ze své počáteční indiference vůči poznanému předmětu do svého prvního aktu, tak vyjádřená smyslová *species*, kterou Suárez uvažuje, na rozdíl od Tomáše Akvinského, nejen v případě rozumového, ale i smyslového poznání, představuje jakousi aktualizaci druhého stupně poznávací mohutnosti. Stává se tak termínem neboli zakončením celého procesu smyslového poznání. K teorii vyjádřené species jako termínu jak rozumového, tak smyslového poznání viz Francisco Suárez, *Commentaria una quaestionibus in libros Aristotelis De anima*, ed. S. Castellote [URL:< http://www.catedraldevalencia.es/castellote/deanimav1.pdf>], disputace 5, otázka 5 (dále budeme citovat v následující podobě: *DA* 5, 5).

percepce. 5) „Zapustíme“ Suárezovu teorii smyslového poznání vnějších smyslů do kontextu středověké a renesanční filosofie.

2. Kognitivní smyslový akt

Podle Suáreze popření samostatné existence kognitivního aktu poukazuje na dvojí chybnou ontologickou redukci. První redukce se týká identifikace aktu smyslového poznání se *species sensibilis*. Ze Suárezova pohledu je toto ztotožnění důsledkem čistě receptivistického přístupu k výkladu smyslové percepce, podle něhož hlavní (v určité radikální verzi také jedinou) příčinou aktu smyslového poznání, např. příčinou aktu vidění (*visio*) konkrétní barvy je vizuální *species* této barvy.[11] Druhá redukce se týká ztotožnění aktu smyslového poznání s kognitivní mohutností, tedy, jak uvidíme níže, s duší poznávajícího (*DA* 5, 3, 3).

První typ redukce nachází Suárez v textech Aristotela, Akvinského a jeho žáků. Cituje z jedenácté kapitoly druhé knihy *O duši*, v níž Aristotelés říká, že „... pociťování čili vnímání jest jakýsi druh trpnosti“ (Aristotle 2000, 424a1). V druhé kapitole třetí knihy formuluje Aristoteles své mínění o zásadní aktivitě poznávaného předmětu. V překladu A. Kříže formulace zní: „Skutečná činnost vnímatelného předmětu i smyslu jest již jakási vazba myšlenek jakoby v jednotu“ (425b25-26; Aristoteles 1996, 86). Podíváme-li se do řeckého originálu, pak přesnější překlad (přijatý i W. S. Hettem, viz Aristotle 2000, 147) zní: „Činnost smyslově vnímatelného předmětu a smyslového vnímání je jedna a táž“.[12] Jinými slovy, tím, co aktivuje mohutnost, a tedy působí akt poznání, je smyslově vnímatelný předmět, resp. jeho zástupce, tj. smyslová *species*. Na kognitivní mohutnost tak v jistém slova smyslu zůstává jen pasivní recepce této *species*. V *Teologické Sumě* 1, 17, 2, ad 1 Akvinský tento postoj vyjadřuje lapidárně: „... sensum affici, est ipsum eius sentire“. Být afikován smyslově vnímatelným objektem, rozuměj smyslovou *species*, znamená *eo ipso* vnímat (*sentire*). Není to jen aristotelsko–tomášovská (tomistická) tradice, jíž je tento přístup blízký. Suárez tento

[11] Jak známo, barva ve scholastice představuje vlastní smyslově vnímatelný předmět (*obiectum per se sensibile*) vizuální mohutnosti.

[12] „Η *δε του αισθητού ενέργεια και της αισθήσεως η αυτή μεν εστι και μία* ...“ (Aristotle 2000, 146).

přístup nachází také u významného renesančního peripatetika Agostina Nifa (1469/70-1538) v jeho *De sensu agente*.[13]

Druhý způsob redukce se týká identifikace aktu poznání s poznávající mohutností, či přesněji, podle Suáreze, s modem (pozorné) duše poznávajícího. Mezi představitele tohoto způsobu redukce Suárez řadí Augustina. Uvádí některá *loci* z jeho *De Trinitate*, v nichž Augustin tvrdí, že „mysl a její poznání jsou jedním“, přičemž na jiném místě téhož spisu identifikuje *mens* s bytností duše. Z těchto tvrzení lze oprávněně vyvodit, že duše a její *notitia* jsou (alespoň v některých) Augustinových textech *De Trinitate* považovány za reálně totéž.[14]

Obojí způsob redukce aktu poznání Suárez odmítá. Kognitivní akt představuje z ontologického hlediska *kvalitu*, a jako takový se reálně liší od potence informované prostřednictvím *species*.[15] Suárez tvrdí, že *sensatio* nemůže být identické se *species*. Platí totiž, že zatímco ke *species* se kognitivní mohutnost vztahuje pasivně, k *sensatio* se má aktivně. Vidění červené barvy nelze identifikovat se *species* této barvy. Kognitivní akt, jako každá jiná *vitální* operace oduševnělé substance,[16] nemůže být hýbán zvenčí na způsob behavioristické „stimulace–reakce“, tj. vnějším poznávaným předmětem či jeho

[13] Nifo explicitně říká, že „… *species et sensatio, ut dicendum, sunt una res simpliciter* …“ (Nifo 1517, 128b). Podle slov samotného autora byl tento text dopsán v červnu roku 1495, a spíše než *De sensu agente* se měl jmenovat *Tractatum de Errore Joannis de Sensu Agente* (srov. Nifo 1517, 129a). Pojednání vyšlo jako poslední část publikace *In librum Destructio Destructionum Averrois commentationes*. Text vyšel poprvé v roce 1497 v Benátkách. My budeme vycházet z druhého (benátského) vydání z roku 1517.

[14] „*Et est quaedam imago trinitatis, ipsa mens et notitia eius, quod est proles eius ac de se ipsa uerbum eius, et amor tertius, et haec tria unum atque una substantia*“ (Augustinus 2006b), *De Trinitate*, lib. 9, n. 18, 962. Na začátku deváté kapitoly pak říká: „*Mens uero et spiritus non relatiue dicuntur sed essentiam demonstrant*“, 902.

[15] *DA* 5, 3, 2: „*Unde sit conclusio: Actus cognoscendi est specialis qualitas realiter distincta a potentia, ut specie informata*“.

[16] Tato skutečnost primárně souvisí s Aristotelovou definicí duše: „… *duše je počátkem uvedených sil a že se vymezuje čtyřmi mohutnostmi, vyživováním, vnímáním, myšlením a pohybem*“ (413b12-14; Aristotelés 1996, 54). Co se týče Suárezovy interpretace této definice a jejího vztahu k jinému Aristotelovu vymezení duše („… *duše je první skutečností přírodního těla, které má v možnosti život*“; 412a28-29; Aristotelés 1996, 51) viz zajímavý text *DA* 1, 4.

„zástupcem", ale musí mít svůj původ (také) v samotné duši, jež je principem všech operací dané substance (*DA* 5, 3, 3).

Bude-li někdo redukovat akt poznání na pouhý aspekt pozorné mohutnosti, bude podle Suáreze redukovat tento akt na určitou modifikaci duše samotné. Takovouto redukci však náš myslitel nepřijímá. V souladu s Tomášovou a tomistickou koncepcí, které jsou jinak obecně v Suárezově *De anima* taktéž patrné, jezuita tvrdí, že takováto redukce je v rozporu se skladebným charakterem stvořeného. Vše konečné je, na rozdíl od Boží jednoduchosti, nějak ontologicky složené, což v případě metafyziky (konečné) mysli obnáší reálnou distinkci duše a jejích mohutností.[17]

Tato redukce však má ještě jeden neméně závažný důsledek. Činí postradatelnou intencionální *species*, čímž vylučuje jakýkoli receptivní aspekt v poznání. V rámci jisté úsporné ontologie pak bude možné za postačující integrální princip aktu smyslové percepce považovat duši a existující vnější předmět. Smyslová *species* se stane redundantní, což je v rozporu se Suárezovým předpokladem, podle něhož *každé* poznání, v tomto pozemském stavu,[18] vyžaduje sjednocení mohutnosti s poznávanou věcí. Oproti intuici operující bez *species*,[19] probíhá smyslové poznání na základě vnitřní determinace kognitivní mohutnosti. Vzhledem k tomu, že vnější objekt sám ze sebe není s to bezprostředně působit na poznávací mohutnost, je třeba vtištěné *species*, která tuto determinaci zajistí (*DA* 5, 4, 5).

3. Vznik *species sensibilis* vnějších smyslů

Teze o existenci samostatných ontologických „jednotek" aktu smyslového poznání a *species sensibilis* v rámci Suárezova výkladu mechanismu smyslového poznání souvisí s jedním důležitým metodologickým předpokladem,

[17] *DA* 5, 3, 4: „… *si actus cognoscendi tantum esset animae attentio, potius esset idem cum ipsa anima, quam cum potentia, immo superflua esset potentia distincta; et ita ipsa anima esset sua potentia; et sua potentia et operatio essent idem, quod repugnat rationi creaturae*".

[18] Bezprostřední spojení bez intencionálních *species* je možné u blažených v nebi: „… *si autem contingat aliquando ipsum obiectum immediate uniri potentiae, ut in beatis fieri credimus, tunc non est necessaria species*" (*DA* 5, 2, 7).

[19] Tento způsob poznání byl rozšířen především u františkánských autorů 14. století v čele s Ockhamem. Ohledně historického milníku ve vývoji středověké epistemologie v postavě Ockhama, viz např. Boler (1982).

který na některé výjimky určuje postup zkoumání renesančních aristoteliků: Otázku příčin vzniku aktu smyslového poznání je třeba považovat za odlišnou od otázky vzniku *species sensibilis.*[20] Podíváme-li se na otázku vzniku *species sensibilis*, není nadsazené konstatovat, že bezprostřední historickou determinantu této otázky v období středověkého a renesančního peripatetismu představuje Averroův komentář k páté kapitole druhé knihy Stagiritova spisu *O duši* (417b22-417b29). V něm arabský filosof říká následující:

> Někdo může říci, že smyslově vnímatelné předměty nehýbou smyslem takovým způsobem, jímž existují mimo duši; hýbají jím totiž podle toho, nakolik jsou intencemi, protože v látce nejsou intencemi v aktu, nýbrž v potenci. A nelze říci, že tato odlišnost nastává prostřednictvím odlišnosti subjektu tak, že intence vznikají kvůli duchovní látce, kterou představuje smysl, a nikoli kvůli nějakému vnějšímu hybateli. Lepší je totiž předpokládat, že příčinou různosti látky je odlišnost forem, nikoli že odlišnost látky je příčinou různosti forem. A protože je tomu takto, je nezbytné v případě smyslů uvažovat nějakého vnějšího hybatele, který je odlišný od smyslově vnímatelných objektů, jako tomu bylo u intelektu. Viděli jsme tedy, že pokud připustíme, že odlišnost forem je příčinou odlišnosti látky, pak je nutné, aby existoval vnější hybatel. Avšak Aristoteles v případě smyslů [o tomto hybateli; D.H.] pomlčel, protože zde [v případě poznání vnějších smyslů; D.H.] je něčím skrytým; a odhalil ho v případě intelektu. Ty však musíš toto uvážit, protože to vyžaduje důkladného zkoumání.[21]

[20] K důrazu na tuto odlišnost viz Kennedy (1966); South (2002). Rozlišení těchto dvou otázek je zcela zásadní taktéž pro Jandunův výklad.

[21] „*Et potest aliquis dicere quod sensibilia non movent sensus illo modo quo existunt extra animam; movent enim sensus secundum quod sunt intentiones, cum in materia non sint intentiones in actu, sed in potentia. Et non potest aliquis dicere quod ista diversitas accidit per diversitatem subiecti, ita quod fiant intentiones propter materiam spiritualem que est sensus, non propter motorem extrinsecum. Melius est enim existimare quod causa in diversitate materie est diversitas formarum, non quod diversitas materie sit causa in diversitate formarum. Et cum ita sit, necesse est ponere motorem extrinsecum in sensibus alium a sensibilibus, sicut fuit necesse in intellectu. Visum est igitur quod, si concesserimus quod diversitas formarum est causa diversitatis materie, quod necesse erit motorem extrinsecum esse. Sed Aristoteles tacuit hoc in sensu, quia latet, et apparet in intellectu. Et tu debes hoc considerare, quoniam indiget perscrutatione*“ (Averroes 1953, 221).

Smyslově vnímatelné předměty tak podle arabského filosofa hýbou smysly jedině jako intence, nikoli jako (materiální) akcidenty. Poznáním zelené barvy se poznávající subjekt nestává zeleným. Specifický (imateriální) způsob přijetí formy v poznání není dán specifickou povahou látky orgánu (mohutnosti), ale specifickou povahou přijaté formy. Ta není důsledkem pouhého „emitování" ze strany smyslově vnímatelného předmětu, ale závisí také na existenci jakéhosi Averroem nespecifikovaného vnějšího hybatele. I když Aristoteles v souvislosti se smyslovým poznáním podle Averroa o žádném hybateli nemluvil (zmínil ho pouze v případě činného rozumu – *intellectus agens*), nelze se domnívat, že v oblasti smyslových intencí žádný takový hybatel neexistuje. Jinými slovy, problém, který v tomto úryvku Averroes nadhodil, lze označit za metafyzický problém tzv. ascendenční kauzality, neboli toho, jak lze vysvětlit produkci ontologicky vyššího z ontologicky nižšího. Má-li být, jak tvrdí receptivisté, poznání čistě recepcí vtištěné *species*, pak *species sensibilis* musí vykazovat rysy určité imateriality související s její intencionalitou. Je však patrné, že materiální smyslově vnímatelný objekt není schopen intencionálně afikovat kognitivní mohutnost. Podle některých exegetů, jak to ostatně naznačuje také arabský filosof sám, je daný předmět tohoto schopen jedině jako nástroj vnějšího činitele, kterého někteří, jak uvidíme níže, ztotožnili se separovanou substancí (hybatelem sfér) či s Prvním hybatelem, Bohem, jiní pak s činným smyslem existujícím v duši, a zase jiní se světlem. Ať je však tímto „motorem" cokoli, v každém případě platí, že jeho základní funkcí je prostřednictvím jakéhosi vycházení (*extramissio*) k smyslově vnímatelným předmětům (ať v podobě světla či jemných částic) „pozvedat" pouze smyslově poznatelný objekt na úroveň intencionální *species sensibilis*, podobně jako činný rozum svou iluminací „pozvedá" (materiální) smyslové obrazy (fantazmata) vnitřního smyslu (fantazie) na úroveň inteligibilních *species,* které mohou být přijaty v imateriálním trpném rozumu.

V kontextu tradiční interpretace tohoto nejednoznačného textu Suárez poukazuje na to, že úvaha o *sensus agens* nepředstavuje ani tak odpověď na otázku po vzniku smyslové *species*, jako spíše odpověď na otázku po produkci *sensatio*. Právě z tohoto důvodu při prezentaci *opiniones* jako možných řešení tohoto problému zmiňuje pouze ty koncepce, podle nichž je onen hybatel vůči duši vnější (viz *DA* 6, 2, 2).

