

РОССИЙСКАЯ АКАДЕМИЯ НАУК  
СИБИРСКОЕ ОТДЕЛЕНИЕ

ИНСТИТУТ ФИЛОСОФИИ И ПРАВА

# ФИЛОСОФИЯ НАУКИ

НАУЧНОЕ ИЗДАНИЕ  
ПО ФИЛОСОФИИ, МЕТОДОЛОГИИ И ЛОГИКЕ  
ЕСТЕСТВЕННЫХ НАУК

**№ 5**    ВЫХОДИТ С АВГУСТА 1995 ГОДА    **2025**  
**(109)**

Новосибирск  
2025

Всероссийский научный журнал «Философия науки»

**Учредители:** Сибирское отделение РАН,  
Институт философии и права СО РАН

Издается с 1995 г.



**РЕДАКЦИОННАЯ КОЛЛЕГИЯ:**

А.В. Хлебалин (главный редактор)  
А.Ю. Сторожук (зам. главного редактора)  
А.С. Зайкова (отв. секретарь)

д.филол.н. Д.В. Винник; д.филол.н., проф. В.П. Горан; д.филол.н., проф. В.С. Диев; акад. РАН Ю.Л. Ершов; д.мед.н., д.филол.н., проф. В.А.Карпин; д.филол.н., проф. В.Г. Кузнецов; к.филол.н. А.А. Лазаревич (Республика Беларусь); проф. Р. Муравски (Польша); к.физ.-мат. н., д.филол.н. И.Е. Прись (Республика Беларусь); к.филол.н., к.физ.-мат.н. В.С. Пронских; д.филол.н., проф. А.Л. Симанов; к.филол.н. А.Н. Спасков (Республика Беларусь); д.филол.н., проф. В.А. Суоровец; д.б.н. О.В. Трапезов; проф. А.Е. Ховланд (Норвегия); д.филол.н., проф. В.В. Целищев; д.филол.н., д.физ.-мат.н. О.В. Шарыпов; акад. РАН В.К. Шумный

Свидетельство о регистрации СМИ  
ПИ № 77-12806 от 31.05.02  
Выдано Министерством Российской Федерации  
по делам печати, телерадиовещания и средств массовых коммуникаций

**Адрес редакции:**

630090, Новосибирск, ул. Николаева, 8,  
Институт философии и права СО РАН, к. 403

Тел. (383) 330-52-35  
E-mail: [phscience@mail.ru](mailto:phscience@mail.ru); [stor71@mail.ru](mailto:stor71@mail.ru)

**Web-сайты:** [philosophy.nsc.ru](http://philosophy.nsc.ru)  
<http://www.sibran.ru>  
[elibrary.ru](http://elibrary.ru)

© Сибирское отделение РАН, 2025  
© Институт философии и права  
СО РАН, 2025



*Аналитическая философия:  
траектории истории и векторы развития-2025*

УДК: 161

DOI: 10.15372/PS20250501

END: YNIPDE

**Saeed Salehi**

**A BIT OF A NOTE ON HALTING PROBABILITIES**

We take a look at the definition of a halting probability from logical and algorithmic information theoretical points of view.

*Keywords:* algorithm, halting probabilities, information.

**Саид Салехи**

**A BIT OF A NOTE ON HALTING PROBABILITIES**

Предложено рассмотрение определения вероятности остановки с точки зрения логической и алгоритмической теории информации.

*Ключевые слова:* алгоритм, вероятности остановки, информация.

**Introduction**

A coin is said to be *fair*, when the probability of getting a head (H) by tossing it is equal to the probability of getting a tail (T). Thus, each probability is  $1/2$  since they should add up to 1 by Kolmogorov's axioms of probability measure. So, the probability of getting

a first tail by tossing a fair coin two times (that is, getting either TH or TT) is again  $1/2$ . Let us toss a fair coin one or two times and ask what the probability of getting either one head (that is, H) or two sides that begin with a tail (that is, TH or TT) is. To choose the number of tosses, we provide an urn containing two similar balls, each with a unique label of 1 or 2. We close our eyes and pick up a ball from the urn. If the ball is labeled 1, then we toss the fair coin once; if it is labeled 2, then we toss twice. What is the probability that we get H, TH, or TT?

Whatever that is, let us consider the complement of this event. What is the probability of getting T, HT, or HH by tossing a fair coin once or twice randomly? The probability of H should be equal to the probability of T; call their common value  $q$ . This is the definition of a fair coin. So should be the probabilities of TH, TT, HT, and HH; call their common value  $r$ . So, the probability of  $E = \{H, TH, TT\}$  is  $q + 2r$ , and so is the probability of its complement  $E^c = \{T, HT, HH\}$ . Thus, the probability of  $E$  and that of  $E^c$  should be  $1/2$ . This holds even if the probability of getting the ball with label 1 out of the urn is not equal to the probability of getting the other ball with label 2. That is to say that the probability of  $E$  is  $1/2$ , no matter the values of  $q$  and  $r$ . Notice that the sample space is  $\{H, T, HH, HT, TH, TT\}$ , so we should have  $2q+4r = 1$ , thus the probability of  $E$  is  $p(E) = q+2r = 2q+4r/2 = 1/2$ . All we assumed here was that (1)  $p(\mathbf{H}) = p(\mathbf{T}) = q$ , and (2)  $p(\mathbf{HH}) = p(\mathbf{HT}) = p(\mathbf{TH}) = p(\mathbf{TT}) = r$ .

Therefore, the probability of our event  $E$  is not equal to 1, but its Omega is so:  $\Omega_E = 1/2 + 1/4 + 1/4 = 1$ . In the literature, the Omega of a set  $S$  of finite sequences of 0's and 1's, so-called *binary strings*, is defined as  $\Omega_S = \sum_{\sigma \in S} 2^{-|\sigma|}$ , where  $|\sigma|$  denotes the length of  $\sigma$  ( $\Omega_S$  is also called the *weight* of the set  $S$ ; see, e.g., [2, p. 201]). Let us note that we have identified  $\mathbf{H}$  with 0 and  $\mathbf{T}$  with 1.

### Fair Probability Measures

This number  $\Omega_S$  bears the grand title of probability, for *probably*, two reasons: (I) If  $S$  is prefix-free, i.e., no element of  $S$  is a proper prefix of another element of  $S$  [2, p. 74], then by Kraft's In-

equality [2, p. 125], we have  $0 \leq \Omega_S \leq 1$ . (II) If  $S$  is prefix-free, then  $\Omega_S$  is equal to the probability that the binary expansion (after zero and dot) of a given real number in the unit interval  $[0, 1]$  contains a string from  $S$  as a prefix (see [2, p. 4]).

These are mathematically proven facts, and we do not dispute them (see [3, Prop. 2.9, Cor. 3.7]). But another interpretation persists in the literature to the effect that  $\Omega_S$  is the probability of getting a string in  $S$  if we toss a fair coin a finite, but indefinite, number of times (see [1]). Notice that  $\tilde{E} = \{0, 10, 11\}$  is prefix-free (and we have  $\Omega_{\tilde{E}} = 1$ ). But how can the probability of getting a string from  $\tilde{E}$  by tossing a fair coin (whose one side is 0 and the other is 1) be equal to 1 (=  $\Omega_{\tilde{E}}$ )? Let us compute the probability of getting a string from  $\tilde{E}$  by tossing a fair coin an indefinitely finite number of times. Our sample space is  $\{0, 1\}^+$ , the set of all nonempty binary strings. Let  $\pi_1$  be the common probability of 0 and 1 (recall the definition of a fair coin). Let  $\pi_2$  be the common probabilities of 00, 01, 10, and 11. For each  $n > 0$ , let  $\pi_n$  be the probability of the binary strings with length  $n$ . Since there are  $2^n$  such strings, then we should have:

$$\begin{aligned} & \infty \\ & \sum_{n=1} 2^n \pi_n = 1. \end{aligned}$$

Now, the probability of  $\tilde{E}$  is  $\pi_1 + 2\pi_2$ , which is much less than 1, since:

$$\pi(\tilde{E}) = \pi_1 + 2\pi_2 = 2\pi_1 + 4\pi_2/2 \leq \sum_{n=1}^{\infty} 2^n \pi_n / 2 = 1/2.$$

This misunderstanding happens for a special prefix-free set: the set of halting programs, that is, the binary codes of the input-free programs that eventually stop after running (and do not loop forever). Call that set  $\mathbf{H}$ . The number  $\Omega$  in the literature is  $\Omega_{\mathbf{H}}$ , which equals to  $\sum_p \text{halts } 2^{-|p|}$ , where  $p$  ranges over the binary codes of input-free programs (see [1]). If  $N_l$  is the number of halting input-free programs with length  $l$ , then  $\Omega = \sum_{l=1}^{\infty} N_l 2^{-l}$  (see [4, p. 1]). If  $\pi_l$  is the probability of getting any binary string with length  $l$ , then, by  $\sum_{l=1}^{\infty} 2^l \pi_l = 1$ , it can be shown that the halting probability measured by  $\pi$  ( $\text{HP}\pi = \sum_{l=1}^{\infty} N_l \pi_l$ ) is less than the celebrated Omega number  $\Omega$  (see [3, Thm. 3.4]). This holds more gen-

erally for every prefix-free set. For a set  $S$  of binary strings, let  $N_l(S)$  be the number of elements of  $S$  with length  $l$ . Then the probability of  $S$  is, by definition,  $\pi(S) = \sum_{l=1}^{\infty} N_l(S) \pi_l$ .

**Theorem 1** ( $\pi(S) < \Omega_S$ )

For every fair probability measure  $\pi$  on  $\{0, 1\}^+$ , that is a sequence  $\{\pi_l\}_{l=1}^{\infty}$  of nonnegative real numbers that satisfy  $\sum_{l=1}^{\infty} 2^{-l} \pi_l = 1$ , and every prefix-free set  $S$  that contains at least two strings with different lengths, we have  $\pi(S) < \Omega_S$ .

**Proof:**

On the one hand, for every  $l$  we have  $N_l(S) \pi_l \leq N_l(S) 2^{-l}$ , and on the other hand, there are at least two  $l$ 's with  $N_l(S) > 0$  and there is at most one  $l$  with  $\pi_l = 2^{-l}$ . Thus, for at least one  $l$  we should have  $N_l(S) \pi_l < N_l(S) 2^{-l}$ . Therefore,  $\pi(S) = \sum_{l=1}^{\infty} N_l(S) \pi_l < \sum_{l=1}^{\infty} N_l(S) 2^{-l} = \Omega_S$ .  $\square$

Indeed,  $\Omega_S$  is the limit of an interesting sequence of probabilities. Fix  $N$  to be a positive natural number. Let us toss a fair coin  $N$  times and compute the probability of getting an end-extension of a member of  $S$ . Our event is, in other words, the set of all binary strings with length  $N$  that contain a member of  $S$  as a prefix. This makes sense only if  $S$  is prefix-free.

**Theorem 2 (probability of end-extensions of  $S$  with fixed length)**

For a fixed  $N > 0$  and a prefix-free set  $S \subseteq \{0, 1\}^+$ , the probability that a member of  $S$  appears as a prefix of the binary string after  $N$  times of tossing a fair coin (with one side 0 and the other 1) is the following number:

$$\begin{aligned} & \sum_{\substack{|\sigma| \leq N \\ \sigma \in S}} 2^{-|\sigma|} \end{aligned}$$

**Proof:**

If  $\{\sigma_i\}_{i=1}^m$  is the set of all string in  $S$  with length  $\leq N$ , then our event consists of all  $\sigma_i \tau$  where  $\tau$  is an arbitrary binary string with length  $N - |\sigma_i|$ , and  $i = 1, \dots, m$ . Thus, there are  $\sum_{\sigma_i \in S} 2^{N - |\sigma_i|}$  binary strings with length  $N$  that have some  $\sigma_i$  as a prefix. Therefore, the probability is:

$$1/2^N \sum_{\sigma_i \in S} 2^{N - |\sigma_i|}, \text{ or } \sum_{\sigma_i \in S} 2^{-|\sigma_i|} \quad \square$$

Let us notice that:

$$\begin{aligned} \Omega_S &= \lim_{N \rightarrow \infty} \sum_{\sigma \in S, |\sigma| \leq N} 2^{-|\sigma|} \\ \sigma &\in S \end{aligned}$$

### Unfair Probability Measures

We saw that it is very improbable that an Omega number could be the probability of getting binary strings from given prefix-free sets by tossing a fair coin for finitely many times. We called a probability measure on the binary strings *fair* when strings with equal lengths have equal probabilities. We denoted a fair probability of the strings with length  $l$  by  $\pi_l$  and noted that  $\sum_{l=1}^{\infty} 2^{-l} \pi_l = 1$ . When, then, could we have  $\pi(S) = \Omega_S$ ? Theorem 1 says not if  $S$  contains two strings with different lengths. So, let  $S$  consist of some binary strings with a fixed length  $\ell$ . Then  $\pi(S) = N_\ell(S) \pi_\ell$ , and  $\Omega_S = N_\ell(S) 2^{-\ell}$ . The equality  $\pi(S) = \Omega_S$  holds if and only if  $\pi_\ell = 2^{-\ell}$ ; thus, by  $\sum_{l=1}^{\infty} 2^{-l} \pi_l = 1$ , for every  $l \neq \ell$  we must have  $\pi_l = 0$ . So, for a non-uniform, but still a fair, probability measure  $\mu$ , defined as  $\mu(\sigma) = 2^{-|\sigma|}$  if  $|\sigma| = \ell$ , and  $\mu(\sigma) = 0$  otherwise, we can have  $\mu(S) = \Omega_S$ , if  $S$  consists of some binary strings with a fixed length  $\ell$  only.

The following table summarizes our observations about  $\Omega_S$  and its approximations for a prefix-free set  $S$  (cf. [3, Lem. 3.1 and Cor. 3.7(1)] and Theorem 2):

$$|\sigma| \leq \ell$$

$\sum_{\sigma \in S, |\sigma| \leq \ell} 2^{-|\sigma|}$  = the probability of getting a member of  $S$  after tossing a fair coin for  $\ell$  times:

$$\sigma \in S$$

$$|\sigma| \leq N$$

$\sum_{\sigma \in S, |\sigma| \leq N} 2^{-|\sigma|}$  = the probability of getting an end-extension of a member of  $S$  after tossing a fair coin  $\sigma \in S$  for  $N$  times

$\sum_{\sigma \in S} 2^{-|\sigma|}$  = the probability of getting an end-extension of a member of  $S$  after tossing a (=  $\Omega_S$ ) fair coin for infinitely many times.