Nejen Akvinský ve svém textu *Quaestiones disputatae de potentia*, q. 5 a. 8 co,[22] a někteří z jeho žáků (především Kajetán[23]), nýbrž také Agostino Nifo, ve svém *De sensu agente* (1497), považovali teorii činného smyslu za nauku týkající se vzniku smyslové *species*. Právě Nifo také výrazně ovlivnil Kajetánovu teorii vzniku *species sensibilis*, kterou později podrobil detailní kritice Suárez (srov. Pattin 1988, 419). Jak v hrubých rysech vypadá Nifova nauka?

V přímé návaznosti na výše uvedený Averroův komentář vychází Nifo z obecné analýzy fyzické změny.[24] V každé změně ze strany činitele existují tři základní faktory: Forma daného činitele; látka, ve které tato forma existuje, a jež je tím, co omezuje a diverzifikuje formy různých činitelů; a konečně univerzální *agens*, díky němuž jsou všichni činitelé operabilní. Vzhledem k propojení přirozených jsoucen sféry sublunární s lunárními hybateli sfér se podle Nifa stávají všichni činitelé sféry podnebeské jakýmisi nástroji vyšších činitelů lunárních. Všechny tyto faktory, tolik Nifo, jsou přítomny také ve výkladu kognitivní smyslové operace, jež není změnou tělesnou, ale intencionální (spirituální). Při této operaci je třeba uvažovat samotnou smyslovou formu (forma červenosti), formu omezenou látkou (forma červenosti omezená na partikulární červenost; smysly poznávají pouze jednotlivé) a konečně univerzálního činitele, jímž je podle Nifa Bůh či První hybatel. Tento univerzální *agens* je činný nejen u fyzických změn (prostřednictvím hýbání), ale také v oblasti změn intencionálních (spirituálních), neboť kromě toho, že je Prvním hybatelem, je také (čirým) Intelektem. Smyslově vnímatelné objekty se tak podle Nifa stávají nástrojem vyššího hybatele, který jim jako první a vzdálená příčina poskytuje potřebné spirituální bytí.

[22] *De potentia*, q. 5 a. 8 co.: "*Sed sciendum quod corpus habet duplicem actionem: unam quidem secundum proprietatem corporis, ut scilicet agat per motum (hoc enim proprium est corporis, ut motum moveat et agat); aliam autem actionem habet, secundum quod attingit ad ordinem substantiarum separatarum, et participat aliquid de modo ipsarum; sicut naturae inferiores consueverunt aliquid participare de proprietate naturae superioris, ut apparet in quibusdam animalibus, quae participant aliquam similitudinem prudentiae, quae propria est hominum. Haec autem est actio corporis, quae non est ad transmutationem materiae, sed ad quamdam diffusionem similitudinis formae in medio secundum similitudinem spiritualis intentionis quae recipitur de re in sensu vel intellectu, et hoc modo sol illuminat aerem, et color speciem suam multiplicat in medio*" (Sancti Thomae de Aquino, 2011).

[23] K výkladu Kajetánovy koncepce vzniku *species sensibilis*, viz také Leijenhorst (2007).

[24] V tomto výkladu, kromě Nifova textu, vycházíme také z Mahoney (1971).

Právě díky tomuto bytí mohou být pak intencionálně přijaty v patřičné poznávací smyslové mohutnosti, kterážto recepce, jak zdůrazňuje Nifo, odpovídá samotnému *sensatio*. Právě instrumentalita smyslově vnímatelných předmětů ve vztahu k vyšší, imateriální příčině je garancí toho, že smyslově vnímatelné předměty jsou s to způsobit účinek, který by jinak ze své vlastní přirozenosti schopny vytvořit nebyly (viz Nifo 1517, 128a-b; Mahoney 1971, 128-131).

Přestože Kajetán sdílí s Nifem základní předpoklad problému spojeného se vzestupnou kauzalitou, varuje se toho, aby smyslové *species* upřel vlastní kauzální činnost. Aniž by citoval Nifa, upozorňuje na nežádoucí tendenci arabských myslitelů a jejich latinských stoupenců upírat jsoucnům sublunární sféry (po způsobu jakéhosi okazionalismu) jejich vlastní kauzalitu.[25] Sám navazuje na zmíněný Akvinského text z *De potentia*, kde nachází odpověď na zmíněný teoreticko-kauzální problém: příčinou intencionálního bytí je smyslově vnímatelný objekt, avšak nikoli nakolik je materiální, ale nakolik jeho forma *participuje* na separovaných formách" [kurzíva; D.H.] (Thomas de Vio Caietan 1583, 112a).

Jaký postoj zaujímá Suárez k těmto dvěma koncepcím? Zcela jednoznačně se vůči nim negativně vymezuje. Jednotně odmítá všechny koncepce, které *ad incognitas causas recurrunt* (viz *DA* 6, 2, 3). Problém spatřuje v Kajetánově pojmu participace. V kritické pasáži věnované této otázce tvrdí, že participovanou přirozeností musí být nějaká entita. Existují dvě základní možnosti. Dokonalost vyšší přirozenosti jakožto participované přirozeností nižší může být buď menší, anebo větší, než je vlastní dokonalost vytvořené *species*. Není-li větší, pak zcela jistě nebude postačovat k vytvoření této *species*, neboť je patrné, že věci jsou činné nikoli podle dokonalosti participované přirozenosti, ale podle toho, nakolik se tato participovaná dokonalost nachází v jejich přirozenosti, tedy v přirozenosti nižší, participující. Přestože člověk a anděl mají účast na Boží přirozenosti, nejsou činní způsobem, jímž je činná samotná Boží přirozenost. Bude-li na druhé straně participovaná dokonalost větší, než je dokonalost *species*, zůstává otázkou, zda je vůči smyslově vnímatelným předmětům vnitřní, či nějak (zvnějšku) přidaná. Bude-li jejich vnitřní součástí, pak celá otázka o instrumentalitě smyslově vnímatelných předmětů a potřebě „výpomoci" prostřednictvím vyšší přiro-

[25] „*Et breviter incideretur in lege maurorum ad subtrahendum rebus causalitates suas* …“ (Thomas de Vio Caietan 1583, 112a).

zenosti ztrácí svůj smysl. Vystačíme si s vnitřní výbavou dané *species*. Bude-li někdo naopak hájit alternativu druhou, Suárez se ptá, v čem ona přidaná dokonalost vlastně spočívá (srov. *DA* 6, 2, 5).

Zamítnutí uvedených teorií vede Suáreze k závěru, podle něhož *species sensibiles* vnějších smyslů jsou vytvořeny samotnými smyslově vnímatelnými objekty, které se samy zmnožují v médiu.[26] Abychom tento Suárezův závěr, který se zdá být *prima facie* ve sporu s výše uvedenou Averroovou pasáží, správně pochopili, musíme si nejdříve představit jezuitovu ontologii intencionální (smyslové) *species*.

4. Ontologie smyslové *species*

Zatímco v otázce existence *species sensibilis* vnějších smyslů se široká paleta středověkých a renesančních aristoteliků až na některé (nikoli nevýznamné) výjimky (Ockham, Durandus) shoduje, v analýze povahy této *species* se názory různí. Ve svém výkladu povahy intencionální *species* Suárez vychází z předpokladu, že intencionální *species* je reálné jsoucno, a klade si následující tři otázky: 1) Je z ontologického hlediska *species sensibilis* substancí, či akcidentem?; 2) Jde o materiální, nebo imateriální entitu?; 3) V jakém vztahu se nachází entitativní a intencionální aspekt této *species*? (srov. *DA* 5, 2, 1).

Species sensibilis vnějšího smyslu není podle Suáreze z ontologického hlediska ničím jiným než akcidentem *kvality*. Tato kvalita, kterou v případě vizuálního vnímání je barva, se ze smyslově vnímatelného předmětu šíří ke smyslovému orgánu, oku, prostřednictvím média, kterým může být vzduch, voda či jiné průhledné médium. Vzhledem k tomu, že toto médium vyplňuje prostor mezi viděným předmětem a okem, a že s ohledem na to, že toto průhledné může být nositelem světla, tak barvy, může k tomuto „přenosu“ smyslově vnímatelné kvality dojít. V obou subjektech tedy zmíněná kvalita inheruje.[27] K jejich funkci, tj. k aktualizaci kognitivní mohutnosti, není tedy nutné, aby byly substancemi. S jejich akcidentální povahou souvisí také způsob jejich sjednocení s mohutností, který nepřekračuje akcidentální spo-

[26] *DA* 6, 2, 6: „*In sensibus exterioribus, species producuntur ab obiectis*“. K Suárezově teorii vzniku *species sensibilis*, ontologii této *species* viz také South (2001, především 226-231).

[27] K Suárezově vizuální teorii zahrnující nejen otázku *species in medio*, ale především teorii průhledného, světla a barvy, viz *DA* 7, 1-4.

jení (*unio accidentalis*) (srov. *DA* 5, 2, 4). Jezuita odmítá mínění Kajetánovo, podle něhož spojením intelektu a *species intelligibilis* vzniká „těsnější“ typ jednoty než je substanciální sjednocení formy a látky. Jak známo, substanciálním spojením látky a formy vzniká cosi „třetího“, totiž materiální kompozitum z látky a formy, jež se od svých „částí“ reálně liší. U kognitivní potence intencionálně sjednocené s formou poznaného předmětu však žádné *tertium quid* nevzniká (srov. Thomas de Vio Caietan, 2000, 57). Tento výklad je Suárezovi cizí. Způsob sjednocení intencionální *species* s kognitivní mohutností se podle Suáreze neliší od jakéhokoli jiného sjednocení akcidentu a substance.

Suárez dále praví, že tyto *species* nejsou akcidenty stejného řádu a druhu, jakými jsou smyslově vnímatelné kvality, které je emitují. Jde o entity, které jsou méně dokonalé, než jsou je emitující objekty. Jsou to jen jakési jejich stopy (*vestigia*) či „deriváty“, které se od nich numericky neliší (*DA* 6, 2, 6).[28] Jako takové představují pouze virtuální reprezentace, které je třeba následně zpracovat a „završit“ v aktivní mohutnosti (duši). Kromě toho platí, že jako stopy nejsou smyslově poznány. Představují tak smysly nepoznatelný princip, který patřičnou mohutnost odvádí ihned „ven“ – k poznání vnějšího předmětu (viz *DA* 5, 2, 6).

Jejich subtilní charakter není v žádném případě důsledkem jejich pravé imateriality. Na rozdíl od svých inteligibilních protějšků jsou zcela materiální a dělitelné. Jejich materiální a rozlehlý charakter je dán jejich korespondencí s mohutností, v níž jsou přijímány. Tyto mohutnosti, na rozdíl od intelektu, jsou tělesné a organické. Svou tezi o rozlehlosti smyslových *species* Suárez dokládá na příkladu se zrcadlem a jeho „mrtvou“ recepcí *species*. Je zřejmé, že v zrcadle dochází k rozlehlé a dělitelné reflexi smyslové *species* vysílané vnějším objektem. Zatímco v celém (nerozbitém) zrcadle dochází k reflexi celé tváře zrcadlené osoby, v zrcadle rozbitém na střepy se objevují jen části její tváře (srov. *DA* 5, 2, 17).

Přestože Suárez ve svém výkladu akcentuje akcidentální, materiální a ve srovnání s vnějšími předměty méně „noblesní“ povahu smyslové *species*, představuje tato *species* současně formální podobu (*similitudo formalis*) těch-

[28] Toto tvrzení o numerické identitě je o to pochopitelnější, pokud vezmeme v úvahu, že pro Suáreze principem individuace akcidentů není subjekt, v němž akcidenty inherují, ale jejich vlastní entita (*entitas tota*). K principu individuace akcidentů viz Heider (2011, 311–332).

to předmětů. Reprezentativní charakter patří podle Suáreze k její „esenci". Z entitativního hlediska totiž *species sensibilis* patří k akcidentu kvality, a to druhu dispozice (*dispositio*), nikoli druhu vnější formy (*figura*). Je-li však intencionální *species* dispozicí, pak její stěžejní funkcí je *disponovat* kognitivní mohutnost k „vyšlehnutí" (*elicitatio*) aktu vnímání. *Species* však nemůže disponovat, pokud nereprezentuje. V Suárezově koncepci hraje důležitou roli to, že reprezentativní aspekt smyslové *species* je reálně totožný s jejím aspektem entitativním. Oproti Tomášovi a jeho stoupencům Suárez odmítá reálnou distinkci mezi těmito aspekty. Právě tato reálná distinkce, *inter alia*, umožňuje Kajetánovi jeho formulaci o „více než substanciálním sjednocení" kognitivní mohutnosti a poznávaného předmětu.[29]

Na ontologickém pozadí Suárezovy, z určitého hlediska naturalistické, teze o původu *species sensibilis* jasně vidíme, proč se jezuita neodvolává ani na činný smysl – poznamenává, že jeho „místo" je spíše v otázce formace *sensatio* –, ani na její participaci na vyšší přirozenosti. Takováto řešení, která Suárez považuje za řešení na způsob *Deus ex machina*, jsou v rozporu s jeho přístupem, podle něhož „subtilita" smyslové *species* není dána její imaterialitou, nýbrž „stopovou" materiální entitou, jež je méně dokonalá než stopu zanechavší reprezentovaná věc.

5. Příčiny produkce smyslového kognitivního aktu

Povaha smyslové *species* podstatně určuje Suárezův postoj k otázce příčin aktu smyslové percepce. Ve čtvrté otázce *Zda akt poznání pochází od potence, v níž je poznán společně se species DA* 5 Suárez, nikoli překvapivě, zamítá dvě verze pasivistického řešení. Pokud by *species* byla celkovou příčinou formace aktu smyslové percepce, pak by tento akt měl původ v něčem vnějším. Tím by se však stal operací tranzitivní, nikoli imanentní. Termín (zakončení) tohoto pohybu by se tak lišil od své příčiny, a „přecházel" by v reálně odlišnou entitu (viz *DA* 5, 4, 3). Nepřijatelná je pro Suáreze také slabší verze receptivismu, podle níž *sensatio* má svou účinnou příčinu v mohutnosti informované prostřednictvím *species*, v tom smyslu, že *ratio agendi* aktu smyslového poznání představuje *species*, podobně jako u teplé vody je tím, co za-

[29] *DA* 5, 2, 24: „*… divisio illa communis, qua species intentionalis distingui solet in esse qualitatis, et in esse repraesentativo, non est propria …*"

hřívá její okolí, nikoli voda jako taková, ale její akcident, tj. teplota (viz *DA* 5, 4, 4). Základním principem Suárezovy argumentace proti těmto míněním, která spojuje se jmény Tomáše a Nifa, je Jandunova teze, podle níž intencionální *species* nemůže eficientně bezprostředně spolupůsobit (*concurrere*) při vzniku kognitivní operace. Za předpokladu materiálního charakteru *species* vnějších smyslů nemůže samotná smyslová *species*, jako ontologicky méně dokonalá entita, účinně spoluvytvářet kvalitu dokonalejší, kterou je akt (kvalita) kognitivní operace.