Let us, finally, neglect the fairness of our coin. One can define a probability measure on the binary strings  $\{0, 1\}^+$  in such a way that the halting probability measured by it becomes  $\Omega$ , or any number strictly between 0 and 1. We recall that a probability measure maps each string  $\sigma \in \{0, 1\}^+$  to a non-negative real number  $\Pi(\sigma)$ , such that the equation  $\sum_{\sigma \in \{0, 1\}^+} \Pi(\sigma) = 1$  is true. If  $\Pi$  is not fair, then we may have  $\Pi(\tau) \neq \Pi(\tau')$  for some  $\tau, \tau' \in \{0, 1\}^+$  with equal lengths.

Fix a real number  $\alpha \in (0, 1)$ . For a binary string  $\sigma$ , define  $\Pi\alpha(\sigma)$  as  $\alpha / \Omega^{2^{-|\sigma|}}$  if it is the binary code of an input-free halting program, and let  $\Pi\alpha(\sigma)$  be  $1 - \alpha / 2^{|\sigma|} N_{|\sigma|}$  otherwise. That is:

$$\alpha / 2^{|\sigma|} / \Omega, \text{ if } \sigma \in H$$

$$\Pi\alpha(\sigma) = 1 - \alpha / 2^{|\sigma|} N_{|\sigma|}, \text{ if } \sigma \notin H.$$

Then the halting probability by  $\Pi\alpha$ , that is  $\Pi\alpha(H)$ , becomes:

$$\sum_{\sigma \in H} \Pi\alpha(\sigma) = \alpha / \Omega \sum_{\sigma \in H} 1 / 2^{|\sigma|} = \alpha$$

This  $\Pi\alpha$  is a probability measure since:

$$\sum_{\sigma \in \{0, 1\}^+} \Pi\alpha(\sigma) = \alpha + (1 - \alpha) \sum_{\sigma \notin H} 1 / 2^{|\sigma|} / 2^{|\sigma|} (N_{|\sigma|} - N_{|\sigma|}) = \alpha + (1 - \alpha) \sum_{l=1}^{\infty} 1 / 2^l = 1.$$

Therefore, by nonstandard probability measures, every real number strictly between zero and one can be a halting probability.

## References

1. *Calude, C.S., Chaitin, G.J.* (2010) What is ... a Halting Probability? // Notices of the American Mathematical Society. № 57(2), P. 236–237.
2. *Downey, R.G., Hirschfeldt, D.R.* (2010) Algorithmic Randomness and Complexity. Springer.
3. *Jalilvand, M., Nikzad, B., Salehi, S.* (2024) How (not) to Compute the Halting Probability, Qeios. // FTEFK8.2. P. 1–16.
4. *Schmidhuber, C.* (2022) Chaitin's Omega and an Algorithmic Phase Transition, Physica A: Statistical Mechanics and its Applications. № 568 (Article 126458) P. 1–15.

## Литература

1. *Calude, C.S., Chaitin, G.J.* What is ... a Halting Probability? // Notices of the American Mathematical Society. 2010. № 57(2), P. 236–237.
2. *Downey, R.G., Hirschfeldt, D.R.* Algorithmic Randomness and Complexity. Springer. 2010.
3. *Jalilvand, M., Nikzad, B., Salehi, S.* How (not) to Compute the Halting Probability, Qeios. // FTEFK8.2. 2024. P. 1–16.
4. *Schmidhuber, C.* Chaitin's Omega and an Algorithmic Phase Transition, Physica A: Statistical Mechanics and its Applications. № 568 (Article 126458). 2022. P. 1–15.

## Information about author

*Salehi Saeed* – Phd, Plaksha University, IT City Road, Sector 101A, Mohali, Punjab 140306, India.  
saesal@gmail.com

## Информация об авторе

*Саид Салехи* – доктор наук, Университет Плакша, IT City Road, сектор 101А, Мохали, Пенджаб 140306, Индия.  
saesal@gmail.com

Дата поступления 17.06.2025  
Принята к публикации 02.10.2025

УДК: 16 + 1(091) + 17 + 51-7 + 510.82 + 512

DOI: 10.15372/PS20250502

EDN: XEUNHC

**V.O. Lobovikov**

**A PRINCIPLE OF EQUIVALENCE OF MODALITIES  
DE-DICTO AND DE-RE UNDER THE CONDITION  
OF A-PRIORI-NESS OF KNOWLEDGE,  
FORMULATED BY AN ARTIFICIAL LANGUAGE  
OF FORMAL MULTIMODAL AXIOMATIC THEORY  $\Phi^{DR}$**

The *subject-matter* of investigation is the nontrivial modal-logic problem of equivalence of *de-dicto* and *de-re* types of modalities. The *target* – explicating, exact formulating and adding the precisely formulated principle of equivalence of the modality-types as a new proper philosophy axiom scheme to the multimodal axiomatic system of formal philosophy  $\Phi+\exists$ . The *scientific novelty*: for realizing the target, the already published formal axiomatic theory  $\Phi+\exists$  has been transformed into a qualitatively different logically formalized multimodal axiomatic system called  $\Phi^{DR}$  due to manifest taking into an account and systematical dealing with the different modality types called *de-dicto* and *de-re*. For accomplishing the transformation, (1) definitions of the alphabets of artificial object-language and meta-language of  $\Phi+\exists$  have been changed: new symbols  $\Omega^D$  and  $\Omega^R$  standing for perfection-modalities (belonging to the types D and R, respectively) have been included into the alphabet of metalanguage of  $\Phi^{DR}$ ; (2) in  $\Phi^{DR}$ , one new axiom-scheme (containing the new symbols  $\Omega^D$  and  $\Omega^R$ ) has been added to the set of axiom-schemes of  $\Phi+\exists$ ; (3) by means of the artificial language of  $\Phi^{DR}$ , a precise formulation of the principle of equivalence of the modality-types *de-dicto* and *de-re* is given, and the concrete epistemic condition, under which the equivalence principle is valid, is defined. For the first time, a nontrivial *multimodal* interpretation is given for the traditional formal logic square of opposition and hexagon. Also, for the first time, these geometric figures have been utilized for visual modeling a system of *formal-axiological* relations among qualitatively different *de-re*-perfection-modalities as *evaluation-functions*.

*Keywords*: modalities; de-dicto; de-re; principle-of-equivalence-of-modalities-de-dicto-and-de-re; logically-formalized-multimodal-axiomatic-philosophy-system.

В.О. Лобовиков

**ПРИНЦИП ЭКВИВАЛЕНТНОСТИ МОДАЛЬНОСТЕЙ  
ДЕ-ДИКТО И ДЕ-РЕ ПРИ УСЛОВИИ  
АПРИОРНОСТИ ЗНАНИЯ, СФОРМУЛИРОВАННЫЙ  
НА ИСКУССТВЕННОМ ЯЗЫКЕ ФОРМАЛЬНОЙ  
МУЛЬТИМОДАЛЬНОЙ АКСИОМАТИЧЕСКОЙ ТЕОРИИ  $\Phi^{DR}$**

*Предмет* исследования – нетривиальная модально-логическая проблема эквивалентности *de-dicto* и *de-re* типов модальностей. *Цель* – прояснение, точная формулировка, и добавление уточненного принципа эквивалентности этих типов модальностей в качестве новой схемы собственных аксиом в мультимодальную аксиоматическую систему формальной философии  $\Phi+\exists$ . *Научная новизна*: для достижения этой цели, уже опубликованная формальная аксиоматическая теория  $\Phi+\exists$  преобразована в некую качественно отличную от нее логически формализованную мультимодальную аксиоматическую систему, названную  $\Phi^{DR}$ , благодаря принятию во внимание и систематическому использованию в ней упомянутых различных типов модальностей, именуемых *dicto* и *de-re*. Для осуществления такой трансформации, (1) изменены определения алфавитов языка-объекта и метаязыка теории  $\Phi+\exists$ : в алфавит метаязыка теории  $\Phi^{DR}$  добавлены новые символы  $\Omega^D$  и  $\Omega^R$ , обозначающие некие (любые) модальности совершенства упомянутых типов D и R, соответственно; (2) ко множеству аксиомных схем теории  $\Phi+\exists$ , в теорию  $\Phi^{DR}$  добавлена одна новая схема аксиом, содержащая символы  $\Omega^D$  и  $\Omega^R$ ; (3) на искусственном языке  $\Phi^{DR}$  дана точная формулировка принципа эквивалентности типов модальностей *dicto* и *de-re*, а также определено то конкретное эпистемическое условие, при котором этот принцип эквивалентности имеет законную силу. Впервые предложена некая нетривиальная *мультимодальная* интерпретация традиционного формально-логического квадрата и гексагона логической оппозиции абстрактных понятий. Также впервые упомянутые геометрические фигуры использованы в качестве наглядных моделей системы *формально-аксиологических* отношений между качественно различными *de-re*-модальностями-совершенства как *ценностными функциями*.

*Ключевые слова*: модальности; *de-dicto*; *de-re*; принцип эквивалентности модальностей *de-dicto* и *de-re*; логически формализованная мультимодальная аксиоматическая система философии.

## 1. Introduction

Many philosophers have defended an account of *de re* belief about an object in terms of having some *de dicto* belief about that object while also bearing a relation of acquaintance to it, that is, while being epistemically *en rapport* with the object [Wierenga 2021].

The most able and influential enemy of modality (both *de dicto* and *de re*) was W. V. Quine, who vigorously defended both the following theses. First, that modality *de dicto* can be understood only in terms of the concept of analyticity (a problematical concept in his view). Secondly, that modality *de re* cannot be understood in terms of analyticity and therefore cannot be understood at all [Inwagen, Sullivan, and Bernstein 2023].

*De re* is less favorable than *de dicto* [Zhang and Davidson 2020, p. 1].

Formal-logic aspect of alethic modalities was noticed and systematically studied yet in ancient times, at least, since that famous epoch when the syllogism theory had been created and elaborated in details [Аристотель 1978; Лукаевич 1959; Маркин 2018]. However, at first, the nontrivial distinction between *de re* and *de dicto* types of the modalities had been not-well-recognized.

Aristotle considered some “mixed” syllogisms with one apodeictic premise, one assertoric premise and apodeictic conclusion to be valid. His pupils Theophrastus and Eudemus introduced the principle that the conclusion always has the same modal character as the weaker of the premises, thereby they rejected all mixed modal syllogisms. In medieval logic, a distinction was made between *de dicto* and *de re* modalities. It was demonstrated that propositions with *de dicto* and *de re* modalities have different deductive characteristics. Aristotle’s apodeictic syllogistic contains both: reasonings valid only under *de dicto*-interpretation of modalities ... and reasonings valid only under *de re*-interpretation (e.g. *modus Barbara*). When we accept the “principle of the weakest premise”, apodeictic syllogistic can be naturally interpreted as containing *de dicto* modalities. The eminent Polish logician Jan Lukasiewicz suggested that both modal syllogistic versions were incorrect. In his opinion all mixed *modi* formed from the valid categorical syllogisms (e.g. *Barbara* rejected by Aristotle) are also valid. Lukasiewicz justified these *modi*

by means of his positive assertoric syllogistic and four-valued modal logic, which contains some theorems unprovable in normal modal calculi [Маркин 2018, с.114].

In Middle Ages, when the distinction between *de re* and *de dicto* types of the modalities had been well-recognized and established as a norm (custom) of intellectually respectable theorizing in logic, systematical discussing knotty interrelations between the two well-separated types of modalities was progressively developed with respect to main themes of philosophizing of that time [Knuutila 2021; Lagerlund 2022; Parsons 2014].

In contemporary literature accepting and following the medieval distinction of *de dicto* and *de re*, profound investigations of the two types of modalities can be found, for example, in [Prior 1952; Quine 1956; Kneale 1962; Kaplan 1968; Sosa 1970; Chisholm 1976; Burge 1977; Целищев 1978; Lewis 1979; McKay 1991; Reimer 1995; Salmon 1997; Jeshion 2002; Stanley 2002; Taylor 2002; Хлебалин 2003; 2004; Turner 2010; Keshet 2010; Roca-Royes 2011; Ламберов 2011; Keshet, and Schwarz 2014; Куайн 2010; 2016; Маркин 2018; Wierenga 2021; Inwagen, Sullivan, and Bernstein 2023; Nelson 2019; 2023; Schwitzgebel 2024] and in many other noteworthy writings on modal logic, metaphysics, and philosophy of logic. Summarizing results of contemporary investigations of the relationship between modality types *de dicto* and *de re*, one is to affirm that *there is no logic equivalence between the two* [Маркин 2018, с.109]. Consequently, it is quite natural to anticipate not the equivalence (interchangeability) of the two, but its negation.

Hence, with respect to the habitual principle of logic separation of *de dicto* and *de re*, the manifestly challenging (surprising and puzzling) title of present article is psychologically unexpected (at least queer; requiring explanations). Therefore, at the very beginning, the introduction to this article must clarify the ambiguous meaning of the title. One of the possible meanings (namely, the implied at first) is certainly false, but there is another meaning (qualitatively different one) to be defined precisely by means of the hitherto never published multimodal axiomatic theory  $\Phi^{\text{DR}}$ .

First necessary step on the way to adequate understanding the paper title (destroying the *illusion* of its paradoxality, countedintui-

tiveness) is attracting attention to the phrase “under the *condition of a-priori-ness* of knowledge”. *Unconditional* affirming the equivalence does make up a shocking paradox, but in the title under discussion, *the equivalence is affirmed not unconditionally*, but under a *very special* (even *extraordinary*) *condition* concerning knowledge. What does this condition mean? How and by which means can one decide whether a knowledge is *a priori* or not? The answer is given by means of the mentioned multimodal formal axiomatic theory  $\Phi^{\text{DR}}$  to be precisely defined below in this article. Certainly, this is not a direct answer, but it is quite sufficient for the introduction. More comprehensible and developed answer is to be given by the present article *as a whole*.

However, already in the introduction, it is worth providing at least some informal tips, vaguely formulated intuitive guesses concerning the main direction in which the discourse is to move in this paper. Which concrete facts and abstract conceptual materials are to be taken into an account and kept in mind while approximating to the promised precise axiomatic definition of the puzzling principle of equivalence of *de re* and *de dicto* types of modalities?

To answer this quite reasonable question, it is worth taking into an account the curious fact (influential tendency started since the beginning of XX century) in history of philosophy of modal logic, namely; (1) there is a propensity to locate *de dicto* modalities into the realm of logic-and-sciences proper: (2) there is a propensity to locate *de re* modalities into the realm of proper metaphysics (proper philosophical ontology), axiology, and even theology [Inwagen, Sullivan, and Bernstein 2023].

Modality *d dicto* is the modality of propositions (‘dictum’ means proposition, or close enough). If modality were coextensive with modality *de dicto*, it would be at least a defensible position that the topic of modality belongs to logic rather than to metaphysics. (Indeed, the study of modal logics goes back to Aristotle’s Prior Analytics.)