Právě Jandunova teorie činného smyslu představuje důležité východisko Suárezovy diskuse. V *Sophisma de sensu agente*, v němž latinský averroista představuje svou doktrínu činného smyslu jakožto potence potřebné pro vznik *sensatio* (nikoli pro formaci *species sensibilis*), formuluje autor čtyři předpoklady (*fundamenta*), které Suárez více méně přijímá. 1) Vše, co je v něčem přijímáno, závisí na účinném principu; 2) Ontologicky vznešenější je to jsoucno, jehož operace je vznešenější; 3) Účinně (aktivně) vytvářet účinek je vznešenější, než tento účinek přijímat; 4) *Species sensibilis* není dokonalejší než kognitivní mohutnost a její akt (viz Pattin 1988, 129-131). V šestnácté otázce druhé knihy *O duši*, kterou cituje Suárez, pak Jandun uvádí osm argumentů ve prospěch existence činného smyslu, přičemž polovina z nich se opírá o aktivitu smyslů vnitřních (srov. Pattin 1988, 226-228): Kogitativní mohutnost (*vis cogitativa*) je potencí, která vytváří partikulární soudy (Sokrates je tento člověk); společný smysl (*sensus communis*) aktivně rozlišuje mezi jednotlivými předměty vnějších smyslů; a konečně paměť (*reminiscentia*) postupuje diskurzivně od známého k neznámému. Druhá polovina těchto argumentů pak v souladu s Jandunovými *fundamenta* poukazuje na bytostně aktivní charakter produkce aktu smyslového vnímání, kterýžto akt je následně přijat pasivním smyslem (*sensus passivus*). Ve svém důsledném aristotelismu, jehož principem je univerzálně spojovat potenci pasivní s aktivní potencí, Jandun pro každý vnější pasivní smysl uvažuje jemu odpovídající smysl aktivní (srov. Pattin 1988, 143-144). Zatímco pasivní smysl v první fázi smyslového poznání recipuje *species* od vnějšího předmětu, který je pouhým disponujícím činitelem (*agens disponens*), nikoli činitelem eficientně dovršujícím (*agens perficiens*) akt percepce, aktivní smysl na základě této dispozice vytváří *sensatio*, které je následně přijato v pasivním smyslu (srov. Pattin 1988, 159-160). Ve vlastním smyslu kognitivní potencí tak není aktivní smysl, ale smysl pasivní (viz Pattin 1988, 157, 230). Tuto teorii činného smyslu lze považovat, jak to také Suárez činí (*DA* 5, 4, 1), za principielně založenou na axiomu „totéž podle téhož hlediska nemůže být činné a být trpné“ (Pattin 1988, 132).

Je nepochybné, že Suárez chápe činný smysl u Janduna jako mohutnost, která se od smyslu pasivního liší *reálně*, což, dlužno podotknout, ne zcela odpovídá liteře Jandunova textu z *De anima*, podle níž se liší spíše *ratione formali* (Pattin, 1988, 152, 234). Proti teorii reálné distinkce mezi činným a pasivním smyslem uvádí Suárez dva argumenty. 1) Akt poznání představuje imanentní operaci. Imanentní operace však nemůže být přijata v potenci, která se (reálně) liší od potence působící tuto operaci, byť je přijata v identické substanci. 2) Která z těchto dvou mohutností bude kognitivní? Ta, jež akt vytváří, anebo ta, která ho přijímá?[30] Podle Suáreze to nemůže být žádná z nich. Nemůže to být pasivní smysl, protože poznání, podobně jako každá operace duše, znamená vitální činnost (*vitaliter operari*). Tedy akt poznání musí vnitřně obsahovat aktivní složku. Nelze si představit, že by kognitivní potence, která není v žádném smyslu činná, nějakým způsobem poznávala. Stejně tak není plausibilní, aby potence, která je pouze činná, byla kognitivní schopností. V souladu se svým aristotelismem Suárez konstatuje nutnost (pasivní) informace, tj. recepce smyslové *species*. Je-li poznání bytostně asimilativní povahy, tj. připodobňuje-li se poznané věci, pak je zcela nezbytná determinace potence ve formě přijetí intencionální *species*. Suárezův postoj k Jandunově koncepci činného smyslu je jednoznačný: *agere* a *recipere* nelze oddělit způsobem, že bychom je připsali dvěma (reálně) odlišným mohutnostem. Poznání je vytvářeno jednou a touž mohutností.[31]

Uvažujeme-li o integrálním principu formace kognitivního aktu, pak, v souladu se Suárezovou odpovědí Jandunovi, je třeba uvažovat jak aktivitu poznávací potence, tak „aktivitu“ (smyslové) *species*. Suárezovo „conclusio“ proto zní: „Integrálním principem vytvářejícím poznání je potence informovaná *species*“ (*DA* 5, 4, 15). Není však tato formulace, o níž Cees Leijenhorst říká, že se nijak neliší od formulace Kajetánovy (viz Leijenhorst 2007, 256),[32] v rozporu s výše uvedeným principem, podle něhož smyslová *species*, jakožto entita méně dokonalá než akt poznání, nemůže *bezprostředně* spolu-

[30] Výše jsme viděli, že pro Janduna je to mohutnost pasivní.

[31] Nikoli náhodou tato charakteristika platí u Suáreze také pro činný a trpný intelekt (viz *DA* 9, 8, 18).

[32] Právě toto je také důvod, proč C. Leijenhorst považuje Suárezovu kritiku Kajetána jakožto „arci-receptivisty“ za neoprávněnou.

vytvářet kognitivní akt? Neměl by Suárez, v duchu jím přijatých principů, spíše uzavřít, že species *nijak* eficientně nespolupůsobí při vzniku poznání a že jedinou aktivitu v tomto případě má poznávací schopnost, čímž by do značné míry zrcadlil, jak ukazuje Leen Spruit v Spruit (2008), *Zeitgeist* renesanční filosofie?[33]

Vtipem Suárezova řešení je to, že tento svým způsobem extrémní názor není nucen přijmout. V odpovědi na námitku Jindřicha z Gentu (cca 1217 – 1293), jehož teorii o výlučné aktivitě poznávací mohutnosti představuje jako pozici extrémně aktivistickou,[34] Suárez vychází z předpokladu, podle něhož akt poznání je dokonalejší entitou než intencionální (materiální) *species*. Tvrdí, že méně dokonalá věc nemůže spolupůsobit při vzniku věci dokonalejší nejen jako i) bezprostřední celková příčina (silný receptivistismus) či ii) jako příčina hlavní s vedlejším doplněním poznávací mohutnosti (slabší receptivismus), ale také ani iii) jako samostatná příčina instrumentální (sic!). Právě tímto svým tvrzením se podle našeho soudu Suárez odlišil nejen od verzí Tomášovy či tomistické teorie, ale (v případě iii)) také od doktríny Jana Dunse Scota (1266 – 1308), podle něhož akt poznání vzniká na základě spolupůsobení dvou částečných příčin, přičemž mohutnost je příčinou hlavní, *species* příčinou instrumentální.[35] Podle Suáreze, a v tom, domníváme se, spočívá originalita jeho řešení, intencionální *species* spoluvytváří kognitivní smyslový akt jako částečný „doplněk" bezprostřední nástrojové příčiny. Jinými slovy, intencionální *species* kauzálně nespolupůsobí při formaci aktu smyslového poznání jako celková instrumentální příčina „v rukou" příčiny hlavní, jíž je poznávací mohutnosti, jako je tomu u Scota, nýbrž jako částečná nástrojová příčina, jež je spojena s jinou částečnou nástrojovou příčinou, totiž kognitivní mohutností. Tyto dvě částečné nástrojové příčiny pak společně vytvářejí celkovou instrumentální účinnou příčinu, se kterou „operuje" příčina hlavní, kterou je *duše*. Tímto řešením, v němž nelze nevidět

[33] K tomuto závěru dospěl Jacobo Zabarella (1533 – 1589), jak ukazuje South (2002).

[34] Podle ní *species* v žádném smyslu nespolupůsobí, a ani nemůže spolupůsobit, při vzniku aktu poznání, a tak jedinou aktivitu poskytuje poznávací mohutnost, popř. duše (srov. *DA* 5, 4, 8).

[35] Ke Scotovi, který hovoří o rozumovém poznání, viz Joannes Duns Scotus (1893) a jeho *Ordinatio* 1, d 3, q 7, 389: „... *sicut causa superior determinatur ad agendum, concurrente aliqua particulari causa inferiori ..., ita intellectus qui est causa superior et causa illimitata, determinatur ad hoc objectum, concurrente causa particulari determinata ... concurrente hac specie.*" Ke Scotově koncepci viz také Chabada (2005).

jistý rys augustinianismu, Suárez v jistém ohledu řeší výše uvedený problém spojený se vzestupnou kauzalitou. Je-li součástí integrální instrumentální příčiny aktu smyslového vnímání nejen (materiální) *species*, ale i (oduševnělá) mohutnost – je patrné, že s*pecies sensibilis* musí být přijata nikoli mechanisticky pouze v tělesném orgánu, ale v (oduševnělé) mohutnosti –, pak tato nástrojová příčina je dokonalejší či stejně dokonalé přirozenosti, než je kognitivní akt. Tato nástrojová příčina totiž obsáhne obě nepostradatelné složky. Jednou bude dokonalá část, která vnímatelný předmět *ne*reprezentuje, druhou pak část méně dokonalá, která ho reprezentuje. Zatímco první složka poskytuje kognitivnímu aktu entitativní dokonalost, druhá mu „propůjčuje“ aspekt reprezentativní. Přestože platí, že co do druhé složky nepřekračuje nástrojová příčina dokonalost kognitivního aktu, v případě spojení s mohutností je tomu jinak (srov. *DA* 5, 4, 16). Můžeme tedy uzavřít, že pokud bychom si dovolili Suárezovu (syntetizující) formulaci poopravit, řekli bychom, že „Integrálním principem vytvářejícím smyslové poznání je potence informovaná *species* a *duše* poznávajícího“.

6. Závěr

Suárezova teorie vzniku smyslové *species* a aktu percepce představuje podle našeho názoru originální nauku spojující různé „ismy“ středověké a renesanční filosofie. Z podstatné části je suareziánská koncepce určena Averroovou výzvou spočívající v samotné formulaci problému ascendenční kauzality a taktéž řešením této otázky prostřednictvím teorie činného smyslu Jana z Jandunu. Suárezova teorie současně obsahuje rysy augustiniánské tradice, která především ve středověké filosofii představovala důležitý „aktivistický“ doplněk tradice peripatetické. Je zřejmé, že ve srovnání s tomášovskou a tomistickou teorií Suárezovo učení představuje posun ke kognitivnímu dynamismu, který do značné míry odpovídá „duchu“ renesanční filosofie. Tento přístup z velké části reflektuje „fenomenologickou“ skutečnost, že pokud nejsme duchem přítomni, samotné přijímání smyslových *species*, které v hojné míře nastává v každém okamžiku,[36] nemá za důsledek „vyšlehnutí“ kognitivního aktu. Tento akcent na fenomenologický rozměr vě-

[36] Přestože když píšu tento text sedím na židli a přijímám taktilní species povrchu židle, nejsem si toho častokrát vědom.

domé smyslové percepce, vyjádřený výše v druhém „testu“ plausibilní filosofické teorie percepce, podle našeho soudu, představuje významný posun suareziánské teorie vůči tradici aristotelsko-tomistické. Současně však, a to je třeba zdůraznit, v žádném případě neznamená odklon od tradiční koncepce, v níž intencionální *species* mají svou nezastupitelnou kauzální roli při vytváření kognitivního aktu. Právě z tohoto důvodu nelze Suárezovu pozici označit za teorii percepce, která by nezohledňovala výše uvedené (první) kritérium epistemologické. Můžeme tedy uzavřít, že suareziánská teorie příčin perceptivního aktu splňuje obě výše uvedené kritéria „úspěšné“ filosofické teorie percepce.

Literatura

ARISTOTELÉS (1996): *O duši*. Praha: Rezek.

ARISTOTLE (2000): *On the Soul, Parva naturalia, On Breath*. In: *Loeb Classical Library*. Transl. by W. S. Hett. Cambridge (Mass.) – London: Harvard University Press.

AUGUSTINUS (2006a): *De Musica*. In: *Migne Patrologia Latina*. Vol. 32. Dostupné: http://www.documentacatholicaomnia.eu/02m/0354-0430,_Augustinus,_De_Musica_Libri_Sex,_MLT.pdf

AUGUSTINUS (2006b): *De Trinitate*. In: *Migne Patrologia Latina*. Vol. 42. Dostupné: http://www.documentacatholicaomnia.eu/02m/0354-0430,_Augustinus,_De_Trinitate,_MLT.pdf

AVERROES (1953): Commentarium magnum in Aristotelis De anima. In: *Corpus Philosophorum medii aevi corpus commentariorum Averrois in Aristotelem*. Vol. VI, 1. The Medieval Academy of America. Dostupné: http://capricorn.bc.edu/siepm/DOCUMENTS/AVERROES/Averroes_DeAnima_Crawford.pdf

BOLER, J. F. (1982): Intuitive and Abstractive Cognition. In: Kretzmann, N. – Kenny, A. – Pinborg, J. (eds.): *The Cambridge History of Later Medieval Philosophy*. Cambridge: Cambridge University Press, 460-478.

DENNETT, D. C. (1969): *Content and Consciousness*. London: Routledge.

DESCARTES, R. (2000): *Regulae ad directionem ingenii*. Přel. V. Balík. Praha: Oikumené.

FISH, W. (2010): *Philosophy of Perception. A Contemporary Introduction*. New York – London: Routledge.

HEIDER, D. (2011): *Suárez a jeho metafyzika. Od pojmu jsoucna přes transcendentální jednotu k druhům transcendentální jednoty*. Praha: Filosofia.

CHABADA, M. (2005): *Cognitio intuitiva et abstractiva. Die ontologische Implikationen der Erkenntnislehre des Johannes Duns Scotus mit Gegenüberstellung zu Aristoteles und I. Kant*. Mönchengladbach: B. Kühlen Verlag.

JOANNES DUNS SCOTUS (1893): *Quaestiones in primum librum Sententiarum*. In: *Opera omnia*. Vol. 9. Parisiis: Apud L. Vives. Dostupné: https://archive.org/stream/operaomni09duns#page/n5/mode/2up.

KENNEDY, L. A. (1966): Sylvester of Ferrara and the Agent Sense. *The New Scholasticism* 40, No. 4, 464-477.

LEIJENHORST, C. (2007): Cajetan and Suarez on Agent Sense: Metaphysics and Epistemology in Late Aristotelian Thought. In: Lagerlund, H. (ed.): *Forming the Mind. Essays on the Internal Senses and the Mind/Body Problem from Avicenna to the Medical Enlightenment*. Dordrecht: Springer, 237-262.

MACCLINTOCK, S. (1956): *Perversity and Error. Studies on the „Averroist" John of Jandun*. Bloomington: Indiana University Press.

MAHONEY, E. P. (1971): Agostino Nifo's De Sensu Agente. In: *Archiv für Geschichte der Philosophie* 53, 119-142.

NIFO, A. (1517): De sensu agente. In: In librum Destructio Destructionum Averrois commentationes. Venetiis, 124-129. Dostupné: http://www.philological.bham.ac.uk/bibliography.

PATTIN, A. (1988): *Pour l´historie du sens agent. La controverse entre Barthélemy de Bruges et Jean de Jandun. Ses antécédents et son évolution*. Leuven: Leuven University Press.

SANCTI THOMAE DE AQUINO (2011): *Quaestiones disputatae de potentia*. Reprint vydání z r. 1953. Dostupné: http://www.corpusthomisticum.org/qdp5.html

SPRUIT, L. (2008): Renaissance Views of Active Perception. In: Knuuttila, S. – Kärkkäinen, P. (eds.): *Theories of Perception in Medieval and Early Modern Philosophy*. Dordrecht: Springer, 203-224.

SOUTH, J. B. (2001): Suárez and the Problem of External Sensation. *Medieval Philosophy and Theology* 10, 217-240.

SOUTH, J. B. (2002): Zabarella and the Intentionality of Sensation. *Rivista di storia della filosofia* 1, 5-25.

SUÁREZ, F. (2014): *Commentaria una cum quaestionibus in libros Aristotelis De anima*. Castellote, S. (ed.). Dostupné: http://www.salvadorcastellote.com/investigacion.htm

THOMAS DE VIO CAIETAN (1583): *In libros Aristotelis de Anima*. Compluti: Apud Ferdinandum Ramirez.

THOMAS DE VIO CAIETAN (2000): *Summa totius theologiae S. Thomae de Aquino cum commentariis*. Reprint vydání z r. 1588. Hildesheim/Zürich/New York: Georg Olms Verlag.

Organon F 22 (2) 2015: 250-254

Response to Peter P. Icke

EUGEN ZELEŇÁK

Katedra filozofie. Filozofická fakulta. Katolícka univerzita v Ružomberku
Hrabovská cesta 1. 034 01 Ružomberok. Slovenská republika
eugen.zelenak@ku.sk

In 2012 Icke published a controversial book *Frank Ankersmit's Lost Historical Cause* about the development in the views of Frank Ankersmit, one of the most influential philosophers of history these days. Recently, I have reviewed this book, which prompted Icke to respond with a short paper criticizing my review. I welcome his response; nevertheless, I must correct Icke's misinterpretation of my claims. Icke raises several issues but I consider two to be particularly important. Icke is convinced that my objection pointing to shallowness in his book is unsubstantiated. Moreover, he feels disappointed because I allegedly focus only on minor points and consequently I neglect the crucial things he has to say. I argue that both of his criticisms are misguided.

Let me begin with a few words about Ankersmit's position, which is the main topic of Icke's book. Ankersmit is one of the leading philosophers of history who is well known for criticizing a naïve view of historical writing. Especially in his earlier writings, for instance in *Narrative Logic* from 1983, he claims that historical works are never pure and simple depictions of what happened in the past, but complicated and sophisticated constructions. According to (this early) Ankersmit, historians never copy the past events but they create their own special tools to explain the past. This position in philosophy of history is sometimes called *narrativism*. However, later in his writings, Ankersmit defends also what seems to be a view incompatible with his narrativism. He claims that it is possible to have some kind of direct experience with the past. In a nutshell, in his earlier works Ankersmit maintains that no direct access to the past events is possible, but in his later works he suggests it is possible. One of the pressing questions is then why Ankersmit developed his position in this direction. How should we explain his “journey” from his earlier to his later views? How should we account for the move from an “early” to “later Ankersmit”?[1]

In his book Icke discusses both the earlier and later views of Ankersmit: he welcomes the crucial points of Ankersmit’s narrativism but rejects Ankersmit’s later views about experience. Icke’s novel contribution to the ongoing debate about Ankersmit seems to be his explanation for Ankersmit’s “journey” from narrativism to experience.[2] Since I find *this explanation* as something new in the discussion about Ankersmit, in my review I focus on Icke’s account (in fact, he provides two explanations – primary and secondary one) of Ankersmit’s journey. The crucial point to be noted is that when I say something is shallow, it is *this explanation* (more specifically Icke’s primary explanation) I am speaking about. Icke, however, misreads my review. He alleges that I claim it is Icke’s *critique of the later views of Ankersmit*, which is shallow. He writes the following about my review:

> ... in his review he characterizes my primary argument(s) – those marshalled against Ankersmit’s proposal(s) for a direct, unmediated form of engagement with the past through (sublime) historical experience – as ‘shallow and not illuminating at all’ (p. 261), ‘just too shallow to explain anything’ (p. 264) and again, lest the charge of *shallowness* be somehow missed, he finds that my writings constitute ‘a very shallow type of explanation’ (p. 267). Yet nowhere in his review does he even begin to address those primary arguments. (Icke 2014, 531)

Only the last sentence is correct. I do not provide any critical examination of Icke’s *arguments “marshalled against Ankersmit’s proposal(s) for a direct, unmediated form of engagement with the past through (sublime) historical experience”*. In fact, this is the aspect of his book I do not discuss in my review. Instead, I concentrate on something else. I explicitly and repeatedly emphasize that I find Icke’s *explanation of Ankersmit’s journey* shallow (not

[1] This is at least one way the issue might be presented. I am not going to consider here whether this is right.

[2] Several other works analyze or criticize the earlier or later views of Ankersmit so this part of Icke’s book is not so original. The most original part focuses on Ankersmit’s journey from narrativism to experience, although this is the part of Icke’s book I consider contentious.

his critique of Ankersmit's later views). If my objections are read in their proper context and not in an *ad hoc* collage constructed by Icke, it is clear that I discuss Icke's "explanation of Ankersmit's route" or Icke's explanation of "Ankersmit's move from language to experience". Let me quote from my review to document this:

> I try to show that [Icke's] so-called secondary explanation of Ankersmit's route is misguided and incoherent with what Icke himself says in some other places of the book. Moreover, his primary explanation is shallow and not illuminating at all. (Zeleňák 2014, 261)
>
> Moreover, his so-called 'primary explanation' is just too shallow to explain anything. (Zeleňák 2014, 264)
>
> Finally, what is Icke's primary explanation of Ankersmit's move from language to experience? ... To put it briefly, Icke's primary explanation tells us that Ankersmit turns to the topic of experience and direct relation with the past because he *wants* and *needs* such a direct access. But this is a very shallow type of explanation. (Zeleňák 2014, 267)

I nowhere criticize Icke's objections "marshalled against Ankersmit's proposal(s) for a direct, unmediated form of engagement with the past through (sublime) historical experience". Hence, Icke attributes to me a completely different type of criticism, which I do not even attempt to give.

Let me turn now to Icke's second point. Icke claims he is "more than a little perplexed by [my] style of argumentation which alights everywhere on the book's relatively minor points while skipping over, or omitting entirely, the vital points about which its central argument turns" (Icke 2014, 531). Hence, he is disappointed that in my review I focus on "secondary matters" (Icke 2014, 533) and ignore the crucial claims of the book. What, according to Icke's response, seem to be those things I should have concentrated on? Supposedly, it should have been mainly his critique of later Ankersmit (discussion of the issue of experience) – recall that Icke is (mistakenly) convinced that I simply label his critique as shallow and do not give any reasons for that. Indeed, he rightly notices that I do not analyze in detail his critique of later Ankersmit. I must admit, as I did already in my review, that my review is selective. Nevertheless, I believe it is not about "secondary matters".

In my review I concentrate on Icke's explanation of Ankersmit's development, i.e., his route from language to experience. If Icke is right in his complaint, this must be a "relatively minor" issue of his book. But is it really the case? One may give several reasons in support of the negative answer. First, let me remind the reader that Icke's book has a subtitle *A Journey from Language to Experience*. Why would an author mention a secondary issue in the subtitle of his book? Second, besides Introduction and Conclusion Icke's book contains four chapters: the first one deals with Ankersmit's narrativism, the last one with the topic of experience, but the remaining two chapters analyze Ankersmit's route from language to experience or questions closely linked to this development. Why would an author devote almost half of the book to "secondary matters"? Third, Icke himself writes in the book:

> And it is this *shift* between theoretical positions [shift from narrativism to the topic of experience] which, I shall argue, precipitates his fall from the *good* to the *lost* Ankersmit that constitutes this book's central theme. (Icke 2012, 68)

If Ankersmit's development is Icke's "central theme", how is it possible that when I focus on Icke's explanation of this shift I suddenly deal with a "minor point"?

Icke discusses various questions in his book. If he thinks that the topic X is the most important one, I am not going to dispute it. However, it must be obvious to anybody who read the book that Icke's account of Ankersmit's journey from language to experience is not a secondary issue, but, on the contrary, one of the key things explored in the book. Unfortunately, as I argue in my review, Icke's explanation is misguided and shallow.

While it is to be welcomed that Peter P. Icke takes notice of the reviews written about his book,[3] I argued that in his response he misinterprets my main objection targeting his (primary) explanation of Ankersmit's move. As a result he does not even address my real critique of his book and it looks like in our exchange we both simply focus on different topics of interest: I concentrate on the explanation of Ankersmit's journey and Icke concentrates on his critique of later Ankersmit. There is nothing necessarily wrong with that. Except, Icke is convinced that I focus on "secondary

[3] I did not realize this before submitting my review article, but now I see that Icke reacted to and criticized each of the other two published reviews of his book I am aware of.

matters", whereas I believe that there are several reasons pointing to the fact that his explanation of Ankersmit's journey is one of the key topics of his book subtitled *A Journey from Language to Experience*.[4]

References

ICKE, P.P. (2012): *Frank Ankersmit's Lost Historical Cause: A Journey from Language to Experience*. New York: Routledge.

ICKE, P.P. (2014): Author's response to Eugen Zeleňák's review of *Frank Ankersmit's Lost Historical Cause*. *Organon F* 21, No. 4, 531-533.

ZELEŇÁK, E. (2014): Review of "Peter. P. Icke: Frank Ankersmit's Lost Historical Cause: A Journey from Language to Experience". *Organon F* 21, No. 2, 261-268.

[4] I would like to thank the members of our Writing group for their helpful comments.

Organon F 22 (2) 2015: 255-261

On Vít Gvoždiak's "John Searle's Theory of Sign"

PHILA MFUNDO MSIMANG

University of KwaZulu-Natal
South Africa
214525839@stu.ukzn.ac.za

1. Overview

Vít Gvoždiak published a reconciliatory analysis of Searle's social ontology with semiotics in Gvoždiak (2012). Without prior knowledge of his paper, I wrote an analysis of the same subject (Msimang 2014). Even though Searle's social ontology is a common point of reference in the formulation of semiotics in these papers, it also serves as a point of departure in their understanding of semiotics and its development.

The semiotic theory expressed in Gvoždiak (2012) is an inherently linguistic (speech act centred) theory, whereas the semiotic theory presented in Msimang (2014) tends more towards a general theory of communicative systems in which social ontology, which follows from speech act theory, is an interesting part. It is my purpose in this note to contrast the two positions of semiotic theory as they appear in the aforementioned papers in reference to their appropriation of Searle's social ontology.

2. Differences in the understanding and formulation of signs

2.1

Gvoždiak's (2012) basic thesis is that "signs are (1) systematically arranged, (2) arbitrary and (3) social" (Gvoždiak 2012, 149). He argues that they are observer-relative, with the 'observer' being a linguistically competent individual in Searle's sense of the speech act. His notion of the sign and representation is thus fundamentally intentional (Gvoždiak 2012, 152-155). The implicit theory of language which Gvoždiak uses to support this understanding, and what relation it has to Searle's own linguistic theory, is not a matter that can be done justice in this note. Here I focus on the notion of sign he draws from it.

Gvoždiak argues that the sign is representational in nature. By its nature, it must be a conveyor of meaning. As Gvoždiak's notion of sign is a speech act centred one, he goes on to say that "representation is a synonym for intentionality and its manifestation is most obvious in language" (Gvoždiak 2012, 152). This point is central to his discussion of signs (Gvoždiak 2012, 156-157).

The explanatory target of such a notion of sign or the symbolic attribution of meanings, following Searle, is the fact that it is possible to give "symbolic functions to objects and also to processes" in the environment (cf. Msimang 2014, 191) which could be seen as just but another way to say that "every brute fact can serve as the X term in a sign function" (Gvoždiak 2012, 152). In contrast to Gvoždiak (2012), the aim of Msimang (2014) is to show how this is not a complete notion of the sign.

Without going into an analysis of intentionality, one could point out that such a narrow definition of signs limited to the context of speech acts would not recognise non-intentional sign action such as the non-intentional representational content the red milk snake gives off to its would-be predators or the representational content a bee gives in its dance. For such representational content to count as a sign even though it is not always intentionally representational, a different kind of understanding of meaning attribution or the sign is required. In the development of semiotics in reference to social ontology in Msimang (2014, 191), it is argued that status functions, which are an intentional class of signs, "are just the anthroposemiotic-level manifestations of biosemiotic meaning attribution", meaning that human intentionality is not a necessary component in the functioning and creation of signs as representations.

From his speech act centred semiotics, Gvoždiak (2012, 150) argues that "the first semiotic finding is that signs are never intrinsic to our physical world" on the basis of Searle's argument about status functions. Illustrated by way of an example, it is argued that "The wink is a sign (X counts as Y in C) but the twitch is not (it is solely X)," so that X cannot count as a sign in any significant sense in any context it is put in. Given that we accept that signs are not intrinsic to the physical world, it does not follow that involuntary actions such as twitching do not carry semiotic content and cannot be construed as signs.

The position that involuntary movements do not contain semiotic content can only be held in the context of a speech act because non-speech acts do not convey linguistically set meanings. Nevertheless, a twitch can carry symbolic content so that a twitch (X) is a sign (Y) in the context of some situation (C). For instance, twitching can indicate pain, or a twitch-like blinking state can indicate that a person is falling asleep.

In contrast to the central thesis put forward by Gvoždiak, the argument in Msimang (2014) is that signs are not inherently social as there are numerous ways in which meaning is codified in human and natural systems which do not necessarily require the high-level codification systems of natural language (and the social institutions thereof). What is argued in Msimang (2014) is that the speech act is just but a sub-domain of communication systems identified in semiotics and, although significant, should be explicitly situated in the larger triadic theoretical construct which brings together the cultural signs of language with the natural signs of biology.