But many philosophers also think there is a second kind of modality, modality *de re* – *modality of things*. (*The* modality of substances, certainly, and perhaps of things in other ontological categories.) The status of modality *de re* is undeniably a metaphysical topic, and we assign it to the “new” metaphysics because, although one can ask

modal questions about things that do not change – God, for example, or universals – a large proportion of the work that has been done in this area concerns the modal features of changing things [Inwagen, Sullivan, and Bernstein 2023].

Thus, there is a curious (even somewhat elegant) *analogy* between the ordered couples <metaphysics-proper; positive-sciences-proper> and <*de-re*-modalities; *de-dicto*-modalities>. Although, in relation to the logic theory of quantification, Quine has a negative attitude to both *de re* and *de dicto* modalities, while reading his writings on the theme, it is easy to feel that his attitude to *de re* is much worse than to *de dicto*: his rejecting *de re* is analogous to his rejecting metaphysics: both are meaningless, from his point of view [Куайн 2010, с. 200-228; 2016, с. 150-177]. In this concrete relation, Quine has been not alone: many positivist-minded logicians have believed that *de re* interpretation of modalities is a manifestation of *essentialism* which (according to them) is a kind of metaphysics in the positivist meaning of the word [Хлебалин 2003, с. 36, 39, 41]. It is highly likely that, at least at the level of subconsciousness, just the anti-metaphysics is underlying Quine's resolutely negative attitude to combining the standard logic of quantification with the logic of modalities. If one (for example [Williamson 2013]) is right, when the one considers *modal logic as metaphysics*, then it is highly likely (and quite natural to anticipate) that radical positivists are to reject if not all then at least some results of combining quantifiers with modalities, especially with *de re* ones [Куайн 2010, с. 213, 215-217, 219-223; 2016, с. 168, 172, 173]. It is not merely curious but also quite relevant to note here that the fact-fixing statement “*de re* is less favorable” has been established (demonstrated statistically) by an *empirical* investigation of belief reports [Zhang and Davidson 2020].

The tendency completely to eliminate (exclude) *de re* type of modalities from meaningful discourse is *analogous* to the positivist tendency to eliminate metaphysics, axiology and theology from scientific philosophy. The above-indicated facts *induce* a psychologically unexpected *hypothesis* that, probably, there is a hidden (not quite evident) but essential link between mutually corresponding formal-logical systems of *de re* modalities and *formal-axiological* systems of

*evaluation-functions* making up the subject-matter of *metaphysics-proper* understood (interpreted) as nothing but abstract *formal axiology* (such an unhabitual hypothesis is formulated and investigated, for instance, in [Лобовиков 2007; Lobovikov 2022]). Obviously, from the traditional point of view, this is a somewhat strange guess. However, in my opinion, this hypothesis is worth taking into an account and investigating carefully by the hypothetic-deductive method (I intend to do this in the present article).

As the long and hard discussion about the significant difference between modality-types *de dicto* and *de re* still goes on, I have decided to submit this paper formulating quite a new idea concerning the problem. In the present paper, the new nontrivial idea (unexpected and somewhat puzzling one) concerning modalities *de dicto* and *de re* is represented by a new scheme of proper axioms of formal philosophy (namely, AX-7) making up the novel formal multimodal axiomatic theory  $\Phi^{\text{DR}}$ .

## 2. Methods

In the present article, the hypothetic-deductive method and discrete mathematical modeling are exploited. Intentional constructing artificial languages and formal axiomatic theories possessing special properties is undertaken systematically. The artificial languages are used for precise defining formal axiomatic theories modeling multimodal proper philosophy systems. For obtaining qualitatively new (hitherto never published) nontrivial scientific results, some significant mutations (creating new species) are made in some already existing (previously created) artificial languages and formal axiomatic theories. For example, alphabets of object-languages are enriched, definitions of terms and formulae are changed, some (at least one) *new axiom-scheme is added* to the set of proper-axiom schemes. Owing to such significant mutations creating new species of formal axiomatic theories, qualitatively new possibilities of formalization and interpretation appear and new application domains are discovered.

## New Scientific Results

### 1.1. A New Logically Formalized Multimodal Axiomatic Philosophy System $\Phi^{DR}$

The hitherto never considered multimodal formal axiomatic theory  $\Phi^{DR}$  is a result of significant mutation of the formal axiomatic *epistemology-and-axiology-and-ontology* theory  $\Phi+\exists$  [Lobovikov, 2024].

For exact defining the formal theory  $\Phi^{DR}$ , it is necessary to provide exact definitions of the notions: “*alphabet of object-language of  $\Phi^{DR}$* ”; “*term of  $\Phi^{DR}$* ”; “*formula of  $\Phi^{DR}$* ”; “*axiom of  $\Phi^{DR}$* ”. Exact definitions of these notions of  $\Phi^{DR}$  are strikingly *analogous* (look *similar*) to the exact definitions of the corresponding notions of  $\Phi+\exists$ , which definitions are already utilized and published in [Lobovikov, 2024]. Therefore, one can feel *déjà vu* (strong sensation of *analogousness*) and fall into the *identity illusion*. Nevertheless, for the sake of rigor, in the present article, it is necessary manifestly to give precise definitions of notions: “*alphabet of object-language of  $\Phi^{DR}$* ”; “*term of  $\Phi^{DR}$* ”; “*formula of  $\Phi^{DR}$* ”; “*axiom of  $\Phi^{DR}$* ”, in spite of the mentioned *analogousness*, as the words “*analogousness*” and “*identity*” are not synonyms. Meanings of these words are not logically equivalent. Notions of  $\Phi^{DR}$  *differ substantially* from the corresponding *analogous (similar)* notions of  $\Phi+\exists$ . Therefore, it is reasonable to start with precise formulating definitions of those notions which are indispensable for adequate comprehending this paper, in spite of the false sensation that they are just repetitions of the already published statements. Let us begin with precise defining the notion “object-language-alphabet of  $\Phi^{DR}$ ”.

According to the below-provided definition (containing 11 items), the alphabet of object-language of  $\Phi^{DR}$  includes all the signs which belong to the alphabets of object-languages of  $\Sigma$ ,  $\Sigma+C$ ,  $\Sigma+2C$ , and  $\Phi+\exists$ . But, the conversion of this judgement is not true, because, in  $\Phi^{DR}$ , some new signs are added to the alphabets of object-languages of  $\Sigma$ ,  $\Sigma+C$ ,  $\Sigma+2C$ , and  $\Phi+\exists$ . The result of such a significant change (complementation) is the below-placed exact definition of the alphabet of object-language of  $\Phi^{DR}$ .

1) The lowercase Latin letters p, q, d (and the same letters possessing lower number indexes) are elements of the object-language-alphabet of  $\Phi^{DR}$ . Such and only such lowercase Latin letters are called

“*dictum variables*”. In the object-language-alphabet of  $\Phi+\exists$ , *not all lowercase Latin letters are called dictum variables* because, according to the provided definition, such lowercase Latin letters which are elements of the set  $\{g, b, e, n, x, y, z, s, h, a, t, f\}$  do not belong to the set of *dictum variables* of object-language of  $\Phi^{\text{DR}}$ .

2) The lowercase Latin letters  $s, h, a$  (and the same letters having lower literal indexes:  $s_m, h_s, a_t$ ) are elements of the object-language-alphabet of  $\Phi^{\text{DR}}$ . The mentioned lowercase Latin letters (and only they) are called “*dictum constants*”.

3) The habitual proper (pure) logic symbols  $\supset, \neg, \leftrightarrow, \&, \vee$  called, respectively, “classical (or ‘material’) implication”, “classical negation”, “classical equivalence”, “classical conjunction”, “classical non-excluding disjunction” are elements of the object-language-alphabet of  $\Phi^{\text{DR}}$ .

4) Elements of the set  $\{I^{\text{D}}, \square^{\text{D}}, \mathcal{K}^{\text{D}}, \text{III}^{\text{D}}, H^{\text{D}}, K^{\text{D}}, A^{\text{D}}, E^{\text{D}}, S^{\text{D}}, Z^{\text{D}}, G^{\text{D}}, W^{\text{D}}, O^{\text{D}}, B^{\text{D}}, C^{\text{D}}, Y^{\text{D}}, T^{\text{D}}, F^{\text{D}}, P^{\text{D}}, \odot^{\text{D}}, D^{\text{D}}, U^{\text{D}}, J^{\text{D}}\}$ , containing the signs  $\square^{\text{D}}, \odot^{\text{D}}$ , the capital Cyrillic letters  $\mathcal{K}$  and  $\text{III}$  having upper literal index D, and some (but not all) capital Latin letters possessing no number indexes but having upper literal index D (indicating to “de dicto”), belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . The mentioned elements of the alphabet are called “de-dicto-modality-symbols” in  $\Phi^{\text{DR}}$ . The de-dicto-modality symbol  $\mathcal{K}^{\text{D}}$  belongs *only* to the object-language-alphabets of  $\Phi+\exists$  and  $\Phi^{\text{DR}}$ . The de-dicto-modality symbols  $I^{\text{D}}, \text{III}^{\text{D}}, H^{\text{D}}, Z^{\text{D}}$ , and  $\odot^{\text{D}}$  belong *exclusively* to the object-language-alphabet of  $\Phi^{\text{DR}}$ .

5) Elements of the set  $\{I^{\text{R}}, \square^{\text{R}}, \mathcal{K}^{\text{R}}, \text{III}^{\text{R}}, H^{\text{R}}, K^{\text{R}}, A^{\text{R}}, E^{\text{R}}, S^{\text{R}}, Z^{\text{R}}, G^{\text{R}}, W^{\text{R}}, O^{\text{R}}, B^{\text{R}}, C^{\text{R}}, Y^{\text{R}}, T^{\text{R}}, F^{\text{R}}, P^{\text{R}}, \odot^{\text{R}}, D^{\text{R}}, U^{\text{R}}, J^{\text{R}}\}$ , containing the signs  $\square^{\text{R}}, \odot^{\text{R}}$ , the capital Cyrillic letters  $\mathcal{K}$  and  $\text{III}$  having upper literal index R, and some (but not all) capital Latin letters possessing no number indexes but having upper literal index R (indicating to “de re”), belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . The mentioned elements of the alphabet are called “de-re-modality-symbols” in  $\Phi^{\text{DR}}$ . The de-re-modality symbols  $\mathcal{K}^{\text{R}}, I^{\text{R}}, \text{III}^{\text{R}}, H^{\text{R}}, Z^{\text{R}}$ , and  $\odot^{\text{R}}$  belong *exclusively* to the object-language-alphabet of  $\Phi^{\text{DR}}$ .

6) The lowercase Latin letters  $z, y, x$  (and the same letters having lower number indexes) belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . They are called “*axiological variables*” in  $\Phi^{\text{DR}}$ .

7) Also the lowercase Latin letters “b” and “g” belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . They are called “*axiological constants*”.

8) The capital Latin letters  $C^1, K^1, K^2, E^1, E^2, C^2, A_k^n, C_j^n, B_i^n, D_m^n, \dots$ , having number-indexes, are elements of the alphabet of object-language of  $\Phi^{\text{DR}}$ . These capital Latin letters are called “*axiological-value-functional-symbols*”. The upper number-index  $n$  (in these symbols) means that the axiological-value-functional-symbol (indexed by number  $n$ ) is  $n$ -placed one. Generally speaking, axiological-value-functional-symbols may possess no lower number-index. However, if axiological-value-functional-symbols possess lower number-indexes, then, if these indexes are different, then the indexed axiological-value-functional-symbols are different signs.

9) The “round brackets”, namely, “(” and “)” belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . These auxiliary symbols are utilized in the present article as usually in symbolic logic, namely, as pure technical signs.

10) The “square brackets”, namely, “[” and “]” belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ . But, it is very important to emphasize here that the “square brackets” (in contrast to the “round ones”,) are exploited in  $\Phi^{\text{DR}}$  not as the auxiliary (pure technical) signs, but as *ontologically meaningful* symbols. Being a psychological surprise, such unhabitual (odd) using the “square brackets” makes some psychological difficulty, because, for common people using natural language in everyday life, square brackets and round ones seem identical and, as a rule (statistic norm) the two kinds of brackets are exploited (in natural language of common people) as synonyms. However, in contrast to natural language of common people, in the artificial object-language of  $\Phi^{\text{DR}}$ , the two species of brackets possess *qualitatively different* meanings. They play *substantially different* roles in  $\Phi^{\text{DR}}$ : round brackets are used as purely technical (auxiliary) signs; on the contrary, usage of square brackets has an *ontological* meaning (very important one). The *ontological* meaning of the inhabitual (even odd) usage of square brackets in the artificial object-language of  $\Phi^{\text{DR}}$ , is exactly defined below in that special part of the present article which (part) is devoted to exact defining *semantics* of object-language of  $\Phi^{\text{DR}}$ . However, even

at the purely syntactic level of the object-language of  $\Phi^{\text{DR}}$ , square-bracketing plays a significant role in giving precise definition of/for the concept “formula of  $\Phi^{\text{DR}}$ ”. (This precise definition is to be given below in the present section of the given article.) Moreover, the unusual utilization of square brackets *plays an important role* also in the precise formulations of some schemes of axioms of  $\Phi^{\text{DR}}$ .” (These formulations of axiom-schemes are placed below also in the present section of the given article.)

11) An inhabitual (odd) complex symbol “=+=” (composed of the habitual symbols) belongs to the object-language-alphabet of  $\Phi^{\text{DR}}$ . The symbol “=+=” is called (a sign of) “*formal-axiological equivalence*”. The strange compound sign “=+=” *plays a very important role* in the precise definition of the notion “formula of  $\Phi^{\text{DR}}$ ”, and also in the precise formulations of some schemes of axioms of  $\Phi^{\text{DR}}$ .

12) A symbol belongs to the object-language-alphabet of  $\Phi^{\text{DR}}$ , when and only when the symbol is an element of that alphabet due to the above-formulated points 1) – 11) of this definition.

Any finite queue (succession) of signs is called “*an expression of  $\Phi^{\text{DR}}$* ”, when and only when that queue contains such and only such signs which belong to the object-language-alphabet of  $\Phi^{\text{DR}}$ .

A precise definition of the concept “*term of  $\Phi^{\text{DR}}$* ”, is the following.

1) the *axiological variables* (mentioned in the above-given definition of object-language-alphabet of  $\Phi^{\text{DR}}$ ) are terms of  $\Phi^{\text{DR}}$ .

2) the *axiological constants* (mentioned in the above-given definition of object-language-alphabet of  $\Phi^{\text{DR}}$ ) are terms of  $\Phi^{\text{DR}}$ .

3) the *dictum variables* (mentioned in the above-provided definition of object-language-alphabet of  $\Phi^{\text{DR}}$ ) are terms of  $\Phi^{\text{DR}}$ .

4) the *dictum constants* (mentioned in the above-provided definition of object-language-alphabet of  $\Phi^{\text{DR}}$ ) are terms of  $\Phi^{\text{DR}}$ .

5) If  $\Phi_k^n$  is an *n-placed value-functional symbol* (belonging to the above-provided definition of object-language-alphabet of  $\Phi^{\text{DR}}$ ), and  $t_i, \dots, t_n$  are *terms of  $\Phi^{\text{DR}}$* , then any expression having the form  $\Phi_k^n t_i, \dots, t_n$  is a term of  $\Phi^{\text{DR}}$ . (Here, it is worth keeping in mind that symbols  $t_i, \dots, t_n$  belong not to the object-language but to the meta-language as they stand for *any* terms of  $\Phi^{\text{DR}}$ ; the similar note is worth keeping in mind with respect to the symbol  $\Phi_k^n$  also belonging to the meta-language of  $\Phi^{\text{DR}}$ .)

6) When  $t_n$  is a term of  $\Phi^{\text{DR}}$ , and the sign  $\Psi$  (belonging to the meta-language of  $\Phi^{\text{DR}}$ ) stands for a (any) modal-symbol belonging to the set  $\{\text{I}^{\text{R}}, \square^{\text{R}}, \mathfrak{K}^{\text{R}}, \text{H}^{\text{R}}, \text{K}^{\text{R}}, \text{A}^{\text{R}}, \text{E}^{\text{R}}, \text{S}^{\text{R}}, \text{Z}^{\text{R}}, \text{T}^{\text{R}}, \text{F}^{\text{R}}, \text{P}^{\text{R}}, \odot^{\text{R}}, \text{D}^{\text{R}}, \text{C}^{\text{R}}, \text{Y}^{\text{R}}, \text{G}^{\text{R}}, \text{W}^{\text{R}}, \text{O}^{\text{R}}, \text{B}^{\text{R}}, \text{U}^{\text{R}}, \text{J}^{\text{R}}\}$ , then all such object-language expressions of  $\Phi^{\text{DR}}$ , which (expressions) have the form  $\Psi t_n$ , are terms of  $\Phi^{\text{DR}}$ . Here, it is worth taking into an account, that, strictly speaking, the expression possessing form  $\Psi t_n$  is not a term of  $\Phi^{\text{DR}}$ , but a scheme of terms of  $\Phi^{\text{DR}}$ . (Thus, de-re modalities applied to terms make up new terms.)

7) Any expression of the object-language of  $\Phi^{\text{DR}}$ , is a term of  $\Phi^{\text{DR}}$ , then and only then, when this is so due to the points 1) – 6) of the given definition.

Now, the *pure-syntax* aspect of the abstract notion “term of  $\Phi^{\text{DR}}$ ”, is perfectly defined. Hence, now it is appropriate to move on to exact defining the *pure-syntax* aspect of the abstract notion “formula of  $\Phi^{\text{DR}}$ ”. To accomplish this move correctly, let us make an agreement that in the given article, lowercase Greek letters  $\omega$ ,  $\beta$ , and  $\alpha$  (belonging to meta-language of  $\Phi^{\text{DR}}$ ) stand for *any* formulae of  $\Phi^{\text{DR}}$ . Taking this agreement into an account, it is possible to formulate the below-placed precise definition of the notion “formula of  $\Phi^{\text{DR}}$ ”.

1) All such lowercase Latin letters which are named “*dictum variables*”, and also all such lowercase Latin letters which are named “*dictum constants*”, are elements of the set of formulae of  $\Phi^{\text{DR}}$ .

2) When  $\alpha$  and  $\beta$  are formulae of  $\Phi^{\text{DR}}$ , then all such object-language expressions of  $\Phi^{\text{DR}}$ , which (expressions) have the forms  $(\alpha \& \beta)$ ,  $(\alpha \vee \beta)$ ,  $(\alpha \supset \beta)$ ,  $(\alpha \leftrightarrow \beta)$ ,  $\neg\alpha$ , are elements of the set of formulae of  $\Phi^{\text{DR}}$  as well.

3) When  $t_i$  and  $t_k$  are terms of  $\Phi^{\text{DR}}$ , then  $(t_i = t_k)$  is a formula of  $\Phi^{\text{DR}}$ .

4) When  $t_i$  is a term of  $\Phi^{\text{DR}}$ , then  $[t_i]$  is a formula of  $\Phi^{\text{DR}}$ .

5) When  $\alpha$  is a formula of  $\Phi^{\text{DR}}$ , and the sign  $\Psi$  (belonging to the meta-language of  $\Phi^{\text{DR}}$ ) stands for a (any) modal-symbol belonging to the set  $\{\text{I}^{\text{D}}, \square^{\text{D}}, \mathfrak{K}^{\text{D}}, \text{III}^{\text{D}}, \text{H}^{\text{D}}, \text{K}^{\text{D}}, \text{A}^{\text{D}}, \text{E}^{\text{D}}, \text{S}^{\text{D}}, \text{Z}^{\text{D}}, \text{T}^{\text{D}}, \text{F}^{\text{D}}, \text{P}^{\text{D}}, \odot^{\text{D}}, \text{D}^{\text{D}}, \text{C}^{\text{D}}, \text{Y}^{\text{D}}, \text{G}^{\text{D}}, \text{W}^{\text{D}}, \text{O}^{\text{D}}, \text{B}^{\text{D}}, \text{U}^{\text{D}}, \text{J}^{\text{D}}\}$ , then all such object-language expressions of  $\Phi^{\text{DR}}$ , which (expressions) have the form  $\Psi\alpha$ , are formulae of  $\Phi^{\text{DR}}$ . Here, it is worth taking into an account, that, strictly speaking, the expression possessing form  $\Psi\alpha$  is not a formula of  $\Phi^{\text{DR}}$ , but

a scheme of formulae of  $\Phi^{\text{DR}}$ . Thus, *de-dicto* modalities applied to formulae make up new formulae. This is a significant *syntactic* difference between the modalities *de dicto* and *de re* (in  $\Phi^{\text{DR}}$ ). The first ones are applied to *formulae* while the second ones are applied to *terms*.

6) Finite successions (queues) of symbols belonging to the object-language-alphabet of  $\Phi^{\text{DR}}$  are formulae of  $\Phi^{\text{DR}}$ , if and only if this is so due to the points 1) – 5) of this definition.

This concrete part of the article is intentionally reduced exclusively to *syntactic* meanings of object-language-expressions of  $\Phi^{\text{DR}}$ . Therefore, the set of *de-dicto*-modal-symbols  $\{I^{\text{D}}, \square^{\text{D}}, \mathcal{K}^{\text{D}}, \text{III}^{\text{D}}, H^{\text{D}}, K^{\text{D}}, A^{\text{D}}, E^{\text{D}}, S^{\text{D}}, Z^{\text{D}}, T^{\text{D}}, F^{\text{D}}, P^{\text{D}}, \odot^{\text{D}}, D^{\text{D}}, C^{\text{D}}, Y^{\text{D}}, G^{\text{D}}, W^{\text{D}}, O^{\text{D}}, B^{\text{D}}, U^{\text{D}}, J^{\text{D}}\}$  and the set of corresponding *de-re*-modal-symbols are considered here as nothing but sets of extremely short *names*. The sign  $I^{\text{D}}$  (Latin letter «I» possessing the upper literal index D informing that the modality belongs to the type called “*de dicto*”) is a name of/for the *newly* introduced philosophical-ontology modality “it is *immutable* (*constant*, does not undergo a change) that ...”. The sign  $\square^{\text{D}}$  is a short name for the well-known *de dicto* modality “it is *necessary* that ...”. The sign  $\mathcal{K}^{\text{D}}$  (Cyrillic letter  $\mathcal{K}$  possessing the upper literal index D) is a short name of/for the *de dicto* philosophical-ontology modality “what is indicated-and-described by the *dictum* ..., *exists*”. (This unhabitual *de-dicto*-modality has been introduced originally in [Lobovikov 2024].) The sign  $\text{III}^{\text{D}}$  (Cyrillic letter III possessing the upper literal index D) is a short name of/for the *de dicto* philosophical modality “it is *essential* (or it is an *essence*) that ...”. (Hence, at least some of the nontrivial problems of the essentialism can be modeled in  $\Phi^{\text{DR}}$ .)

Here it is worth highlighting that the symbols  $\mathcal{K}^{\text{D}}$  and  $I^{\text{D}}$  (denoting the ontological modalities *de dicto*) do not belong to the object-language-alphabets of  $\Sigma$ ,  $\Sigma+C$ ,  $\Sigma+2C$ ,  $\Phi$ . The formal theory  $\Phi^{\text{DR}}$  is an outcome of adding the symbols  $I^{\text{D}}$ ,  $\text{III}^{\text{D}}$ ,  $H^{\text{D}}$ ,  $Z^{\text{D}}$ , and  $\odot^{\text{D}}$  to the set of *de-dicto*-modality-symbols belonging to the object-language-alphabet of  $\Phi+\exists$ . Also, it is worth highlighting here that, in contrast with the object-language-alphabets of  $\Sigma$ ,  $\Sigma+C$ ,  $\Sigma+2C$ ,  $\Phi$ , and  $\Phi+\exists$ , the modality signs in  $\Phi^{\text{DR}}$  possess upper literal indexes D or R informing that the indexed symbols stand, respectively, for *de-dicto* or *de-re* modality types.

The symbols  $H^D$ ,  $K^D$ ,  $E^D$ ,  $A^D$ ,  $S^D$ ,  $Z^D$ ,  $T^D$ ,  $F^D$ ,  $P^D$ ,  $\odot^D$ ,  $D^D$ ,  $Y^D$ ,  $C^D$ , respectively, denote the following modal expressions: “it is *imperseparable* (*sensually unverifiable, inaccessible to the senses*) that...”; “agent *Knows* that...”; “agent *Empirically (a-posteriori) knows* that...”; “agent *A-priori knows* that...”; “in some fixed time-and-space, under some special conditions, an agent has a *Sensation* (either by means of some instruments, or immediately), that...”; “in some fixed time-and-space, under some special conditions, (either immediately, or by means of some instruments) an agent *changes* what is described by the dictum that ...”. “it is *True* that...”; “agent has *Faith* that... (or person believes that...)”; “in a consistent theory, it is *Provable* that...”; “it is *Constructive* that...”; “it is *Decidable* that...”; “it is *Complete* that ...”; “it is *Consistent* that...”.

The symbols  $G^D$ ,  $W^D$ ,  $O^D$ ,  $B^D$ ,  $U^D$ ,  $J^D$ , respectively, denote the following modal expressions: “it is *Good* (perfect in moral sense) that...”; “it is *Wicked* (bad in moral sense) that...”; “it is *Obligatory* (or it is a duty) that ...”; “it is *Beautiful* (perfect in aesthetic sense) that ...”; “it is *Useful* (beneficial) that ...”; “it is a *Joy* (pleasure, gladness, happiness) that ...”. In this concrete part of the article, purely *syntactic* meanings of the de-dicto-modality-symbols are precisely defined by the below-placed schemes of proper-formal-philosophy-axioms of the multimodal system  $\Phi^{DR}$ . Obviously, the axiomatic definition is not manifest (direct) one, but, nevertheless, it is quite exact and sufficient for precise philosophizing.

In the logically formalized multimodal philosophy system  $\Phi^{DR}$ , the proper axioms of universal philosophical *epistemology*, formal philosophical *ontology*, and abstract formal *axiology* are combined with pure logic axioms which (axioms of logic-proper) are obviously *analogous* to the axioms of classical propositional logic (let it be called PL). Thus, pure-logic axioms and logical derivation rules of  $\Phi^{DR}$  are substantially *analogous* to the ones of classical PL calculus. Certainly, the analogousness is not an absolute identity as sets of formulae of PL and  $\Phi^{DR}$  are different.

The well-known proper logic meanings of the classical propositional-logic-connectives represented by habitual signs  $\neg$ ,  $\&$ ,  $\leftrightarrow$ ,  $\vee$ , and  $\supset$ , which (signs) are exploited in the below-provided axiom-schemes of  $\Phi^{DR}$ , are precisely defined, for instance, in [Kleene 1957;

1973]. Schemes of axioms and logic-derivation rules of classical PL calculus are also well-defined in [Kleene 1957; 1973]. Therefore, let us abstain from repetition of the well-known, and concentrate on actually original aspects of  $\Phi^{\text{DR}}$ .

While comparing formal theories  $\Phi+\exists$  and  $\Phi^{\text{DR}}$ , one can come to the *truthlike* conclusion that corresponding definitions of their basic notions are completely identical. But, strictly speaking, being actually *analogous* they are not identical. *Analogousness* is *conjunction* of *similarity* and *difference*. (This very important statement is well-demonstrated visually by the relevant hexagon and square of opposition modeling graphically the logical relations among notions “difference”, “identity”, “similarity”, and “analogousness” [Beziau 2015].) Being *similar*,  $\Phi+\exists$  and  $\Phi^{\text{DR}}$  have *different* alphabets of their object-languages and, hence, *different* sets of artificial-language expressions, *different* sets of terms (and, hence, *different* sets of formulae), *different* sets of proper axioms, and, hence, *different* sets of theorems. Thus, even at the syntaxis level,  $\Phi+\exists$  and  $\Phi^{\text{DR}}$  are qualitatively *different* axiomatic theories.

In the present section of the paper, the *syntactic* meanings of the modality signs and of the other symbols belonging to the alphabet of object-language of  $\Phi^{\text{DR}}$  are defined precisely by the below-located list (AX-1 – AX-11) of schemes of proper-philosophy (ontology, epistemology, axiology) axioms of  $\Phi^{\text{DR}}$ . (Obviously, the *axiomatic* definition of proper-philosophy notions, namely, universal epistemological, ontological and axiological concepts is *not a manifest* definition. Notwithstanding, it is *perfectly exact* one.) When  $\alpha, \beta, \omega$  are formulae of  $\Phi^{\text{DR}}$ , then all such (and only such) expressions of the object-language of  $\Phi^{\text{DR}}$ , which have the following logic forms, are *proper-axioms* of  $\Phi^{\text{DR}}$ .

Axiom scheme AX-1:  $A^{\text{D}}\alpha \supset (\Omega^{\text{D}}\beta \supset \beta)$ .

Axiom scheme AX-2:  $A^{\text{D}}\alpha \supset (\Omega^{\text{D}}(\alpha \supset \beta) \supset (\Omega^{\text{D}}\alpha \supset \Omega^{\text{D}}\beta))$ .