The argument made is that an 'X' may count as a sign in lower order iterations of meaning below that of social construction so that phenomena like twitches fall into the domain of the involuntary communication systems of the body. The point that Gvoždiak is making about these kinds of behaviours being involuntary and so not part of the semiotics of speech acts is valid, but this finding does not extend to justifying his claim that all signs are social signs. Being involuntary does not exclude a phenomenon from being a sign even if it does exclude it from particular kinds of expressive domains (e.g., speech acts).

Gvoždiak (2012, 150) gives a narrow definition of semiotics as "the study of every possible thing that can be used for lying", implying that signs are inherently intentional or voluntary and that what is not intentional or voluntary cannot count as a sign. But the voluntary/involuntary demarcation is insufficient for separating signs from non-signs. The use of signs in deceit (creating false impressions) in the animal kingdom is profuse and ranges from the somewhat intentional acts of some species of birds limping to convince predators they are easier to catch because of injury (see Gochfeld 1984) or performing other displays to distract predators from the nest which are examples of somewhat voluntary and intentional behaviours animals perform to purposely deceive (Walters 1990), and there are also natural affordances that animals have used to communicate false information such as in Batesian mimicry and camouflage which are examples of non-voluntary signs used for 'lying' or deceit though the animals themselves need not have any intention (let alone any comprehension) of the act to have successfully deceived other agents (see Pasteur 1982). The point of these examples is to show that being voluntary or being involuntary is not a requirement of a sign as signs are manifestly both. Thus, such a distinction cannot be taken as the hallmark of the sign.

Msimang (2014, 187-196) attempts to introduce the reader to the theoretical underpinnings of this general aspect of signs by showing how the logical form 'X counts as Y in C' can be amended so that it can be incorporated in the general definition of meaning – the subject which semiotics claims to deal with – with speech act theory fitting into this picture as a sub-domain of the larger semiotic enterprise (*ibid*). The basic claim is that

> To say that some particular thing, x, counts as something other than itself, y, in some specific context, c, is a semiotic statement describing a "codified" relationship between a sign, x, the symbolic meaning the sign engenders, y, within some system of signs, c. (Msimang 2014, 189)

Contrary to Gvoždiak (2012), Msimang (2014) argues that the context in which signs operate is not limited to human social communication systems but extends into even animal and cellular communication systems (see Sebeok 1962; Emmeche 1999) with some structural qualifications in detail needed depending on the degree of intentionality or the kind of meaning attribution that is at play (Msimang 2014, 196).

2.2

Gvoždiak's (2012, 154) view that "[s]emiotics often concerns itself with the economical nature of the expression plane while striving to find a similar principle on the content plane", is not the way semiotics is understood, at least in the biosemiotic tradition. Semiotics is not generally seen to be a study concerning itself with the efficiency in which expressions are made and communicated, let alone is it seen to be the study of the most efficient solution to the codification and expression problems of communication. Semiotics is construed as the study of communication systems in general (with the exception of Saussurean semiotics which concerns itself with 'psychosocial' signs) and shows specific interest in the workings of actual – rather than ideal – communication systems. The main communication systems of interest in semiotics are found in the natural world and in society at large (see Sebeok – Danesi 2000). It may be that Gvoždiak's notion of the semiotic enterprise in this regard can be traced back to his view of language, reflected in his position that as "language users we are truly *homo economicus*". In contradistinction to this, he also states that

> Since general semiotics involves both closed systems (words) and open systems (sentences, texts), the question of representation arises regardless of whether our language is economical or not and to reduce the sign problem to its economy means to give up the notion of sign as a function. (Gvoždiak 2012, 154)

Instead of trying to understand Gvoždiak's formulation of the sign in terms of his specific linguistic thesis, we might make progress by focusing on the definition of a sign he gives and what that would mean for any formulation of semiotics.

Gvoždiak (2012, 154) argues that "every person can, individually, impose a function arbitrarily upon whichever object they desire" but that such an imposition "however, is not a sign". To make the point, Gvoždiak uses examples in which the representational content needs be set by a community of speakers such as the meaning of a wink or the meaning of the phrase "The President of the Czech Republic" as attributed to his father. It is true that any such statement "completely lacks the collective dimension [necessary for the meaning it is meant to convey because] it would not constitute an institutional fact" (Gvoždiak 2012, 155). It is not clear how this is to preclude non-collective facts from being signs except if institutional meaning was what was being proposed as a new definition of the sign. If a sign is that which stands in place of something else by being in a meaningful relation to it, then individual attributions of meaning would still count as signs. For one, an individual may create a new kind of marking to remember some particular thing so that every time the individual is impressed by the marking it stimulates some particular meaningful memory in that individual in spite of any recognition of that sign as having symbolic significance by a community of speakers. Even though social facts abound, private facts or non-collective meaning attributions (e.g., the wooden ruler that is kept as a back-scratcher) can still count as signs (e.g. wooden ruler = BACK-SCRATCHER).

3. Concluding remarks

Gvoždiak (2012) makes an argument for a collectively intentional and speech act centred construal of the sign and semiotics. Through his interpretation of Searle's social ontology, he came to the view that all signs are necessarily social signs and that the representational content of signs must be intentional in nature and not only intentional in interpretation. He says that "basic semiotic terms suggest that Searle's philosophy offers an explanatory framework to key semiotic questions, namely the differentiation of non-signs and signs, the place of intentionality in semiotic description, and the nature of sign correlations".

On the other hand, the claim in Msimang (2014) is that social ontology is defective as a theory of sign because it is inherently intentional (viz., based on speech acts). I argued that although the iteration of sign-functions in social ontology can continue up indefinitely, their iteration downwards reaches a threshold at the level of the speech act. From that point it is argued that signs, in the form X counts as Y in C, have purchase in the natural world only if signs are not defined in terms of intentionality but rather in terms of the representational structure of the XYC relation.

References

EMMECHE, C. (1999): The Sarkar Challenge to Biosemiotics: Is There Any Information in a Cell? *Semiotica* 127, 273-293.

GOCHFELD, M. (1984): Antipredator Behaviour: Aggressive and Distraction Displays of Shorebirds. In: Burger, J. – Olla, B. (eds): *Shorebirds: Breeding Behaviour and Populations. Behaviour of Marine Mammals.* New York: Plenum Press, 289-377.

GVOŽDIAK, V. (2012): John Searle's Theory of Sign. *Organon F* 19, Supp. Issue 2, 148-160.

MSIMANG, P. (2014): Living in One World: Searle's Social Ontology and Semiotics. *Signs and Society* 2, No. 2, 173-202.

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

SEBEOK, T. – DANESI, M. (2000): *The Forms of Meaning: Modeling Systems Theory and Semiotic Analysis.* Berlin: Mouton de Gruyter.

SEBEOK, T. (1962): Coding in the Evolution of Signalling Behavior. *Behavioral Science* 7, 430-442.

WALTERS, J. (1990): Anti-Predatory Behavior of Lapwings: Field Evidence of Discriminative Abilities. *Wilson Bulletin* 102, No. 1, 49-70.

Organon F 22 (2) 2015: 262-274

Geoffrey Brennan, Lina Eriksson, Robert E. Goodin, Nicholas Southwood: *Explaining Norms* Oxford University Press, Oxford, 2013, x+290 pp.

The aim of this book is nothing less than "explaining what norms are; explaining how and why they emerge, persist and change; and explaining how and indeed to what extent they themselves are capable of explaining our actions, attitudes, and modes of deliberation." Given the number of recent books focused on norms and normativity, this is a bold task, but apart from my reservations – which I will venture later – I must say that on the whole the book is successful. The authors concentrate mainly on norms as resulting from social configurations, putting forward important classifications of such norms (while managing to avoid an overload of classificatory concepts) and the result is an illuminating exposition of the phenomenon of norms.

What do the authors say that norms are? After rejecting some alternative proposals (norms as practices and norms as desires), the authors present their own answer: norms are *clusters of normative attitudes*. Their explanation runs as follows:

> A normative principle P is a norm within a group G if and only if: (i) A significant proportion of the members of G have P-corresponding normative attitudes; and (ii) A significant proportion of the members of G know that a significant proportion of the members of G have P-corresponding normative attitudes.

This accords with my own convictions (and in my recent book I made the very same proposal; see Peregrin 2014). And I am convinced that this is the approach which facilitates the proper understanding of norms.

After the above clarification, the first part of the book moves on to present a distinction between what the authors call *formal* and *non-formal* norms, a distinction which I find highly insightful – not to be confused with that between explicit (or written) and implicit (or unwritten) norms. Formal norms, according to the authors, are characterized: i) by being accompanied by secondary norms (which are norms determining how primary norms are followed), ii) by being enforced in specific ways (whereas non-formal norms cannot be enforced in an institutional way), iii) by being constituted by specific kinds of normative attitudes (while formal norms must be acknowledged individually, informal ones can be acknowledged simply as whatever is issued by an acknowledged institution), and iv) by making demand upon specific kinds of phenomena (while formal norms are aimed at actions, non-formal norms may concern also attitudes etc.). The authors finish this part of their book by considering the two principle kinds of non-formal norms: moral norms and what they call social norms (social norms differing from moral ones only by their greater arbitrariness or conventionality).

In the second part of the book, the authors try to elucidate how norms might emerge, persist and possibly fade away. They give suggestions about how norms may conceivably bootstrap themselves into existence, about which mechanisms may sustain them, and discuss the manners in which they might be abandoned. A key topic for this part of the book is explaining norms by "rational reconstruction", i.e. in terms of identifying the purpose(s) they serve. The authors point out that unless the purpose can be identified also as a reason capable of persuading some agents to adopt the norm, then "rational reconstruction" can perhap gives us an answer to the question *why* the norm exists, but not yet to the question *how* it came into existence. It follows that we must be very careful in assessing exactly what various (for example game-theoretical) models of the emergence of norms are qualified to disclose – and which questions are beyond their scope.

The second part of the book then continues with a chapter devoted to the process in which norms make social life "meaningful", in particular by creating social roles and individual identities and in conferring meanings on various kinds of social actions. According to the authors, identities are just creatures of social actions but what social roles are is somewhat unclear (I do not understand why; I would think that roles are constituted by rules no less than identities; indeed I do not see any clear break between the two). In the final chapter of this part, the authors discuss some illustrative cases of the persistence of bad norms and offer suggestions as to why such things can happen.

In the book's third and last part, attention is focused on specifically how norms may be used to explain human conduct and social reality. The authors categorize three ways in which a norm can be seen to influence human behavior: *norm following* (where an actor does something for the reason that a norm directly tells her to do it), *norm conforming* (where the actor conforms with the norm not directly because of it, but rather because she has other reasons, that are, nevertheless connected to the norm), and *norm breaching* (where the norm makes the actor, in a typical case, do something which goes against the norm). In the final chapter of the book they turn their attention to the psychological aspect of the acknowledgement of norms.

Let me now highlight some places where I feel the authors could have made more of their book. This should be seen as sympathetic criticism – I agree with their "norms as clusters of attitudes" approach, only I think that they not always managed to pursue it to its full consequences.

The first problem I see concerns the explanation of the normative attitudes that are so crucial for the authors' theory. What is their nature? Initially the authors do not say much about this, save citing H.L.A. Hart's notion of "reflective critical attitude to certain behavior". At the very least, I think it should have been clarified whether these attitudes are propositional. (To use the authors' own example, whether a normative attitude towards wearing headscarves is a *conviction that* one should or should not wear headscarves.) Later in the book it slowly emerges that the authors do want to identify normative attitudes with propositional attitudes, indeed with *judgments*. Thus, on pp. 57-8 they write:

> The obvious thing to say, then, is that moral norms and social norms are different in virtue of being constituted by clusters of *different kinds of normative attitudes*. Moral norms are clusters of *moral judgments*. Social norms are clusters of normative attitudes of some other kind – *social judgments*, as we might say.

I think this is problematic: on pain of a vicious circle, such an identification would be acceptable only if judging were not itself a norm-governed activity. But I cannot imagine any account of judging not based on norms.

In Chapter 5, the authors consider how norms can emerge, concentrating on "how, in the absence of any secondary rules, primary rules are imposed, interpreted, applied, and altered". Their unsurprising answer is that such rules must somehow bootstrap themselves into existence. But my point is that, given such bootstrapping must be responsible for there being any rules whatsoever, then there must be normative attitudes that are not propositional attitudes, attitudes that help propositions and propositional attitudes, as specific creatures of norms (in this case of norms of broadly conceived logic) into existence.

A further problem concerns the authors' response to the question of the *function* of norms, of the explanation why something such as norms ever emerged within the evolution of us humans. They reject the *prima facie* obvious answers, namely that norms facilitate coordination and/or cooperation, as not giving the "core function" of norms, and they suggest instead that the core function of norms is "to make us *accountable* to one another". Although this answer need not be wrong, I think it is essentially incomplete. In fact, it seems to me to beg the question: in order to be able to explain what the function of norms is, we would have to know what the function of accountability is (and, unlike cooperation or coordination, the usefulness of which is straightforward, the usefulness of accountability is not). And this is not something the authors really explain.

The most problematic part of the book – by my lights – comes in the beginning of part III, where the authors extensively discuss "internalizing norms". This whole discussion gives the impression that they are abandoning what they had earlier taken a norm to be (*viz.* a complex of attitudes) and shifting towards the understanding of a norm as some kind of *judgment*. (Note that this is something over and above the identification of an individual normative attitude with a judgment, to which the authors had subscribed before – now it is the whole *cluster* of normative attitudes, amounting to a norm, which mutates into a judgment.) For otherwise it would be quite difficult to make sense of what they say: "An individual can be said to have internalized a norm when she is ... disposed ... to treat the norm as a non-instrumental reason to act in accordance with a norm." A cluster of attitudes does not seem to be the sort of item that could meaningfully serve as a reason. (As Davidson (1986, 310) famously claimed "nothing can count as a *reason* for holding a *belief* except *another belief*".)

In my view, if a norm is a cluster of normative attitudes, then "internalizing a norm" would seem to amount to adopting the relevant normative attitude (plus perhaps some kind of acknowledging that others assume the same normative attitude). Then it would seem possible that members of the community in question need not initially be capable of grasping the 'normative facts' in terms of judgments or propositions. In an initial state, their "internalizing of norms" would amount to practical participation in a "normative setting"; and then later they could reflect on this and acquire real, propositional beliefs about the setting, which could subsequently take part in their reasoning.

But despite these reservations, I think this is a good book, recommendable to anybody interested in norms and normativity.