Axiom scheme AX-3:

$A^{\text{D}}\alpha \leftrightarrow (K^{\text{D}}\alpha \ \& \ (\neg\Diamond^{\text{D}}\neg\alpha \ \& \ \neg\Diamond^{\text{D}}S\alpha \ \& \ \Box^{\text{D}}(\beta \leftrightarrow \Omega^{\text{D}}\beta)))$ .

Axiom scheme AX-4:

$E^{\text{D}}\alpha \leftrightarrow (K^{\text{D}}\alpha \ \& \ (\Diamond^{\text{D}}\neg\alpha \ \vee \ \Diamond^{\text{D}}S\alpha \ \vee \ \neg\Box^{\text{D}}(\beta \leftrightarrow \Omega^{\text{D}}\beta)))$ .

Axiom scheme AX-5:  $\Omega^{\text{D}}\alpha \supset \Diamond^{\text{D}}\alpha$ .

Axiom scheme AX-6:  $(\Box^D\beta \ \& \ \Box^D\Omega^D\beta) \supset \beta$ .

Axiom scheme AX-7:  $A^D\alpha \supset (\Omega^D[t_k] \leftrightarrow [\Omega^R t_k])$ .

Axiom scheme AX-8:  $(t_i =+ = t_k) \leftrightarrow (G^D[t_i] \leftrightarrow G^D[t_k])$ .

Axiom scheme AX-9:  $(t_i =+ = g) \supset \Box^D G^D[t_i]$ .

Axiom scheme AX-10:  $(t_i =+ = b) \supset \Box^D W^D[t_i]$ .

Axiom scheme AX-11:  $(G^D\alpha \leftrightarrow \neg W^D\alpha)$ .

Definition scheme DF-1: when  $\alpha$  is a formula of  $\Phi^{DR}$ , then  $\diamond^D\alpha$  is a name of/for  $\neg\Box^D\neg\alpha$ .

In AX-1, AX-2, AX-3, AX-4, AX-5, AX-6, and AX-7, the sign  $\Omega^D$  (belonging to the meta-language of  $\Phi^{DR}$ ) denotes a (any) “*perfection modality de dicto*” exclusively. Not all the above-mentioned *de dicto* modalities are called “*perfection ones*”. (It is possible to call them merely “*perfections de dicto*”.) The set  $\Delta^D$  of symbols denoting *perfection-modalities de dicto* is the following  $\{\mathcal{K}^D, \odot^D, K^D, D^D, F^D, C^D, Y^D, P^D, J^D, T^D, B^D, G^D, U^D, O^D, \text{III}^D, H^D, I^D, \Box^D\}$ . Obviously,  $\Delta^D$  is only a subset of the above-mentioned set of all symbols denoting *de dicto* modalities under consideration in this article. For instance,  $W^D, S^D, Z^D$ , and  $\diamond^D$  are signs of such *de dicto* modalities, which are not *de dicto* perfections.

In AX-7, the sign  $\Omega^R$  (belonging to the meta-language of  $\Phi^{DR}$ ) denotes a (any) “*perfection modality de re*” exclusively. The set  $\Delta^R$  of symbols denoting *perfection-modalities de re* is the following  $\{\mathcal{K}^R, \odot^R, K^R, D^R, F^R, C^R, Y^R, P^R, J^R, T^R, B^R, G^R, U^R, O^R, \text{III}^R, H^R, I^R, \Box^R\}$ . It is easy to see that there is a *one-to-one correspondence* between elements of the sets  $\Delta^D$  and  $\Delta^R$  of perfection modalities.

Evidently, all the above-said is *semantically meaningless*; in any relation to such things, which are external to  $\Phi^{DR}$ , all the above-given exact definitions make no sense as they are *purely syntactic* ones. However, this is not an outcome of absentmindedness (or a delinquency committed by negligence) but such a scientific abstraction which is deliberately accepted on purpose. The intentionally admitted scientific abstraction is perfectly resonable, when and only when, its adequateness domain is well-defined. Therefore, to make the present article text as a whole perfectly meaningful, now I have to leave the above-defined *syntaxis* of the artificial language of  $\Phi^{DR}$  for a hitherto completely unknown (absolutely undefined) *semantics* of that artificial language.

## 1.2. Moving from Syntaxis of the New Formal Multimodal Axiomatic Theory $\Phi^{DR}$ to its Semantics

The above-placed section 3.1 of this paper, presents the *purely syntactic* definition of  $\Phi^{DR}$ , which has been intentionally deprived of its relevant philosophical contents (due to the accepted scientific abstraction). The formal-philosophy axiomatic system  $\Phi^{DR}$  is a *multimodal* one, but hitherto concrete contents of the modalities under consideration have been revealed not sufficiently; the theory  $\Phi^{DR}$  has been considered as an exactly *formal* theory. Now, in the given part of the paper, namely, in the section 3.2, I am to relax the formality of  $\Phi^{DR}$  by shifting immediately to *concrete philosophical contents* of the above-mentioned modalities studied in  $\Phi^{DR}$ . In the present paper it is implied that semantic meanings of the habitual artificial-language signs of classical symbolic logic are already introduced and well-defined owing to relevant handbooks. As the quite clear semantic meanings of the relevant proper-logic symbols are well-known, it is redundant to define them here. But, such unusual (inhabitual, perhaps, very odd) signs of the artificial language of  $\Phi^{DR}$ , which are exploited systematically in the proper-philosophy-axioms (ones of epistemology, ontology, etc.), require special introduction and precise definition of their semantic meanings.

Meanings of the lowercase Latin letters q, p, d (and of the same letters possessing lower number indexes) named “*dictum-variables*” are analogous to the meanings of the habitual “*propositional variables*”. But there is a substantial difference: in  $\Phi^{DR}$ , values of “*dictum variables*” belong to the set of *dictums*, to which (set) not only all true or false *sentences* (statements) but also all true or false theories (logically organized systems of propositions) belong. Thus, generally speaking, the dictum-variables range over the set of either true or false dictums. If an interpretation of  $\Phi^{DR}$  is provided (well-defined), then a *dictum-constant* means (in the given interpretation) quite a definite (perfectly fixed) element from the set of dictums, namely, either a concrete true or false sentence (statement) or a concrete true or false theory.

According to the habitual (statistical) linguistic norm (custom rule), from the Latin language, “*dictum*” is to be translated as “an expression of (a thought ...) in words”, for example, as “a proposition (sentence) q”, or “an affirmation of (...)”. But, there is a heuristically important possibility deliberately to shift from “affirmation of (a proposition ...)”

to a *significantly more general* “affirmation of (a proposition ..., or a theory ...)” as along with uttering separate statements, one can affirm also a theory (logically organized system of statements). As the indicated innovative generalization is accepted, in this paper, it is presumed (as a hypothesis worthy of investigation) that a theory is also a dictum. Hence, attaching *de-dicto*-modalities Ж (existence), С (consistency), Т (truth), Y (completeness), and D (decidability) to theories is vindicated in  $\Phi^{\text{DR}}$ . Relevant information of modalities *de-dicto* and *de-re*, along with interesting philosophical discussing their interconnections, can be found, for example, in [Кнеале, 1962; Целищев, 1978].

Defining semantic meanings of formal-language-expressions is defining an *interpretation-function*. For defining the interpretation-function, it is necessary to define precisely: (1) such a set which is called “realm (or domain) of interpretation” (hereafter the letter M denotes the set which is the domain of interpretation); (2) an “assessor (valuator)” V. If a standard interpretation of  $\Phi^{\text{DR}}$  is fixed, then, by definition, M is such a set, each element of which possesses: (1) one and only one proper *axiological value* belonging to the set {good, bad}, and (2) one and only one proper *ontological value* belonging to the set {exists, not-exists}.

The *axiological variables* ( $x, y, z, x_k, y_m, z_i, \dots$ ) take their values from the domain of interpretation (M).

The *axiological constants* “b” and “g” denote abstract *axiological values* “bad” and “good”, respectively.

Valuating an element belonging to M by a definite (fixed) assessor V is nothing but ascribing an *axiological value* [either g (good), or b (bad)] to the element. The assessor V may be either an individual or a collective (it does not matter). Certainly, any change of V can result in a change of some (relative) evaluations, nevertheless, such mutations cannot change the set of absolutely immutable formal-axiological laws of two-valued algebra of metaphysics (as formal axiology), which absolutely universal laws are not relative but absolute moral evaluations. The laws in question are *constant valuation-functions* possessing the axiological value “good” under any combination of the values of their arguments. Although V is such a *variable*, values of which belong to the set of all possible assessors (interpreters), any well-defined interpretation of  $\Phi^{\text{DR}}$  necessarily implies that

the value of assessor-*variable*  $V$  is fixed. Any change of value of  $V$  means a change of interpretation.

In the given paper, “e” and “n” denote “... exists” and “... not-exists”, respectively. The lowercase Latin letters “e” and “n” are called “*ontological constants*”. In any standard interpretation of  $\Phi^{\text{DR}}$ , by definition, one and only one element of the four-element-ed set  $\{\{g, n\}, \{g, e\}, \{b, n\}, \{b, e\}\}$  corresponds to every element of the domain of interpretation (the above-introduced set  $M$ ). That is why  $\Phi^{\text{DR}}$  may be considered as a formal semantic representation (discrete mathematical model) of an important truth existing in “Meinong’s jungles” [Meinong, 1960], [Russell, 1905; 1941; 1992], [Jacquette, 1996; 1997; 2015], [Marek, 2022], [Parsons, 1980; 1982], [Perszyk, 1993], [Berto, 2012], [Berto, and Priest, 2014], [Routley, 1980], [Zalta, 1983]. The lowercase Latin letters “e” and “n” belong to the alphabet of meta-language. They do not belong to the alphabet of object-language of  $\Phi^{\text{DR}}$ , according to the above-given definition (of the alphabet). Notwithstanding, “e” and “n” are represented in the object-language of  $\Phi^{\text{DR}}$  *indirectly* by means of *square-bracketing*: the *ontological* statement-form “ $t_i$  exists” is represented by formula  $[t_i]$ ; the *ontological* statement-form “ $t_i$  does not exist” is represented by formula  $\neg[t_i]$ . This implies that square-bracketing is a very important aspect of precise defining the proper philosophical *semantics* of  $\Phi^{\text{DR}}$ .

From the viewpoint of *formal modal logic* of values, preferences, and assessments, the axiom schemes AX-11 and AX-12 are quite clear and obvious. In contrast to them, the almost unknown (extraordinary, aunnhabitual) nontrivial axiom-schemes AX-8, AX-9, AX-10 represent not the symbolic *formal logic of (evaluative modal) judgements* but a symbolic *formal axiology* – general theory of abstract-*value-forms of any* (either existing or not-existing) *things*. (This is an option of systematical rationalizing Meinongianism, or a special kind of its being quite consistent.) The concept “symbolic *formal-logic*” is not identical (logically) to the concept “symbolic *formal-axiology*”, hence, “formal-logic inconsistency” and “formal-axiological inconsistency” are not synonyms.

In my opinion, the above-said (of “ $t_i$  exists” and “ $t_i$  does not exist”) is an adequate formal semantic representation of the extraordinary doctrine uniting existent and nonexistent objects in one system of philosophical ontology [Fine, 1984; 1985], [Hintikka,

1984], [Jacquette, 1996; 1997; 2015], [Marek, 2022], [Parsons, 1980; 1982], [Perszyk, 1993], [Priest, 2005], [Reicher, 2022], [Smith, 1985], [Zalta, 1983].

$N$ -placed terms of  $\Phi^{\text{DR}}$  are interpreted as  $N$ -placed *evaluation-functions* defined on the set  $M$ . To exclude possible misunderstandings produced by concept confusions, it is worth noting here that the word combination “*evaluation function*” is used in this paper in three *qualitatively different* meanings which are precisely defined as follows.

(I) any function is called an “*evaluation-function*”, if and only if the function takes its values from the set of *axiological values*  $\{g$  (good),  $b$  (bad) $\}$ .

(II) a function is called “*mixed evaluation-function*”, if and only if the function takes values from the set of *axiological values*  $\{g$  (good),  $b$  (bad) $\}$ , and its variables take their values from the set  $M$ . Hereafter the symbol  $\surd$  shall stand for the *mixed evaluation-functions* (mappings in the proper mathematical meaning of the word “mapping”):  $M^N \rightarrow \{g$  (good),  $b$  (bad) $\}$ .

(III) a function is called “*pure evaluation-function*”, if and only if the function takes its values from the set of *axiological values*  $\{g$  (good),  $b$  (bad) $\}$ , and variables of the function take their values from the same set.

Speaking of exactly the *pure evaluation-functions* I mean the following mappings:

$\{g, b\} \rightarrow \{g, b\}$ , if one speaks of the evaluation-functions determined by *one* evaluation-argument;

$\{g, b\} \times \{g, b\} \rightarrow \{g, b\}$ , where “ $\times$ ” stands for the Cartesian product of sets, if one speaks of the evaluation-functions determined by *two* evaluation-arguments;

$\{g, b\}^N \rightarrow \{g, b\}$ , if one speaks of the evaluation-functions determined by  $N$  evaluation-arguments, where  $N$  is a finite positive integer.

It is well-known that in the two-valued algebraic system, there are exactly 4 *one-placed pure evaluation-functions* and exactly 16 *two-placed pure evaluation-functions*. But it *seems* that this well-known information contradicts logically to the below-located tables 1 and 2 possessing more than 4 columns. The illusion of paradox is a result of the fact that the number of columns is equal to the number of *mixed*

evaluation-functions under definition by the tables. Identifying and substituting for each other the *pure* and the *mixed* (evaluation functions) is the cause of/for the paradox illusion to be excluded. The *logically different* notions of the *pure* and the *mixed* (evaluation functions) *must be distinguished* systematically.

The concept “one-placed evaluation-function” is instantiated below by Table 1. (Here it is relevant to note that hereafter the upper literal index R standing immediately after a capital letter implies that this letter stands for such an evaluation-function which is a modality *de re*.) Precise dtfning the interpretation of *de re* modalities as *one-placed evaluation-functions* is accomplished by the below-located tables 1 and 2.

Table 1

Interpretation of *de re* modalities as unary evaluation-functions

$\sqrt{x}$	$\sqrt{\mathcal{K}^R x}$	$\sqrt{N^R x}$	$\sqrt{\square^R x}$	$\sqrt{\diamond^R x}$	$\sqrt{V^R x}$	$\sqrt{\odot^R x}$	$\sqrt{K^R x}$	$\sqrt{P^R x}$	$\sqrt{F^R x}$	$\sqrt{D^R x}$	$\sqrt{G^R x}$	$\sqrt{B^R x}$
g	g	b	g	g	b	g	g	g	g	g	g	g
b	b	g	b	b	g	b	b	b	b	b	b	b

By Table 1, the one-placed term  $\mathcal{K}^R x$  is interpreted as *one-placed evaluation-function* “*existence, being of* (what, whom) *x*”. The one-placed term  $N^R x$  is interpreted as one-placed evaluation-function “*nonexistence, nonbeing of* (what, whom) *x*”. The sign  $\square^R x$  stands in standard interpretation for “*necessity of* (what, whom) *x*”. The sign  $\diamond^R x$  stands in standard interpretation for “*possibility of* (what, whom) *x*”. The symbol  $V^R x$  – “*volatility, mutability, changeability of* (what, whom) *x*”, or “*possibility of a change of* (what, whom) *x* by an agent, under some conditions”.  $\odot^R x$  – “*constructivity or constructiveness of x, i.e. possibility of construction of x*”.  $K^R x$  – “*knowledge of* (what, whom) *x, i.e. acquaintance with* (what, whom) *x*”.  $P^R x$  – “*proof (evidence) of* (what) *x*”, or “*basis for accepting x (means of/for forcing to accept) x*”.  $F^R x$  – “*faith, belief, trust in* (what, whom) *x*”.  $D^R x$  – “*algorithm of/for construction of* (what, whom) *x*”.  $G^R x$  – “*positive moral value (goodness) of* (what, whom) *x*”.  $B^R x$  – “*positive aesthetic value (beauty) of* (what, whom) *x*”.

Table 2

Interpretation of *de re* modalities as one-placed valuation-functions

$\sqrt{x}$	$\sqrt{WRx}$	$\sqrt{URx}$	$\sqrt{JRx}$	$\sqrt{TRx}$	$\sqrt{CRx}$	$\sqrt{YRx}$	$\sqrt{IRx}$	$\sqrt{SRx}$	$\sqrt{HRx}$	$\sqrt{ZRx}$	$\sqrt{IIIx}$	$\sqrt{RX}$
g	b	g	g	g	g	g	g	b	g	b	g	b
b	g	b	b	b	b	b	b	g	b	g	b	g

By Table 2, the one-placed term  $W^R x$  is interpreted as *unary evaluation-function* “negative moral value (badness, evilness, wickedness) of (what, whom)  $x$ ”.  $U^R x$  – “positive utilitarian value (utility) of (what, whom)  $x$ ”.  $J^R x$  – “positive hedonistic value (pleasantness) of (what, whom)  $x$ ”, in other words, “happiness with, or joy from (what, whom)  $x$ ”.  $T^R x$  – “truth of  $x$ ”, or “truthfulness (adequateness, rightness, authenticity) of  $x$ ”.  $C^R x$  – “consistency of  $x$ ”.  $Y^R x$  – “completeness of  $x$ ”.  $I^R x$  – “immutability, unchangeability, constantness of (what, whom)  $x$ ”, or “impossibility of a (any) change of (what, whom)  $x$  by an agent”.  $S^R x$  – “an agent’s sensation of (what, whom)  $x$ ”, or “feeling (what, whom)  $x$  by an agent (under some conditions, either immediately, or by means of some instruments)”.  $H^R x$  – “imperceptibility of (what, whom)  $x$ ”, or “ $x$ ’s being imperceptible (sensually unverifiable, inaccessible to the senses) for/by any agents under any conditions”. (In other words,  $H^R x$  is interpreted as “impossibility of  $S^R x$ ”.)  $Z^R x$  – “a change of (what, whom)  $x$  by an agent (either immediately, or by means of some instruments), in some fixed time-and-space, under some special conditions”. (Hence,  $I^R x$  is interpreted as “impossibility of  $Z^R x$ ”.)  $III^R x$  – “essence of (what, whom)  $x$ ”.  $R^R x$  – “appearance, phenomenon of (what, whom)  $x$ ”.

Although today logicians are quite reconciled with (and even used to) talks of truth as modality *de dicto* [Wright 1996; Wolenski 2016], they are not used to talks of truth as modality *de re*. (Therefore,  $T^R x$  is an odd thing for them.) However, sometimes not only “grass-roots” but also celebrated thinkers use the word “truth” in such meaning which is *de-re*-modality one. Look, for example, into the following wonderful citation from G. W. Leibniz’ “On the Ultimate Origination of Things”, interesting comments to which (citation) can be found in [Geretto 2016, p. 617-632].

“These considerations must be held to be not only pleasing and consoling, but most true. I think that in the universe nothing is truer than happiness, nor is anything happier or sweeter than truth” [Leibniz 1989, p. 154]. This wonderful citation demonstrates that not only the above-defined (by Table 2) *de re* modaliteies  $T^R$  and  $J^R$  understood as evaluation-functions, but also their compositions  $T^R J^R$  and  $J^R T^R$  have been discussed by Leibniz. Thus, not only talks of truth as a modality *de dicto* but also talks of truth as a modality *de re* are vindicated.

The concept “binary evaluation-function” is exemplified below by Table 3. In the given article, the upper index 2 standing immediately after a capital letter implies that the letter denotes a binary evaluation-function.

Table 3

Definition of interpretation of the binary evaluation-functions

$\surd x$	$\surd y$	$\surd N^2xy$	$\surd K^2xy$	$\surd S^2xy$	$\surd E^2xy$	$\surd V^2xy$	$\surd A^2xy$	$\surd X^2xy$	$\surd T^2xy$	$\surd Z^2xy$	$\surd P^2xy$	$\surd C^2xy$
g	g	b	g	b	g	b	g	b	b	b	g	g
g	b	b	b	g	b	g	g	b	b	b	g	b
b	g	b	b	g	b	g	g	g	g	g	b	g
b	b	g	b	g	g	b	b	b	b	b	g	g

In Table 3, the two-placed term  $N^2xy$  is interpreted as binary evaluation-function “realizing neither  $x$  nor  $y$ ”; the two-placed term  $K^2xy$  is interpreted as binary evaluation-function “uniting  $x$  and  $y$ ”, or “joint being of  $x$  with  $y$ ”, or “being of both  $x$  and  $y$  together”. The two-placed term  $S^2xy$  is interpreted as binary evaluation-function “split, separation, disengagement, discommunication between  $x$  and  $y$ ”.  $E^2xy$  – “equivalence (identity of values) of  $x$  and  $y$ ”.  $V^2xy$  – “choosing and realizing such and only such an element of the set  $\{x, y\}$ , which is: (1) the best one, when both  $x$  and  $y$  are good; (2) the least bad one, when both  $x$  and  $y$  are bad; (3) the good one, when  $x$  and  $y$  have opposite values. (Thus,  $V^2xy$  means an excluding choice and realization of only the optimal between  $x$  and  $y$ .) The term  $A^2xy$  is interpreted as the binary evaluation-function “realizing a not-excluding-choice result, i.e. (1) realizing  $K^2xy$  when both  $x$  and  $y$  are

good, and (2) realizing  $V^2xy$  otherwise".  $X^2xy$  – evaluation-function “joint being of  $y$  with nonbeing of  $x$ ”, or “ $y$ 's being without  $x$ ”.  $T^2xy$  – “termination of  $x$  by  $y$ ”.  $Z^2xy$  – “ $y$ 's contradiction to (with)  $x$ ”.  $P^2xy$  – “preservation, conservation, protection of  $x$  by  $y$ ”.  $C^2xy$  – evaluation-function “ $y$ 's existence, presence in  $x$ ”. Additional instantiations of the concept “binary evaluation-function” can be found in [Лобовиков 2007; Lobovikov, 2022].

For excluding possible psychological-linguistic illusions and conceptual confusions, now it is quite opportune and perfectly relevant to highlight that in a standard interpretation of  $\Phi^{\text{DR}}$ , the symbols  $\mathcal{X}^R x$ ,  $N^R x$ ,  $B^R x$ ,  $U^R x$ ,  $J^R x$ ,  $N^2xy$ ,  $K^2xy$ ,  $S^2xy$ ,  $A^2xy$  denote not some predicates but some  $n$ -placed *evaluation-functions*. When a standard interpretation of  $\Phi^{\text{DR}}$  is provided, then such object-language expressions of  $\Phi^{\text{DR}}$ , which (expressions) have the forms  $(t_i = + = t_k)$ ,  $(t_i = + = g)$ ,  $(t_i = + = b)$ , represent some *predicates* in  $\Phi^{\text{DR}}$ .

According to semantics of  $\Phi^{\text{DR}}$ , when  $t_i$  is a term of  $\Phi^{\text{DR}}$ , then, if a formula of  $\Phi^{\text{DR}}$ , possessing the form  $[t_i]$ , is interpreted, then the mentioned formula represents (in the given interpretation) an *either false or true proposition* having the form “ $t_i$  exists”. Thus, according to the definition, in any standard interpretation, any formula  $[t_i]$  is true then and only then, when  $t_i$  possesses the *ontological value* “ $e$  (exists)” in the given interpretation. Also, in any standard interpretation of  $\Phi^{\text{DR}}$ , any formula  $[t_i]$  is false, then and only then, when  $t_i$  possesses the ontological value “ $n$  (not-exists)” in the given interpretation.

By definition of semantics of  $\Phi^{\text{DR}}$ , in a standard interpretation of  $\Phi^{\text{DR}}$ , the formula scheme  $(t_i = + = t_k)$  is a proposition possessing the form “ $t_i$  is *formally-axiologically equivalent* to  $t_k$ ”; this proposition is true if and only if (in that interpretation) the terms  $t_i$  and  $t_k$  obtain identical *axiological values* (from the set {good, bad}) under any possible combination of *axiological values* of their *axiological variables*. In other words, by definition,  $(t_i = + = t_k)$  iff  $(\surd t_i = \surd t_k)$ .

By definition of semantics of  $\Phi^{\text{DR}}$ , in a standard interpretation of  $\Phi^{\text{DR}}$ , the formula scheme  $(t_i = + = b)$  is a proposition having the form “ $t_i$  is a *formal-axiological contradiction*” (or “ $t_i$  is *formally-axiologically, or invariantly, or absolutely bad*”); this proposition is true if and only if (in that interpretation) the term  $t_i$  acquires axiological value “bad” under any possible combination of axiological

values of the axiological variables. In other words, by definition,  $(t_i =+= b)$  iff  $(\sqrt{t_i} = b)$ .

By definition of semantics of  $\Phi^{DR}$ , in a standard interpretation of  $\Phi^{DR}$ , the formula scheme  $(t_i =+= g)$  is a proposition having the form “ $t_i$  is a *formal-axiological law*” (or “ $t_i$  is *formally-axiologically, or invariantly, or absolutely good*”); this proposition is true if and only if (in the interpretation) the term  $t_i$  acquires *axiological value* “good” under any possible combination of axiological values of the axiological variables. In other words, by definition,  $(t_i =+= g)$  iff  $(\sqrt{t_i} = g)$ .

Concerning the above-given definition of semantic meaning of  $(t_i =+= t_k)$  in  $\Phi^{DR}$ , it is indispensable to highlight the important linguistic fact of homonymy of the words “is”, “means”, “implies”, “entails”, “equivalence” in natural language. On the one hand, in natural language, these words may have the well-known formal logic meanings. On the other hand, in natural language, the same words may stand for the above-defined *formal-axiological-equivalence* relation “=+=”. This ambiguity of natural language is to be taken into an account; the different meanings of the homonyms are to be separated systematically; otherwise the homonymy can head to logic-linguistic illusions of paradoxes.

Owing to the above-provided definition of proper philosophical semantics (formal-ontological-and-formal-axiological one) of/for the formal theory  $\Phi^{DR}$ , readers can easily notice that the above-defined algebraic system of metaphysics as formal axiology plays the role of such a *theory of relativity* of assessments, in which (relativity theory), the *laws (formal-axiological ones)* of that algebraic system are nothing but constantly-good evaluation-functions. Speaking differently, the absolutely universal and immutable laws of valuation-relativity are *invariants* in relation to all possible changes of valuator (interpreter) V.

Systematically using the above-provided definitions, one can generate (and examine) the following philosophically interesting formal inference in  $\Phi^{DR}$  from the assumption  $A^D\alpha$ .

- 1)  $A^D\alpha \supset (\Omega^D[t_k] \leftrightarrow [\Omega^R t_k])$ : axiom scheme AX-7.
- 2)  $A^D\alpha$ : assumption.
- 3)  $(\Omega^D[t_k] \leftrightarrow [\Omega^R t_k])$ : from 1 and 2 by modus ponens.

4)  $\Box^D[x] \leftrightarrow [\Box^R x]$ : from 3 by substituting:  $\Box$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 4) is translated as follows: it is necessary that  $x$  exists, if and only if there is a necessity of  $x$ .

5)  $K^D[x] \leftrightarrow [K^R x]$ : from 3 by substituting:  $K$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 5) is translated as follows: an agent knows that  $x$  exists, iff the agent's knowledge of  $x$  exists.

6)  $F^D[x] \leftrightarrow [F^R x]$ : from 3 by substituting:  $F$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 6) is translated as follows: an agent believes that  $x$  exists, iff there is the agent's belief in  $x$ .

7)  $D^D[x] \leftrightarrow [D^R x]$ : from 3 by substituting:  $F$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 7) is translated as follows: there is an algorithm of/for deciding that ( $x$  exists), iff there is an algorithm of/for implementation (construction) of  $x$ .

8)  $G^D[x] \leftrightarrow [G^R x]$ : from 3 by substituting:  $G$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 8) is translated as follows: it is (morally) good that ( $x$  exists), iff a moral goodness (positive moral value) of  $x$  exists.

9)  $B^D[x] \leftrightarrow [B^R x]$ : from 3 by substituting:  $B$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 9) is translated as follows: it is beautiful that ( $x$  exists), iff a beauty (positive aesthetic value) of  $x$  exists.

10)  $U^D[x] \leftrightarrow [U^R x]$ : from 3 by substituting:  $U$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 10) is translated as follows: it is useful that ( $x$  exists), iff a use (utility) of/from  $x$  exists.

11)  $J^D[x] \leftrightarrow [J^R x]$ : from 3 by substituting:  $J$  for  $\Omega$ ;  $x$  for  $t_k$ . Into the natural language, 11) is translated as follows: it is a pleasure, joy, happiness that ( $x$  exists), iff there is a happiness, joy, pleasure of/from  $x$ .

In my opinion, from the viewpoint of formal philosophy proper, the above-considered (hitherto never published elsewhere) option (direction) of investigating interconnection of *de dicto* and *de re* types of modalities is worth developing further.