Jaroslav Peregrin
jarda@peregrin.cz

References

Davidson, D. (1986): A Coherence Theory of Truth and Knowledge. In: LePore, E. (ed.): *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell, 307–319. Reprinted (with added "Afterthoughts") in Davidson, D.: *Subjective, intersubjective, objective*. New York: Oxford University Press, 137-153.

Peregrin, J. (2014): *Inferentialism: Why Rules Matter*. Basingstoke: Palgrave.

Petr Kuchyňka – Jiří Raclavský: *Pojmy a vědecké teorie*
Brno: Masarykova univerzita 2014, 139 strán

Recenzovaná kniha *Pojmy a vedecké teórie* nadväzuje na bohatý teoretický odkaz Pavla Tichého ako aj na výsledky práce jeho nasledovníkov – preto neprekvapí, že zvoleným teoretickým rámcom skúmaní bola *Transparentná intenzionálna logika* (TIL). Kniha tematicky spadá najmä do oblasti filozofie jazyka, čo je však vzhľadom na štandardné zameranie skúmaní v TIL zriedkavejšie, aj do oblasti filozofie vedy.

Ako napovedá názov, ústrednou témou knihy sú pojmy a (vedecké) teórie, pričom podľa autorov existuje vzťah vzájomnej závislosti medzi tým, ako pojmovo uchopujeme svet, a tým, aké vedecké teórie uprednostňujeme. Autori postupne vysvetľujú, čo je jazyk, aké sú jeho nedostatky a akými metódami sa tieto nedostatky odstraňujú. V súlade s bežnou praxou TIL chápu významy ako konštrukcie a aj pojmy explikujú ako konštrukcie určitého druhu. Treba poznamenať, že napriek kriticky malému priestoru sa nevyhýbajú ani mnohým náročným a dlho diskutovaným otázkam a problémom (nielen) analytickej filozofie. Usilujú sa napríklad vysvetliť, čo je pravda, fakt, vedecká teória či pravdeblízkosť (*verisimilitude*), a riešia viacero problémov a paradoxov. Útla knižka pozostáva zo siedmich kapitol, ktoré sú ďalej prehľadne členené na podkapitoly.

Prvá kapitola začína úvahami o tom, že pri hľadaní najlepších vedeckých teórií predkladáme protichodné požiadavky: silu a jednoduchosť. Autori považujú cestu za najlepšou teóriou zároveň za cestu za najlepším jazykom, preto je dôležité vysvetliť, čo je vlastne jazyk. Zavádzajú dôležité rozlíšenie medzi jazykom$^!$ v zmysle normatívneho systému, ktorý reguluje správanie členov nejakého spoločenstva s cieľom umožniť komunikáciu, a jazykom$^\#$ v zmysle kódu, konkrétnejšie parciálnej funkcie z množiny semivýrazov do množiny významov. Zavedené rozlíšenie následne efektívne aplikujú na riešenie niektorých známych problémov či paradoxov analytickej filozofie. Druhá polovica kapitoly je venovaná nedostatkom jazyka$^!$, medzi ktoré patrí nejednoznačnosť, vágnosť a nekonzistentnosť.

Prirodzene, existujú metódy, ako nedostatky jazyka$^!$ odstraňovať. Druhá kapitola sa venuje práve takýmto metódam. Prvou z nich je zavedenie (analytickej či syntetickej) definície. Niektoré termíny však musia byť primitívne (nedefinované), aj tie však môžeme špecifikovať pomocou implicitných definícií či pomocou metódy redukcie. A napokon, nepresné pojmy možno explikovať, teda nahradiť nejakým presným pojmom.

V tretej kapitole autori načrtávajú teoretický rámec TIL spolu s jeho najdôležitejším prvkom – pojmom konštrukcie. Základné definície a vymedzenia pochádzajú ešte od Tichého, no autori ich zaujímavo dopĺňajú a následne efektívne aplikujú na riešenie problémov. Zavádzajú napríklad termín *primárny atribút*, pomocou ktorého riešia Quinov problém s (údajnou) nevymedziteľnosťou referencie ako aj Goodmanovu *novú záhadu indukcie*. Čitateľa, ktorý sa venuje TIL, navyše iste potešia niektoré sympatické detaily. Napríklad na s. 46 autori píšu, že do bázy patria súbory *explikátov* indivíduí, afirmatívnych kvalít Áno a Nie, intuitívnych možných svetov a časových okamihov. V literatúre venovanej TIL sa pritom pomerne bežne píše, že ide o súbor indivíduí, pravdivostných hodnôt, atď. – čo je nepresný spôsob vyjadrovania. Vyzdvihnúť možno aj začlenenie Tichého systému dedukcie, ktorý je následne využitý na objektuálne vysvetlenie definícií (hoci o týchto výsledkoch vieme už zo starších prác Jiřího Raclavského).

V nadväznosti na tradíciu založenú Pavlom Maternom autori v štvrtej kapitole explikujú pojmy ako konštrukcie určitého druhu. Autori vychádzajú z (Maternovho) pojmu pojmového systému a ukazujú, že bude výhodnejšie pracovať s iným pojmom – s pojmom derivačného systému, pretože je plodnejší a všeobecnejší ako pojem pojmového systému. Za zmienku stojí aj fakt, že autori explicitne diskutujú o kumulativite v Tichého teórii typov (pozri s. 65).

Piata kapitola sa zaoberá najmä pojmami pravdy, pravdivosti a faktov. Podľa autorov sa pojem pravdy týka primárne propozícií (t. j. funkcií definovaných na možných svetoch a časoch), ide teda o mimojazykovú záležitosť. Pravdivosť konštrukcií propozícií je odvodená od pravdivosti propozícií. A napokon pravdivosť jazykových výrazov je odvodená od pravdivosti konštrukcií propozícií. Zároveň ukazujú, že takáto koncepcia je imúnna voči každej verzii paradoxu luhára. Následne navrhujú chápať fakty ako propozičné konštrukcie, ktoré môžu byť pravdivé, čím sa vyhýbajú viacerým problémom, ktorým musia čeliť koncepcie faktov ako neštruktúrovaných entít.

Šiesta kapitola sa venuje pojmu teórie a zrejme ju možno považovať za najoriginálnejšiu kapitolu knihy. Autori definujú vedeckú teóriu ako (neúplnú) odpoveď na otázku „Čo je pravda" (resp. „Ktorá najsilnejšia propozícia je pravdivá"), ktorá komprimuje nejaký súbor empirických alebo teoretických hypotéz a je v súlade s doterajšími skúsenosťami a výsledkami relevantných experimentov.

Posledná, siedma kapitola sa zaoberá spôsobom porovnávania vedeckých teórií podľa toho, nakoľko sa blížia k pravde. Autori zastávajú modifikovanú verziu Tichého a Oddieho koncepcie a ukazujú, že ak si uvedomíme, že pojem pravdeblízkosti je (zjednodušene povedané) závislý od pojmu derivačného systému, vyhneme sa Millerovej námietke voči Tichého koncepcii.

Teraz uvediem niekoľko pripomienok k textu. Autori na s. 16 píšu, že výraz určitého jazyka[1] je sekundárne vágny práve vtedy, keď je zložený a niektorá z jeho zložiek je primárne vágna. Výraz jazyka$^{!}$ pritom definovali ako určitú *množinu* semivýrazov (s. 14-15). Množina však zložená nie je a nemá zložky, iba prvky či podmnožiny. Vzhľadom na predošlý text knihy môžu byť zložené semivýrazy, tie však nie sú tým, čomu sa pripisuje vágnosť či nejednoznačnosť. Preto sa domnievam, že by bolo potrebné buď revidovať vymedzenie výrazu, alebo vymedzenie sekundárnej vágnosti, pričom by bolo zrejme prospešné využiť rozšírenie Cmorejovej teórie semivýrazov z práce Zouhar (2011, 144-150), ktoré je aj tak potrebné na vysvetlenie rozlíšenia lexikálnej a syntaktickej nejednoznačnosti, ktoré sa v recenzovanej knihe zavádza (s. 15).

Za problematické považujem aj chápanie nepresného pojmu ako *množiny* možných významov nejakého vágneho či nejednoznačného výrazu (s. 35). Nepresné pojmy potom vôbec nie sú pojmami, čo je značne neintuitívne. Úlohu intuícií vo filozofii v žiadnom prípade nepreceňujem, no sami autori v úvode práce vyhlasujú, že úloha hľadania významu nie je ľubovoľná a že jej nesprávne riešenie odporuje jazykovým intuíciám (s. 5). Preto si dovolím uvádzať aj pripomienky tohto druhu. Závažnejší problém však vzniká, keď prejdeme v texte o pár riadkov nižšie: Nepresný pojem má totiž figurovať ako vstupná entita pre explikáciu, teda ako *explikandum*. Spresňovaný pojem, explikandum, teda opäť nie je pojem. To však vedie automaticky k problému s prvým kritériom adekvátnej explikácie, podľa ktorej má byť explikát podobný explikandu v tom zmysle, že ho možno použiť vo väčšine prípadov namiesto neho (s. 35). Lenže ak explikandum nie je ani len konštrukciou, vôbec ho nemožno použiť. Prvé kritérium adekvátnej explikácie sa potom stáva triviálnym, a teda úplne zbytočným. Je tiež otázne, či možno prijať takúto analýzu v prípade postojov indivíduí k nepresným pojmom (najmä v súvislosti s argumentáciou z Tichý 1988, 12-13). Z týchto dôvodov sa domnievam, že by bolo predsa len vhodnejšie považovať nepresné pojmy za otvorené konštrukcie určitého druhu. Vzhľadom na prijatú teóriu pojmov by síce opäť neboli pojmami, no vyhli by sme sa problému s prvým kritériom adekvátnej explikácie aj problémom, ktoré sa vynoria na obzore hneď, ako začneme uvažovať o postojoch k nepresným pojmom.

Jednu okrajovú pripomienku možno uviesť aj k tvrdeniu, že možnosvetové propozície chápané ako významy viet v intenzionálnej sémantike vystihujú intuíciu o logickej sile tvrdenia, pretože propozícia je tým silnejšia, v čím menej možných svetoch a časoch platí, teda čím viac okolností vylučuje (s. 39). Toto tvrdenie je bez ďalšieho vyjasnenia problematické. Ak pracujeme s množinami dvojíc možný svet – časový okamih (t. j. svetamihov) s kardinalitou aspoň $\aleph_0$, môže nastať situácia, v ktorej sa nám intuitívne *zdá*, že jedna takáto množina má viac prvkov ako druhá, no fakticky majú rovnakú kardinalitu. Problém však možno ľahko vyriešiť tým, že povieme, že propozícia A je silnejšia ako propozícia B práve vtedy, keď propozícia A platí vo všetkých svetamihoch, v ktorých platí B, no existuje aspoň jeden svetamih, v ktorom B platí, ale A neplatí.

Nie je mi celkom jasné, prečo sa v definíciách na s. 46, 53 a 54 píše o súbore totálnych a parciálnych funkcií, keď totálne funkcie sú prípadom parciálnych.

Autori, nasledujúc Tichého, tvrdia, že veličiny možno explikovať ako intenzie, ktoré majú ako svoje hodnoty čísla, a ako príklad uvádzajú výšku priemerného obyvateľa planéty (s. 48). Domnievam sa, že by bolo potrebné do analýzy veličín nejakým spôsobom zaradiť aj jednotky, ak má byť prijateľná. Ak sa totiž napríklad pýtam, koľko meriam, nie je mi jasné, aké reálne číslo by malo byť správnou odpoveďou: 1770, 177, 17,7 či 1,77 (atď.)? Treba však upozorniť, že Tichý (1988, 176) jednotky explicitne spomínal.

Navrhovaná objektuálna teória definícií (s. 55-58) je možno prijateľnou teóriou analytických definícií, formálne uchopenie syntetických definícií by však vyžadovalo komplikovanejšiu teóriu, keďže sa v nich prechádza z jedného derivačného systému do druhého.

Text miestami dopláca na rozsah. Napríklad pri práve spomínaných definíciách absentuje reflexia aktuálnej literatúry. Ďalší príklad neblahých dôsledkov malého rozsahu knihy možno nájsť na s. 63, kde autori vysvetľujú analýzu doxastických viet. Autori propagujú výhody chápania doxastických (či epistemických) postojov ako postojov ku konštrukciám (vyhneme sa tak paradoxu vševedúcnosti), no úplne ignorujú diskusiu, ktorá prebehla v Duží – Materna (2001). Z tejto diskusie je pritom zrejmé, že aj chápanie doxastických postojov ako postojov k propozíciám má určité nezanedbateľné výhody, a preto by stálo za zváženie, či nepracovať radšej s postojmi oboch druhov (t. j. s explicitnými aj implicitnými postojmi). Prílišná stručnosť je podľa mňa na škodu aj na s. 66-67. Tvrdenie, že otvorené konštrukcie nemôžu byť pojmami, by vyžadovalo presvedčivejšie zdôvodnenie ako to, že sa „nezdá", že by existoval pojem, ktorý obsahuje voľnú premennú. Čo napríklad s protichodnou intuíciou, ktorá hovorí, že nepresné pojmy (chápané ako otvorené konštrukcie) by mali byť pojmami?

Za problematickú považujem aj autormi predkladanú definíciu faktov (s. 85). Ako som už písala, autori za fakt považujú propozičnú konštrukciu, ktorá *môže* byť pravdivá. Dôsledkom je to, že niektoré fakty nie sú pravdivé, resp. že existujú aj neplatné fakty. To je však značne neintuitívne, pretože sa zdá, že „neplatný fakt" je *contradictio in adiecto*. A opäť pripomínam, že autori sa zaviazali k dodržiavaniu intuícií. Ďalším kontraintuitívnym dôsledkom takejto koncepcie je to, že každá empirická veta vlastne vyjadruje ako svoj význam niečo, čo je faktom. Autori tiež píšu, že ich návrh vykazuje pozoruhodnú zhodu s Fregeho názorom, ktorý však tvrdil, že fakt je myšlienka, ktorá *je* pravdivá, nie tá, ktorá môže byť pravdivá.

Ďalej na s. 91 píšu, že propozíciu, ktorá charakterizuje určitý možný svet, možno konštruovať nekonečne mnohými konštrukciami, no iba jedna z nich je vzhľadom na derivačný systém zložená iba z primárnych faktov. Čo však napríklad konštrukcia, ktorá vznikne vymenením poradia konjunktov v takejto konštrukcii?

Na s. 100 autori píšu, že priesek propozícií (t. j. prienik definovaný pre propozície) P a Q je propozícia pravdivá práve v tých svetamihoch, v ktorých je pravdivá P alebo Q (a nepravdivá v ostatných), a spojenie propozícií (t. j. zjednotenie definované pre propozície) P a Q je propozícia, ktorá je pravdivá práve v tých svetamihoch, v ktorých je pravdivá P aj Q (a nepravdivá v ostatných) – v prvom prípade by však malo ísť práve o spojenie (zjednotenie) a v druhom o priesek (prienik).