### 1.3. Conditions of Truth of/for Formulae of $\Phi DR$ in Standard Model of/for It

Concerning the important problem of proper-logic *sematic* status (*truth*-related one) of moral-legal *evaluations*, i.e. statements of moral-legal *value* (which *evaluations* are also *either good or bad moral-legal*

acts according to the bivalent *ethics and theory of the natural law*), it is worth taking into an account formal-axiological equivalences ( $\mathcal{K}^R x = + = x$ ) and ( $C^R x = + = x$ ), in which meanings of terms  $x$ ,  $\mathcal{K}^R x$  and  $C^R x$  belong to the above-defined set M (named “interpretation domain”). When an interpretation of  $\Phi^{DR}$  is defined, then every element of M obtains one and only one of the values belonging to the four-elemented set of two-elemented sets  $\{\{g, n\}, \{g, e\}, \{b, e\}, \{b, n\}\}$ . Probably, it is worth recalling here that: “g” stands for “good”; “b” stands for “bad”; “e” stands for “exists”; “n” stands for “not exists”.

From the above-defined viewpoint, it is evident that formal-axiological equivalences ( $\mathcal{K}^R x = + = x$ ) and ( $C^R x = + = x$ ) are essentially *similar* (significantly *analogous*) to the well-known formal-logical *tantamount-ness* ((It is true that  $p$ )  $\equiv$   $p$ ), considered systematically in the logical semantics by Tarsky. However, in this relation, it is very important to take into an account that, strictly speaking, *analogy (similarity) is not an identity* relation. Generally speaking, *similarity (analogousness) is not transitive*, hence, it is not a relation of equivalence. According to the logical positivism, strictly speaking, *statements of values* (aesthetic, religious, moral-legal) *are neither true nor false*, hence, in relation to *proper logic* semantics (i.e. *logic as a theory of truth*), they are semantically *meaningless*.

However,  $\Phi^{DR}$  is not a *pure* formal logic theory; it is not *only* logic system.  $\Phi^{DR}$  is a result of *application* of formal logic to what is not formal logic but possesses *own proper* axioms, for instance, epistemic, axiological ones.

Moreover,  $\Phi^{DR}$  is a *multimodal* theory (truth and falsity are considered in it along with many other substantially different kinds of modalities, for instance, along with the *evaluative* modalities (moral goodness, wickedness, etc.).

To avoid possible misinterpretations of the below-placed formulations, here it is worth emphasizing that “t” possessing no indexes stands in the present paper for “true”, and “t” possessing a lower index (literal one) stands for a *term*.

In logic semantics of the above-defined artificial language of  $\Phi^{DR}$ , truth of moral-legal judgement  $G^D[t_i]$  is quite exactly defined by the following *formal-logical equivalence*: ((assessment  $G^D[t_i]$  has proper logic value “t”)  $\equiv$  (term  $t_i$  has axiological value “g”). The arti-

ficial symbol “ $\equiv$ ” stands here for the habitual relation of *formal-logical tantamount-ness* (coincidence of proper logic values). This *tantamount-ness* is a quite precise fundamental *definition* of proper logic semantic meaning of  $G^D[t_i]$  in the language of  $\Phi^{DR}$ .

Keeping in mind all the above-given definitions, conditions of truth of/for formulae of  $\Phi^{DR}$  (in its standard model) are defined in the present paper as follows:

I. ((formula  $[t_i]$  possesses proper *logic* value “t”)  $\equiv$  (term  $t_i$  possesses *ontological* value “e”).

II. ((formula  $G^D[t_i]$  possesses proper *logic* value “t”)  $\equiv$  (term  $t_i$  possesses *axiological* value “g”).

III. ((formula  $W^D[t_i]$  possesses proper *logic* value “t”)  $\equiv$   $\bar{\bar{}}$  (term  $t_i$  possesses *axiological* value “g”).

IV. For any formula  $\alpha$  of  $\Phi^{DR}$ , formula  $K^D\alpha$  possesses proper *logic* value “t”, if and only if, either it is true that  $E^D\alpha$ , or it is true that  $A^D\alpha$ . Here, “either ..., or ...” stands for the strict (excluding) disjunction. Consequently, it is quite relevant to use the below-placed hexagon containing the logic opposition square for visualizing (modelling geometrically) the system of proper formal-logic semantic (truth-related) logical interrelations among *de dicto epistemic* modalities  $K^D$ ,  $A^D$ ,  $E^D$ , while precisely defining conditions of/for truth of formulae of  $\Phi^{DR}$  in standard model this theory.

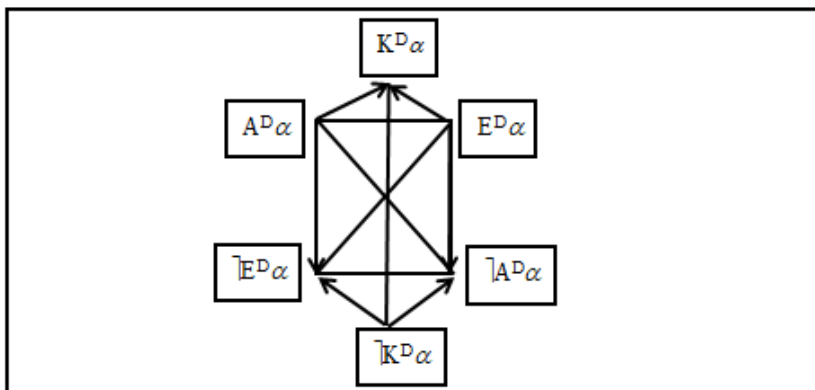


Fig. 1. The logical opposition-square-and-hexagon of epistemic modalities de dicto

According to the so-called “*traditional formal logic*”, in the above visual image, the upper horizontal line represents geometrically the well-known *contrariety* relation; in its turn, the bottom horizontal line in Fig. 1 represents geometrically the well-known *sub-contrariety* relation; the three lines crossing the square geometrically model the well-known *contradictoriness* relations. The arrows model visually the well-known relations of *subordination* (formal-logic *consequence*). With respect to the epistemic modalities *de dicto*, all the truth-related formal-logic rules of the well-known traditional formal-logic opposition square are valid. The heuristically and pedagogically fruitful idea of geometric modelling proper formal-logic aspect of conceptual systems by the square of conceptual opposition has been known and exploited systematically since ancient times to our ones. Nowadays, the traditional square of opposition is essentially generalized by the hexagon and many *qualitatively novel* interpretations of/for the old idea (of visual geometric modeling proper logic relations and rules) have been invented and exploited.

According to the given paper, in standard model of  $\Phi^{\text{DR}}$ , conditions of truth of/for formulae containing the formula  $E\alpha$ , are exactly defined by the following statements:

V. For every formula  $\alpha$ , when it is true that  $E^D\alpha$ , then it is also true that  $K^D\alpha$ .

VI. For every formula  $\alpha$ , when it is true that  $\diamond^D S\alpha$ , then it is also true that  $E^D\alpha$ . This represents (in truth-related semantics of/for  $\Phi^{\text{DR}}$ ) the positivist requirement of science methodology emphasizing that proper *empirical* knowledge is to be *sensually verifiable* either actually or potentially (in principle). Obviously, the suggested representation (model) of the sensual verification-ism substantially differs from its original actually realized in history of philosophy of science (here I mean the substantial difference of *sufficient* and *necessary* conditions of/for quite rational affirming that knowledge is actually *empirical one*).

VII. For every formula  $\alpha$ , when it is true that  $\diamond^D \lceil \alpha$ , then it is also true that  $E^D\alpha$ . This represents (in truth-related semantics of/for  $\Phi^{\text{DR}}$ ) exactly that positivist principle of science methodology (i.e. indispensable condition of/for scientific-ness of knowledge), according to which (principle), *actually scientific* (=empirical) knowledge has to be *falsifiable* either actually or potentially (in prin-

ciple). Evidently, the suggested representation (model) of the falsification-ism substantially differs from its original actually realized in history of philosophy of science. Here I imply the noteworthy difference between *sufficient* and *necessary* conditions of/for *empirical-ness* of knowledge (the indicated situation is *analogous*, i.e. *different but similar*, to the one described in the above-formulated item VI). Special attention to the potential falsifiability of proper scientific knowledge had been paid in XX century. However, some outstanding specialists in science philosophy of XX century have systematically criticized the overly obsessive (pseudo-rationalistic) tendency to oust and replace the limitedly respectable criterion of *sensual verifiability* (potential verification by sensation) with the limitedly respectable criterion of *falsifiability* (potential refutation), in rational philosophy, methodology and logic of science proper. The opponents reasonably pointed out that, according to history and philosophy of science, neither verifiability, nor falsifiability, nor their non-excluding disjunction are *necessary* conditions of/for actually *empirical* (=proper *scientific*) knowledge; the stormy discussion implies a nontrivial hypothesis that, perhaps, there are also some additional not yet recognized nontrivial principles (criteria) of perfect empirical-ness (actual scientific-ness) of human knowledge. Which still not well-recognized principles (criteria) are meant here? In this paper, a nontrivial hypothetic answer to this hard question is given by the below-placed statement VIII representing also a *sufficient* (but also a not necessary) condition of/for quite rational affirming actual *empirical-ness* (=scientific-ness) of human knowledge.

VIII. For any formulae  $\alpha$  and  $\omega$  of  $\Phi^{\text{DR}}$ , if it is possible that proper logic values of the expressions (formula  $\Omega^D \omega$  possesses logic value “t”) and (formula  $\omega$  possesses logic value “t”) are different, then it is true that  $E^D \alpha$ . Here it is worth recalling that  $\Omega^D$  stands for an (arbitrarily taken) element of the set of *de dicto* perfection-modalities which (set) is defined above. In standard model of  $\Phi^{\text{DR}}$ , the *necessary and sufficient* condition (truth-related one) of *empirical-ness* of knowledge is the below-formulated condition IX.

IX. For every formulae  $\alpha$  and  $\omega$ , when and only when, in a standard model of  $\Phi^{\text{DR}}$ , it is true that  $E^D \alpha$ , then, in the standard model of  $\Phi^{\text{DR}}$ , either (1) it is true that  $\diamond^D \neg \alpha$ , or (2) it is true that  $\diamond^D S \alpha$ , or (3) it is

true that it is possible that it is false that proper logic values (truth-ones) of formulae  $\omega$  and  $\Omega^D \omega$  are identical.

Taking the above-formulated conditions VIII and IX seriously, one can discover (notice and recognize) that, generally speaking, knowledge can be actually *empirical* one even then, when it is neither sensually *verifiable* nor actually *falsifiable*. Being psychologically unexpected this discovery destroys the habitual paradigm as it goes far beyond traditional tenets of the classical empiricism. Nevertheless, this substantial reconstruction of precise formulation of the criterion (=necessary-and-sufficient-condition) of *empirical-ness* of knowledge makes it possible easily to understand and simply to explain some hitherto mysterious facts of cognition history.

Truth-conditions for formulae of  $\Phi^{\text{DR}}$ , containing the expression  $A^D \alpha$ , are defined in this article as follows:

X. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when it is true that  $A^D \alpha$ , then it is true that  $K^D \alpha$ .

XI. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when it is true that  $A^D \alpha$ , then it is true that  $\alpha$ .

XII. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when it is true that  $A^D \alpha$ , then it is true that  $\Box^D \alpha$ .

XIII. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when it is true that  $A^D \alpha$ , then it is true that  $\bigvee^D S \alpha$ .

XIV. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when  $A^D \alpha$ , then for every formula  $\omega$  of  $\Phi^{\text{DR}}$ , the formal-logic tantamount-ness ( $(\Omega^D \omega) \equiv \omega$ ) is valid, or, in other words, ((formula  $\Omega^D \omega$  possesses logic value “t”), when and only when ( $\omega$  possesses logic value “t”)), where symbol  $\Omega^D$  stands for any (arbitrarily taken) element of the above-defined set of *de dicto* perfection-modalities.

XV. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when  $A^D \alpha$ , then, for every formula  $\omega$  of  $\Phi^{\text{DR}}$ , it is true that  $(\Xi^D \omega \equiv \Omega^D \omega)$ , where symbols  $\Xi^D$  and  $\Omega^D$  stand for any (arbitrarily taken) elements of the above-defined set of *de dicto* perfection-modalities.

XVI. For every formula  $\alpha$  of  $\Phi^{\text{DR}}$ , when  $A^D \alpha$ , then, for every modality  $\Omega$  and every term  $t_k$  of  $\Phi^{\text{DR}}$ , it is true that proper logic values (truth ones) of  $\Omega^D[t_k]$  and  $[\Omega^R t_k]$  are identical (do coincide).

The sentence {If  $A^D\alpha$ , then, for every term  $t_i$ , ((formula  $T^D[t_i]$  possesses logic value “t”)  $\equiv$  (formula  $[t_i]$  possesses logic value “t”))} is nothing but a concrete *particular* case (instance) of the above-formulated condition XIV.

The below-located statements a) and b) are quite representative particular cases (instances) of the condition XV:

a) If  $A^D\alpha$ , then, for every formula  $\omega$  of  $\Phi^{DR}$ , it is true that ( $T^D\omega \equiv C^D\omega$ ), where “ $T^D$ ” stands for *de dicto* modality “it is *true* that ...”, while “ $C^D$ ” stands for *de dicto* modality “it is *consistent* that ...”.

b) If  $A^D\alpha$ , then, for every formula  $\omega$  of  $\Phi^{DR}$ , it is true that ( $\mathcal{K}^D\omega \equiv C^D\omega$ ), where “ $\mathcal{K}^D$ ” stands for the *existence de dicto* modality, while “ $C^D$ ” stands for the *consistency de dicto* modality.

Keeping all the above-said in mind, one can arrive to the conclusion that, when the truth condition of/for  $A^D\alpha$  is fulfilled (certainly, this case is *extraordinary* one), then the notorious *modal collapse* happens, namely: *all* the perfection-modalities *de-dicto* are formally-logically tantamount to each other and, consequently, may be resolutely eliminated from relevant expressions; without any change of truth-values of these expressions, hence, under this *extraordinary* condition,  $\Phi^{DR}$  turns from the *multimodal* theory into the classical propositional calculus which is consistent and complete one. It is an almost common belief that deductive provability of the modal collapse is a grave defect (existentially significant flaw) of/for every formal axiomatic theory of modalities. In other words, when the modal collapse is a theorem of a general theory of modalities, then the general theory of modalities ceases to exist as such. Evidently, this is really so, if one speaks in general. But, it is worth emphasizing here that, generally speaking, the formula implying modal-collapse is not deductively provable in  $\Phi^{DR}$ . The modal collapse is formally-logically (deductively) derivable in  $\Phi^{DR}$  *not in general*, but only from the very strong *assumption* (*extraordinary hypothesis*) that  $A^D\alpha$ . If  $E^D\alpha$ , then the modal collapse does not exist in  $\Phi^{DR}$ . Consequently, in general, as a whole,  $\Phi^{DR}$  is free of the modal-collapse paradox (while, in that *extraordinary* particular case, when  $A^D\alpha$ , the modal collapse is neither a flaw nor a paradox but quite a *normal* condition).