Drobným formálnym nedostatkom je (miestami) trochu nedôsledné citovanie. Napríklad na s. 103 odkazujú autori na Tichého a Oddieho článok z roku 1982, správne by však mal byť uvedený rok 1983 (ktorý je uvedený aj v literatúre, s. 135). Na s. 111 zase odkazujú na Tichého stať označenú „(1978a)“, práca, na ktorú chceli odkazovať, však figuruje v literatúre pod označením „(1978)“. Ďalším príkladom sú Tichého práce z roku 1986 či práce z roku 1994, ktoré nie sú rozlíšené (s. 134). Na stranách 91 či 113 sa objavili preklepy („větě“ namiesto „světě“ a „*h-r-w*-ština“ uvedená dvakrát, v druhom prípade však malo ísť o „*h-m-a*-štinu“). To sú však len drobné formálne detaily.

Knižka je možno až príliš stručná, napriek tomu však veľmi obsažná, nápaditá a inšpiratívna. Čitateľ znalý TIL v nej môže nájsť mnoho inšpirácií pre ďalšiu prácu a čitateľ, ktorý sa TIL-ke nevenuje, v nej môže nájsť mnoho motivácií, prečo sa týmto systémom začať zaoberať – prečítanie knižky teda v každom prípade odporúčam.[1]

Daniela Glavaničová
dada.baudelaire@gmail.com

Literatúra

DUŽÍ, M. – MATERNA, P. (2001): Propositional attitudes revised. In: Childers, T. (ed.): *The LOGICA Yearbook* 2000. Praha: FILOSOFIA, 163-173.

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

ZOUHAR, M. (2011): *Význam v kontexte*. Bratislava: aleph.

[1] Ďakujem M. Kostercovi a M. Zouharovi za pripomienky k predchádzajúcej verzii recenzie.

Stanislav Sousedík: *Kosmologický důkaz Boží existence v živote a myšlení* Praha: Vyšehrad 2014, 176 strán

Zvýšený záujem o kozmologické argumenty pre Božie jestvovanie vyvolal na konci minulého storočia pomerne populárny kalámsky kozmologický argument „najlepšieho vysvetlenia“ od W. L. Craiga a pravdepodobnostný argument „najjednoduchšieho vysvetlenia“ od R. Swinburna. V publikácii profesora S. Sousedíka ide o iný druh argumentu, doslova o metafyzikálny dôkaz, ktorý je menej diskutovaný, menej známy a ťažšie pochopiteľný v dnešnej kultúre a súčasnej filozofii. O to vzácnejšie je úsilie autora o jeho priblíženie a obhajobu, a vyvoláva zvedavosť, ako sa mu podarí odpovedať na námietky, ktoré spôsobili nepopulárnosť tohto klasického dôkazu a niekedy aj jeho explicitné odmietnutie zo strany filozofov.

Kozmologický dôkaz predstavený v publikácii je vsadený skôr do kontextu kontinentálnej filozofickej tradície ako analytickej, a je pomerne originálne rozpracovanou treťou „cestou“ Tomáša Akvinského (z náhodilosti vecí k nutnej príčine). Dôkaz je rozpracovaný v líniách tradičnej aristotelovskej metafyziky a je blízky Leibnizovmu argumentu, pričom jeho autor zohladňuje (okrem iného) kritiku D. Huma a I. Kanta a používa niektoré myšlienky z diel M. Heideggera a G. Fregeho.

Kniha je rozdelená do troch častí, pričom prvá a posledná časť hovoria o tom, ako tento dôkaz zapadá do bežného života človeka. V prvej časti autor opisuje detský svet, ktorý je zaujímavý, priateľský a v istom zmysle nadčasový, akoby mal „trvať večne“. Detský „večný“ svet sa neskôr v dospelosti rozpadá, pod vplyvom reálnych starostí, povinností, neopätovanej lásky a zvlášť straty (smrti) blízkych. Dospelý človek si s úzkosťou uvedomuje konečnosť a nestálosť svojho sveta, ktorý pre neho práve ako dôsledok konečnosti, akoby strácal zmysel. Svet sa mu začína javiť mŕtvy, bezvýznamný, niekedy až krutý. V takomto stave keď človek stráca zmysel svojho života (víziu jeho plného a definitívneho naplnenia), môže zaujať postoj zúfalstva (beznádeje) alebo hľadania. Hľadajúci človek sa zamýšľa nad tým, či existuje také dobro, ktoré by vrátilo jeho životu zmysel. Skôr ako sa nad tým zamyslí autor knihy (vo svojom dôkaze), odpovedá na otázku, aké formálne vlastnosti by malo mať toto dobro, aby napĺňalo danú úlohu v živote človeka. Autor zdôvodňuje, že dobro by malo byť (1) večné, malo by mať pozitívne nepomíňajúci charakter, (2) nemá byť prvkom sveta, pretože svet ako celok sa stal cez úzkosť problémom (ktorému treba prinavrátiť zmysel) a (3) nemá byť imanentné ľudskému vedomiu, pretože to je vnorené do úzkosti (s. 46-47). V prvej časti knihy sú najdôležitejšie práve tieto tri vlastnosti (alebo podmienky) dobra, ktoré môže človeku prinavrátiť celkový zmysel života a objasnenie, prečo je kozmologický dôkaz, ktorý hovorí o existencii tohto dobra, dôležitý pre každú ľudskú bytosť.

Druhá časť začína objasnením predpokladov dôkazu, špeciálne pojmu existencie. Dôležitá je konkrétna, „absolútna“ existencia súcien, ktorá ich aj nejakým spôsobom charakterizuje, nielen „relatívna“ existencia, ktorú pripisujeme pojmom na základe existencie súcien, ku ktorým sa vzťahujú (s. 78). Autor ukazuje, že Fregeho kritika takto chápanej existencie vedie k nekonečnému regresu, a preto ak neexistuje nijaký argument proti, je rozumné zostať pri takomto chápaní existencie, ktoré je implementované v prirodzenom jazyku. Druhým dôležitým predpokladom dôkazu je ontologická podvojná štruktúra súcien pozostávajúca z možnosti (*potence*) a aktu, ktoré sú odvodené z „konkrétneho obsahu“ (esencie) a „konkrétnej existencie“. Na túto podvojnú štruktúru (a náhodilosť) súcien sa vzťahuje princíp príčinnosti, že „každé svetské (t. j. náhodilé) súcno má nutne príčinu, totiž príčinu toho, že existuje“ (s. 117), pretože pasívna možnosť nenadobúda akt existovania sama zo seba. Nutnosť tejto príčinnosti na rozdiel od príčinnosti v prírodných vedách vyplýva práve z podvojnej ontologickej štruktúry súcna, o ktorej prírodovedecká príčinnosť nehovorí, a preto aj jej úspech pri objasňovaní príčin je len relatívny alebo pravdepodobnostný.

Východisková štruktúra dôkazu je táto: 1. Ak je niektoré svetské súcno náhodilé, v reálnom svete jestvuje aj nejaké (najmenej jedno) súcno, ktoré je (absolútne, logicky) nutné. 2. Avšak niektoré svetské súcna sú náhodilé. Preto, v reálnom svete jestvuje súcno, ktoré je (absolútne, logicky) nutné (s. 127). Reálny svet v sylogizme nezahŕňa len veci s priestorovými vzťahmi, ale je otvorený aj iným možnostiam. Rozhodujúcou vlastnosťou je to, že existuje nezávisle od ľudského poznania (s. 20-21). Náhodilosť súcna v druhej premise znamená náhodilosť následnú a simultánnu, čo znamená, že veci vo svete nie len vznikajú a zanikajú, ale medzi sebou interagujú a môžeme ich aj ovplyvňovať (s. 129). Tieto spresnenia významov a druhá premisa sú prijateľné a možno súhlasiť s autorom, že niet dôvodu na ich odmietnutie.

Logická implikácia v prvej premise je však problematickejšia. Autor pri hľadaní vysvetlenia náhodilosti súcna, zvažuje tri možnosti: materiálny svet ako celok, svet ako množinu konečných súcien a množinu nekonečného počtu náhodilých súcien. V každom prípade sa pri hľadaní príčiny náhodilého sveta alebo množiny náhodilých súcien dopracováva k nutnému súcnu. Iná možnosť neexistuje. V celkovom zdôvodnení argumentu sa predpokladá princíp príčinnosti, podvojná štruktúra možnosti a aktu, a to, že kontingentnosť nie je kumulatívnou vlastnosťou (ako napríklad hmotnosť). Autor odpovedá na námietky D. Huma a B. Russella, pričom zdôrazňuje, že pri kozmologickom argumente nejde len o akúsi množinu súcien, ale o reálne existujúci celok, ktorého príčina musí jestvovať.

Ďalším krokom je priblíženie klasického pojmu Boha, ktorý je po zodpovedaní Kantových námietok identifikovaný s absolútnym, logicky nutným súcnom zo záveru dôkazu. Toto súcno je úplne jednoduché a je subsistujúcim jestvovaním (bytím), je nepomíňajúce a je čírym aktom. Úplne v zhode s autorovým chápaním existencie súcien, Božie jestvovanie mu vychádza, ako „dokonalosť všetkých dokonalostí", v ktorej sa završuje všetka dokonalosť súcien (s. 156). O ďalších Božích vlastnostiach sa autor v knihe nezmieňuje, pretože by sa tým otvorila nová problematika analógie a celkove reči o Bohu, čo nie je potrebné pre ciele publikácie. Zdôrazňuje však, že logickým dôsledkom argumentu je to, že číry akt jestvovania nemôže byť ničím nútený k tvoreniu a pritom nemožno pochybovať, že je činný. Preto je navonok činný slobodne a správne o ňom môžeme hovoriť, ako o slobodnom Stvoriteľovi.

Pri analýze náhodilosti a nutnosti autor definuje logicky nutné súcno takto:

> *Za absolútne, teda logicky nutné, budeme považovať indivíduum x práve vtedy, keď 1. x existuje v nekonečne dlhom časovom úseku tak, že 2. v žiadnej jeho ľubovoľne dlhej časti by nemohlo ani následne ani simultánne neexistovať.* (s. 123)

Podľa tejto definície by číry logicky nutný akt jestvovania bol večne trvajúcim, teda v akomsi čase, aj napriek tomu, že má v sebe všetku plnosť a dokonalosť. Ide o pozoruhodnú úpravu v porovnaní s tradičným bezčasovým pojmom.

V tretej, najkratšej, časti autor konštatuje, že Božie jestvovanie spĺňa druhú a tretiu podmienku, aby mohlo dať ľudskému životu zmysel (nie je prvkom materiálneho sveta a nie je bezprostredne zakusovaným v ľudskom vedomí, s. 167) a skúma splnenie prvej podmienky, či mu zmysel aj skutočne dáva. Odpoveďou však je prekvapivé „nie". Boh, ku ktorému sa dopracovávame v tomto argumente, Boh filozofov všeobecne, túto podmienku nespĺňa. Avšak „súcno troch podmienok" by ju mohlo spĺňať, mohlo by sa stať večným zmyslom ľudského života, ak by sa k tomu slobodne rozhodlo (s. 170). Má k tomu všetky predpoklady. Filozofia podľa autora knihy v tomto ohľade mlčí. Pravdepodobne sa myslí aristotelovská alebo systematická filozofia, ale nie filozofia náboženstva všeobecne.

Celkovo sa kniha číta veľmi dobre, autor používa zrozumiteľný jazyk a zaujímavé príklady zo života i literatúry (Máchov *Máj*, *Babička* B. Nemcovej). Pri objasnení predpokladov a premís dôkazu je čítanie náročnejšie a vyžaduje zvýšenú pozornosť. Štruktúra prístupu je pomerne tradičná, najprv dôležitosť témy, potom pojem Boha a nakoniec argument, v tomto prípade priamo dôkaz Božieho jestvovania. Dôležitým prínosom je vsadenie problematiky do kontextu života bežného človeka a jemu zrozumiteľného jazyka vo forme objasnenia „troch podmienok“. Z hľadiska filozofického je zaujímavá syntéza odpovedí na námietky moderných a súčasných kritikov kozmologických argumentov a ako autor optimalizuje svoj dôkaz, aby sa vyhol zbytočným nedorozumeniam a kritike. Dobrým príkladom je vyhnutie sa otázke vzťahu času a bezčasovosti.

Vážnejší kritik by mohol namietať proti tomu, ako je objasnený princíp príčinnosti a Božie atribúty. Ohľadom príčinnosti, autor pri „derivácii“ účinku z príčiny hovorí o metafore, ktorú sa len pokúša naznačiť, ale ju ďalej neobjasňuje (s. 118). Podobne autor neobjasňuje, ako by bolo možné bližšie zadefinovať slobodu Božiu, slobodu číreho aktu jestvovania. Tieto dve komplikované témy by pravdepodobne podľa autora vyžadovali ďalšiu štúdiu, a preto sa im hlbšie nevenuje. Pravdou je však aj to, že práve na nich závisí, či tento dôkaz je prijateľnejším, než súčasné pravdepodobnostné kozmologické argumenty W. L. Craiga a R. Swinburna. V celku však ide o dôležitý príspevok do rodiny súčasných kozmologických argumentov, inšpiratívny pre všetkých aristotelikov a tomistov, ako aj pre každého čitateľa, ktorý si dá námahu a pochopí základné tézy dôkazu.

Ľuboš Rojka
lubosrojka@gmail.com

Organon F 22 (2) 2015: 275-280

Petra Vopěnky kopernikovský obrat aneb matematika mezi náboženstvím a empirií

věnováno památce profesora Petra Vopěnky (1935-2015)

Pokud má mít otázka po ospravedlnění nějakého opatření dobrý smysl,
který přesahuje důkaz bezespornosti, pak spočívá jedině v tom,
zda toto opatření vede k odpovídajícímu úspěchu.
Úspěch je zde nejvyšší instancí, před kterou se musí každý sklonit.
David Hilbert

Matematické nekonečno je převzato ze zkušenosti, i když nevědomky.
Může tedy být vysvětleno jen ze skutečnosti a ne ze sebe sama,
z matematické abstrakce.
Friedrich Engels

Věcí všech měrou je člověk…
Prótagorás z Abdér

Evropská tradice praví, že matematika povýšila na vědu díky pýthagorejcům, kteří ji vysvobodili ze služebného postavení v obchodě, zeměměřičství, architektuře a řemesle. Přemístili ji do světa nemateriálního, světa viditelného jen naším vnitřním náhledem – teorií. V tomto vyšším světě – „pýthagorejském ráji" – přebývá matematika dodnes. Poté, co se matematika odpoutala od svých kořenů, tj. od světa jevů, přestala důsledně rozlišovat mezi tím, co bylo součástí kvantitativního popisu jevů tohoto světa a tím, co si matematikové vymyslili navíc, aby ony jevy uspořádali do konzistentního systému, matematického obrazu. Matematika se stala hrou v ideálním světě, i když hrou ne zcela samoúčelnou. Z pragmatických důvodů si ponechávala vazbu na svět reálných jevů, na fyziku, na aplikovanou geometrii i na svět financí.

V půli dvacátého století zavedl Hans Reichenbach (1891-1953) velmi užitečnou leč dosud nedoceňovanou klasifikaci prvků epistemologického systému.

Rozdělil je na fenomény a interfenomény. Fenomény jsou vnímatelné, nějakým způsobem se nám „samy" dávají, jeví. Při troše kritičnosti musíme dospět k závěru, že je nevědomě sami tvoříme z našich počitků a zkušeností. (Mohli bychom je považovat i za dílo onoho milostivého Boha, který nám už nadělil jeden veledůležitý jev, totiž celá čísla, čili počty nějakých věcí. Viz slavný výrok Kroneckerův: „Celá čísla stvořil náš milostivý Bůh, ostatní je dílo lidské."). Naproti tomu interfenomény se nám přímo nejeví, nejsou viditelné ani hmatatelné. Jeví se jen nepřímo tak, že si je zavádíme vědomě sami – a činíme tak často se značným intelektuálním úsilím. Jejich účelem je udělat mezi jevy pořádek, stmelit je v řád, v jednotný teoretický systém, model reality. Tento systém je primárně povahy epistemologické, protože nám umožňuje chápat jevy v souvislosti celku. Umožňuje nám tak vytvořit pochopitelný model patřičné části světa. Když ztotožníme tento model s realitou, můžeme náš systém považovat za „ontologický", tvrdící „jak věci jsou". Na rozdíl od fenoménů mají interfenomény zřetelně metafyzický charakter, jsou to v podstatě naše vědomé konstrukty. Přisoudit jim „objektivní" (a intersubjektivní) existenci je pouze na nás, jejich tvůrcích. Je to záležitost našeho rozhodnutí, opírajícího se o naší vědomou víru i o naše nevědomé předsudky. Konečně jako veškerá ontologie.

Hranice mezi světem fenoménů a interfenoménů není pevná, podobně jako hranice mezi vědomím a nevědomím. Jde spíše o jakýsi epistemologický horizont. Existence každého jevu tedy vychází z nějakých apriorních předpokladů, z rámce nějaké, třeba neuvědomované „teorie". V případě společenských věd si býváme obvykle oné subjektivity vytváření fenoménů a interfenoménů vědomi. Uvědomujeme si, že historické či politické události lze vysvětlovat různě, často i protichůdně. V matematice máme ale obvykle zato, že se pohybujeme na jisté půdě. (Konečně Holanďané nazývají matematiku „wiskunde" – uměním jistoty.) Avšak kdybychom uvažovali kritičtěji, tak bychom museli i matematickým entitám přisoudit různý stupeň „jevovosti" či reality. Parafrázujeme-li Platóna (jeho podobenství o úsečce, dialog *Ústava*), tak by se v případě matematiky na nejvyšším stupni reality ocitla ona malá celá čísla (počty nějakých věcí), pak snad geometrické útvary. Avšak složité pojmy včetně nekonečných množin by stály na pólu opačném. Nekonečna, jak je uvažují filosofové a matematikové, nejsou jevy. Nikdo je nikdy neviděl, nenapočítal ani jinak nevnímal, maximálně je s jistým mrazením v zádech tušil. Bezprostředně se nám nejeví, jsou našimi konstrukty, interfenomény. Z tohoto důvodu neuznávali nekonečno pýthagorejci ani Aristotelés.[1] Změnu přineslo až křesťanství. Byl to svatý Augustin,

[1] Pod nekonečnem mám na mysli nekonečno aktuální. Potenciální nekonečno není nijaký objekt, který by šlo zkoumat, je to jen nemožnost dospět k hranici.

který se obrátil na vševědoucího Boha jako na ručitele existence nekonečna. Tento „Bůh matematiků“ však nebyl tak milostiv, aby aktuální nekonečno vyjevil nám smrtelníkům, aby z něj učinil opravdový jev. Situaci moc nezměnilo, ani když se v rámci teorie množin explicitní zmínka o Bohu vypustila a víra v nekonečno byla vtělena do systému axiomů. Bůh jen sestoupil z vědomí do matematického nevědomí. Matematikům tak nezbývá, než spoléhat na nějaký axiomatický systém, soubor „zjevených pravd“, dogmat, na „matematické náboženství“. Navíc, ať už Bůh existuje nebo ne, těžko by bylo odporovat tomu, že nekonečno ve skutečnosti nezaručuje Bůh, ale jen lidská představa o Bohu.

K zlomovému okamžiku došlo, když si Petr Vopěnka uvědomil, že onen přebujelý interfenomenální charakter nekonečných množin odvádí matematiku od reálného, jevového světa a posiluje v ní prvek samoúčelnosti. Uvědomil si, že je to reziduum skrytého náboženského předsudku, od kterého je prospěšné matematiku osvobodit. Vzal Occamovu břitvu a „matematickému světu“ aktuální nekonečna amputoval. Inspirován Husserlovou fenomenologií založil svou Alternativní teorii množin na více fenomenální, více reálné koncepci obzoru a „přirozeného nekonečna“.[2] Prolomil „zakladatelský mýtus“ teorie klasické a ukázal tak na ontologickou relativnost základů matematiky.[3] Tím, že svou teorii nazval „alternativní“, dal najevo, že uznává teorie dvě. (Teorii svou a „klasickou“, která vychází z Cantorových koncepcí.) Tyto teorie se liší především v ontologickém základu, v tom, co definují jako skutečné, existující, jazykem teorie množin „aktualizovatelné“.[4]

Vopěnkův počin představoval skutečný kopernikovský obrat. Kopernik svého času ukázal, že přesun středu vesmíru je nejen možný, ale že představuje z hlediska pochopení „funkce vesmíru“ (tj. z hlediska epistemologie) zásadní zjednodušení. Ukázal však ještě víc, než si byl vědom: Zpochybnil střed jako takový. Vopěnka analogicky ukázal, že matematiku je možné založit i jinak, jednodušeji a srozumitelněji. Avšak stejně jako Kopernik, ani Vopěnka plně nedomyslil důsledky svého činu, tj. relativitu onoho založení. Připustil sice více teorií množin, avšak „jedna z nich musí být ta pravá, pravdivá“ – jak píše hned na několika místech.

[2] Navázal tím na starší koncepty intuicionistů a konstruktivistů (L. Brouwer).

[3] Skoro souběžně vzniklá nestandardní analýza A. Robinsona dochází k podobným výsledkům, její založení však ještě respektuje přístup klasický.

[4] Dle mého náhledu by byl vhodnější název „teorie alternativních množin“. Ony množiny jsou totiž jiné než u teorií „klasických“. V anglickém překladu (Alternative Set Theory, AST) se ovšem rozdíl stírá.

V roce 2011 učinil Petr Vopěnka další krok: Pro údajnou nekonzistenci vyloučil ze hry klasickou (cantorovskou) teorii množin a svou „alternativní" teorii přejmenoval na „Novou".[5] Provokativně pak deklaroval, že veškerá infinitní matematika vycházející z klasické teorie množin je iluzorní. Nutno se prý vrátit o sto let zpátky. Totiž: „*Existence množiny všech přirozených čísel je tím, na čem stojí a s čím padá téměř veškerá infinitní matematika dvacátého století.*" Má Vopěnka v tomto bodě pravdu? Může být matematika „špatně"?

Podle mne by matematiku bylo možno založit na jakémkoli metamatematickém (konzistentním) systému (teorii množin) za podmínky, že její aplikovaná část bude funkční: bude ve shodě s popisovanými (vypočítanými) jevy a bude plodná, tedy schopná nové jevy předpovídat – vypočítat. Ani to, že matematika vychází z chybného základu (jak se pokouší dokázat Vopěnka), ještě neznamená, že je „špatně". Z logického hlediska i z nepravdy plyne pravda. Navíc ona chyba může být (a nejspíš bude) jen okrajová, tedy bez zásadního přepracování opravitelná, podobně jako byly opravitelné nekonzistence v Cantorově „naivní" teorii množin. (Oprava spočívala ve vyloučení některých typů množin pomocí systému axiomů, aby se tak zabránilo sporům typu Cantorova nebo Russellova paradoxu.)

Základním kritériem správnosti matematiky (a jakékoli vědy) totiž nejsou její „ontologické základy", tedy na čem je formálně vybudována. Kritériem správnosti není ani konzistentnost teorie množin, ani otázka „aktualizovatelností" či „neaktualizovatelností" různých nekonečných souborů. Kritériem správnosti je náležité fungování v praxi. Mnohem důležitější než ontologický základ matematiky je její základ epistemologický a pragmatický. Epistemologie je skutečným základem veškeré vědy.[6] Důležité tedy je, jak matematiku (vědu) ze zkušenosti s fyzickým světem vyvozujeme a zpětně jak se tato matematika (věda) v tomto světě osvědčuje. Ontologické základy (včetně axiomů teorie množin) bývají konstruovány až dodatečně, jak z hlediska historie kolektivní (jak lidstvo k matematice dospělo), tak i z hlediska historie individuální (jak se lidé s matematikou seznamují). Sám Vopěnka tyto ontologické základy (teorii množin se svými nekonečny) s oblibou nazýval „okrasnou novobarokní nadstavbou". Dával tím najevo, že bez nich se stavba může klidně obejít a dále i to, že ony „okrasné nadstavby" můžeme měnit podle dobového vkusu.

[5] Co se týká názvů, sám Vopěnka vnesl do terminologie zmatek („Nová, dříve alternativní teorie"). Navíc název by neměl řešit nejen novost teorie, ani to, zda je teorie jedna, dvě (dvě alternativy), nebo více.

[6] Mnozí novodobí myslitelé (např. Quine 1968, Bateson 2006) dokonce ontologii s epistemologií ztotožňují.

Bolzano, theologie a matematika

Zajímavou historickou paralelu tvoří myšlenkový vývoj Bernarda Bolzana. Ovlivněn osvícenstvím a racionalismem pochybuje Bolzano o božském původu biblické zvěsti. Vše může být jen mýtus či blud. Zdálo by se, že pochybuje o samotném základu víry. Své rozpaky však překonává a stává se z něj dokonce hlasatel oné víry. Pochopil, že údajný „božský původ“ není na víře to nejdůležitější. Podstatné je dobro, které víra lidstvu přináší. To, co se zdá být základem skutečným, je jen základem formálním, pravý základ spočívá jinde. Tedy „ne pravda, ale užitečnost“,[7] přesně tak, jak hlásá filosofie pragmatismu. Kdyby se žádné zázraky nestaly, církev by to nesložilo. (Konečně protestanti jsou k existenci zázraků skeptičtí.) Jestliže však náboženství nebude mít lidem co říci, nebude jim poskytovat ono „dobro“, z chrámů se stanou mrtvé památky a mrtvá bude i sama církev. Síla i slabost víry spočívá v životní praxi a praxe, která „lidi povznáší“, která je jim prospěšná, vede je k úspěchu, je i tím pravým základem a argumentem víry a náboženství.

> *… je celkem docela lhostejné, vznikla-li určitá nauka církve teprve později, ba i nepřispěl-li k jejímu vzniku a rozšíření svým dílem nějaký blud.*
>
> *Nebyl jsem dokonce ani úplně přesvědčen o pravdivosti a božském původu náboženství, jehož hlasatelem jsem se měl stát…*
>
> *… v náboženství, zejména v božím zjevení, nejde vůbec o to, jaká je věc sama o sobě, ale naopak jen o to, jaká představa o ní nás nejvíce povznáší.*
>
> B. Bolzano[8]

> *…objektivní pravda, v níž funkce uspokojování lidských přání vůbec nehraje žádnou úlohu, neexistuje. … nezávislá pravda představuje jen mrtvé srdce prázdného stromu.*
>
> W. James, *Pragmatismus*

Podobně i v matematice je podstatné to, co matematika lidstvu přináší. Kdyby matematikové zjistili, že teorie množin je blud, stavba matematiky se nezhroutí. Kdyby ale matematika nebyla úspěšná v praxi, kdyby neposkytovala lidstvu „dobro“, stala by se z ní samoúčelná hra pro pár fandů, něco jako je šach, dáma či go. Síla matematiky je v životní praxi a praxe – tedy aplikovaná matematika – je tím nejskutečnějším základem, nejopravdovějším „bohem matematiky“.

[7] Není jistě bez zajímavosti, že mezi Bolzanovy žáky patřil i Václav Hanka (1791-1861) a Josef Linda (1789?-1834), pravděpodobní autoři *Rukopisu královédvorského* a *Rukopisu zelenohorského*.

[8] Srov. Bolzano (manuskript, 37-38).

Závěr

Teorie množin, ono formální podbudování matematiky, je monumentální stavbou, na které si matematici pocvičili své umění. Úsilí, které této stavbě věnovali a „ráj", který si tím otevřeli, tvoří psychologickou překážku, která brání pochopit, že tato teorie je základem pouze formálním. Podstatnější než toto ontologické založení je založení epistemologické, které vychází ze vzájemné reflexe matematiky a vezdejšího světa. Měli bychom proto být otevřeni k různým možným pojetím ontologických základů matematiky a tedy i k různým teoriím množin. Na to právě významně upozornila alternativní teorie množin Petra Vopěnky. Avšak neměli bychom ztrácet čas a hledat tu jedinou „pravdivou" teorii. Pravd může být totiž více. A neměli bychom úspěšné teorie zavrhovat hned, jakmile najdeme drobný rozpor. Může být totiž snadno opravitelný.

Peter Zamarovský
zamarovs@fel.cvut.cz

Literatura

BATESON, G. (2006): *Mysl a příroda, nezbytná jednota*. Praha: Malvern.

BERKA, K. (1981): *Bernard Bolzano*. Praha: Horizont.

BOLZANO, B. (manuskript): *Vlastní životopis*. Dostupné: http://www.cs.cas.cz/bolzano/node/221.

ENGELS, B. (1952): *Pana Evžena Dühringa převrat vědy*. Praha: Svoboda.

LOUŽIL, J. (1978): *Bernard Bolzano*. Praha: Melantrich.

QUINE, V. O. (1968): Ontological Relativity. *Journal of Philosophy* 45, No. 7, 185-212.

REICHENBACH, H. (1973/1951): *The Rise of Scientific Philosophy*. University of California Press.

VOPĚNKA, P. (2004): *Vyprávění o kráse novobarokní matematiky*. Praha: Práh.

VOPĚNKA, P. (2004): *Horizonty nekonečna*. Praha: Moraviapress.

VOPĚNKA, P. (2011): *Velká iluze matematiky XX. století a nové základy*. Plzeň: Vydavatelství Západočeské univerzity v Plzni, Koniáš.

VOPĚNKA, P. (2014): *Prolegomena*. Praha: Karolinum.