I faith that, in future, the powerful tendency to multimodality in philosophical logic (in particular, the strong propensity to think of

many-ness of qualitatively different kinds of *de dicto* modalities) instantiated in this paper by formulae  $G^D a$ ,  $T^D a$ ,  $C^D a$ ,  $\mathcal{K}^D a$ , shall lead philosophers to a significantly clearer recognition of the inevitable ambiguity, polysemy, and essential many-valued-ness of natural human language semantics. The given paper is the start of/for promising scientific investigations in the indicated direction.

#### **1.4. Utilizing the Logical Square of Concept Opposition and Blanché Hexagon as Graphic Models (Means of/for Visualization) of formal-logical interconnections among perfection-modalities de-dicto and of formal-axiological interconnections among perfection-modalities de-re, in the Formal Philosophy System $\Phi DR$**

One of theoretically interesting options and prospective directions of investigating *perfection-modalities de-dicto* and *de-re*, is attempting to utilize the hexagon [Blanché 1966] containing the square of opposition for graphic modelling (visualizing) the systems of formal interconnections among the modalities belonging to the two qualitatively different types under consideration. In [Lobovikov 2024, p. 34], the opposition square included into Blanché hexagon has been used for graphic modelling (visualizing) the systems of *formal-logical* interconnections among the *epistemic* modalities  $K$ ,  $A$ ,  $E$ , which are *de dicto* ones. Hereunder I suggest quite a new *formal-logical* interpretation of the geometric figures. The *formal-logic-square-and-hexagon* is exploited below as a *graphic model (means of/for visualization) of formal-logical interconnections among any possible perfection-modalities de-dicto*.

With respect to *all* perfection-modalities *de dicto*, all the truth-related formal-logic rules of the traditional formal-logic-square-and-hexagon are valid. Fig. 2 models *proper logic* aspect of *de dicto* modal notion systems. According to [Inwagen, Sullivan, and Bernstein 2023], it is highly likely that discourses of *de-dicto* modalities belong to the subject-matter of logic, while discourses of *de-re* modalities belong to the subject-matter of metaphysics. If this is really so, then, in the present article, it is worth undertaking a hypothetic-deductive investigation of an extraordinary possibility of inventing (intentional constructing by analogy) a hitherto never considered *formal-*

*axiological* square-and-hexagon for modeling graphically not formal-logical but formal-axiological (i.e. proper metaphysical) aspect of relations among *de re* modalities (objects of metaphysics proper).

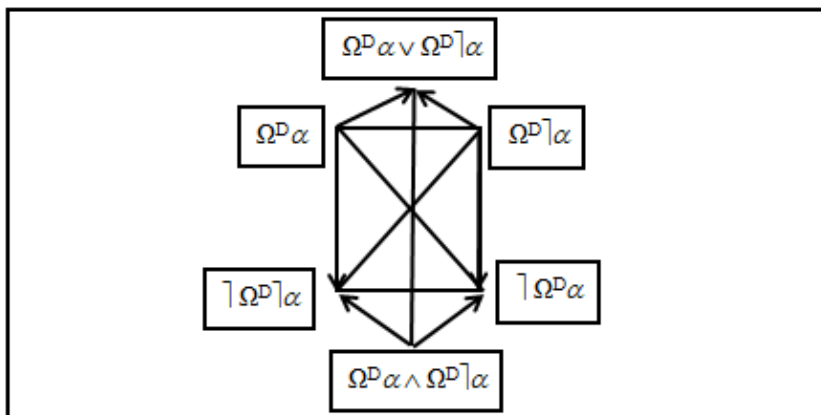


Fig. 2. The multimodal square and hexagon of opposition visually modeling *formal-logical* interconnections among any perfection-modalities *de-dicto*<sup>1</sup>

If the statement that *de re* modalities belong to the subject-matter of metaphysics is true, then, if the above-mentioned challenging *formal-axiological* conception of metaphysics [Лобовиков 2007; Lobovikov 2022] is accepted (at least as a hypothesis worthy of investigation), then the relations among modalities *de re* are to be *formal-axiological* ones (*formal-axiological* consequence, *formal-axiological* contradiction, etc.).

I think that the pedagogically and heuristically fruitful idea of graphic modelling abstract relations among abstract objects in abstract systems of any nature (well-known and used systematically since ancient times) can be exploited successfully (by analogy) also in connec-

<sup>1</sup> Probably, it is worth recalling here that, in harmony with standard handbooks of the so-called traditional formal logic, in Fig. 1, the bottom horizontal line represents graphically the *sub-contrariety* relation; the upper horizontal line represents visually the *contrariety* relation; the *formal-logic-consequence* relations are graphically modeled by arrows; the lines crossing the square model graphically the *contradictoriness* relations.

tion with the *formal-axiological* systems. This abstract hypothetical statement is exemplified below by Fig. 3, in which square and hexagon model graphically not the habitual proper *formal-logical* relations but unhabitual (qualitatively new) *formal-axiological* ones among *modalities-de-re* as *evaluation-functions*. To exclude possible misunderstanding, here it is worth highlighting that, in Fig. 3, the artificial language expressions  $\Omega^R t_k$ ,  $N^R \Omega^R t_k$ ,  $\Omega^R N^R t_k$ ,  $K^2 \Omega^R t_k \Omega^R N^R t_k$ , and other mean not “formulae” but “terms” of the theory  $\Phi^{DR}$ . Thus, in Fig. 3, the lines visually model the relations not among formulae but among terms of  $\Phi^{DR}$ .

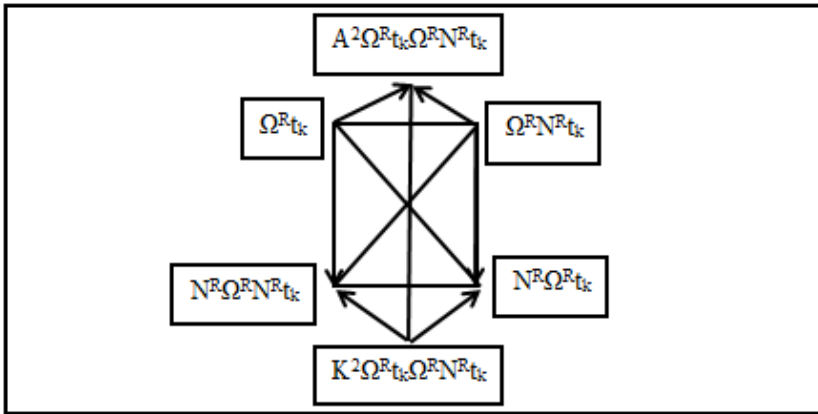


Fig. 3. The multimodal square and hexagon of opposition visually modeling *formal-axiological* interconnections among any perfection-modalities *de-re*

To exclude possible misunderstanding, it is necessary to emphasize here that, in Fig. 3, arrows model graphically the relations of “*formal-axiological* consequence”; the lines crossing the square model graphically the relations of “*formal-axiological* contradictoriness”; the upper horizontal line represents visually the relation of “*formal-axiological* contrariety”; the bottom horizontal line represents graphically the relation of “*formal-axiological* subcontrariety”. From the psychological point of view, it is highly likely that people can mistake the “*formal-axiological*” relations for the

corresponding “*formal-logical*” ones. However, from the proper logic and theoretical-philosophy points of view, this gravely dangerous mistake must be excluded due to perfect recognizing that “*formal-axiological*” and “*formal-logical*” are not synonyms: they have not identical but qualitatively different meanings. (In this connection, readers are advised accurately to compare the relevant definitions, which are strikingly *analogous* but *not identical*; and recognizing *their difference is very important* for dismissing possible illusions of paradoxes. I recall that relevant precise definitions of “*formal-axiological contradiction*”, “*formal-axiological equivalence*”, et al are provided above in this article.)

The *multimodal* opposition squares and hexagons represented above by Fig. 2 and Fig. 3 have been never published before. Certainly, various qualitatively different concrete interpretations (very interesting and *important* particular cases) of Fig. 2 have been already either discovered (contingently found) or deliberately invented (intentionally constructed) much earlier, especially by Robert Blanché and Jean-Yves Béziau [Blanché1966; Béziau 2003; 2012a; 2012b; Béziau, Buchsbaum, Rey 2018]. But the *multimodal* graphic scheme located in Fig. 2 visually represents for the first time *a universal* of/for the mentioned interesting and important *particular* cases: the *alethic* modalities make up an important *particular* case of Fig. 2; the *deontic* modalities make up an important *particular* case of Fig. 2 as well, and so on; the *volatile* vistas are *potentially infinite* in the indicated respect. This is promising original fruitful applications of the *universal* (visually modeled by the figures) to unexpected contexts and problems to be encountered in future.

## Discussion and Exemplification of the Novel Results

Let us begin with discussing the equivalence ( $\Box^D[x] \leftrightarrow [\Box^R x]$ ) formally derived above (in  $\Phi^{DR}$ ) from the presumption of *a-priori*-ness of knowledge. While talking and writing of ontology in physics and metaphysics, thinkers deal with two kinds of necessity: *de dicto* and *de re*. The sentence “It is *necessary* that there is conservation of energy” exploits the *necessity de dicto*, while the sentence “There

is *necessity of conservation of energy*” exploits the *necessity de re*. Let us fix and discuss such an *interpretation* of  $(\Box^D[x] \leftrightarrow [\Box^R x])$ , in which *conservation-of-energy* is a value of the variable  $x$ . In this fixed interpretation,  $[x]$  is a dictum (proposition) “*conservation-of-energy exists*”, and  $\Box^D[x]$  is a *de dicto* modal proposition (modal statement about the dictum  $[x]$ ). Along with this, in the fixed interpretation under discussion,  $\Box^R x$  is *not a dictum* but a physical *object*: either a physical state of matter, or a material process, or a material thing, i.e. *re*. Both  $x$  and  $\Box^R x$  belong to the material world: they are not dictums. Consequently, in the concrete interpretation under discussion,  $[\Box^R x]$  is a *de re* modal proposition – modal statement about what exists (in the real world). It is easy to see that, in the given interpretation, the equivalence  $(\Box^D[x] \leftrightarrow [\Box^R x])$ , is true. Hence, with respect to physics, the axiom-scheme AX-7 (of  $\Phi^{DR}$ ) looks quite adequate.

According to history of philosophy, exactly physics had been “native land” (basic professional education) of/for plenty of positivist-minded philosophers negatively treating metaphysics, axiology, and theology. In this connection, it is curious that the equivalence  $(\Box^D[x] \leftrightarrow [\Box^R x])$ , is true not only in proper physical but also in proper theological interpretation. If such a theological interpretation of  $(\Box^D[x] \leftrightarrow [\Box^R x])$ , is suggested, in which a name of God is substituted for  $x$ , then, being theologically interpreted,  $(\Box^D[x] \leftrightarrow [\Box^R x])$  is translated into the natural language by the following biconditional: (it is *necessary that* (God exists)), if and only if (*necessity of* God) exists). Here, (*necessity of* God) is not a proposition but a *re*. Certainly, from the classical theology viewpoint, the biconditional is true.

Now let us consider the sentence “*In God we trust*” located on the green banknotes. If it was written “*We trust that (God exists)*”, then “*trust (that ...)*” would be a modality *de dicto*. But the “*trust in (what, whom)*” is a modality *de re*. In the sentence “*There is (our trust in God)*”, or “*(Our trust in God) exists*”, God is considered as a Divine *Re*. For true believers, it is certain that the biconditional “*We trust that (God exists), if and only if (Our trust in God) exists*” is true. This very special biconditional sentence is a concrete exemplification of the abstract equivalence  $(F^D[t_k] \leftrightarrow [F^R t_k])$ . The true positivists hating talks about God and ignoring *de re* modalities

(as metaphysical ones) shall be not happy with the equivalence ( $F^D[t_k] \leftrightarrow [F^R t_k]$ ) exemplified theologically. But, in my opinion, for unpedjudiced formalist-minded philosophers and abstractly thinking logicians this quite formal *equivalence of de dicto and de re beliefs* (not necessarily having theological interpretation) is of some professional interest.

In mathematical logic proper, there are quite natural and fruitful discourses of *provability as modality* [Кузнецов и Муравицкий 1980; Boolos and Sambin 1991]. However, as a rule, the modality discourses of provability are reduced to *provability as modality de dicto*, exclusively. In addition to these writings on proper logic, let us try to consider evidence (proof) and proving (making evident) as modalities *de re*. Obviously, this is something unusual (not habitual) for logicians proper, but lawyers (especially the ones taking part in litigations) are used to the so-called “*real evidences*”. For litigants, judges, and investigators, “proving thing (*re*)” has value. Hence, *axiology* (dealing with moral-legal value) is quite relevant to *evidence (proof) as modality de re*. Often, during legal trials, making up (or ruining) a proof (finding or destroying an evidence) are operations with objects (things); in this relation, evidences (proofs) are *objects* of an activity of appropriate lawyers, hence, in this relation, proofs (evidences) are nothing but things (*re*), to which the activity is applied.

Let us consider the following interpretation: term *d* is interpreted as John Fitzgerald Kennedy; term *h* is interpreted as Lee Harvey Oswald; the two-placed term (applied to *d* and *h*) is interpreted as (termination of J. F. Kennedy by L. H. Oswald). In this interpretation, term  $T^2dh$  means the act of termination of Kennedy by Oswald, which act is (certainly) not a proposition but an event (*re*). (Here it is worth recalling that the *evaluation-function*  $T^2xy$  – “*termination of x by y*” is defined above by Table 2.) In the interpretation under consideration, term  $P^R T^2dh$  (made by attaching  $P^R$  to term  $T^2dh$ ) means *de-re-modality-term* “*evidence (proof) of (what) T<sup>2</sup>dh*”. Thus, in the interpretation under discussion, terms *d*, *h*,  $T^2dh$ ,  $P^R T^2dh$  are not propositions (dictums), but things (res): persons, actions, events, states of affairs, real evidences, which are not proper logic phenomena. Nonetheless, in the fixed interpretation,  $[P^R T^2dh]$  denotes the either true or false statement (dictum) that there is an evidence that Kennedy has

been killed by Oswald. In its turn,  $[T^2dh]$  is also an either true or false proposition (dictum), namely, the statement that Kennedy has been killed by Oswald. Consequently, both formulae  $[T^2dh]$  and  $[P^RT^2dh]$  belong to the subject-matter of logic proper. According to the above-given definition of “formula of  $\Phi^{DR}$ ”, if  $[T^2dh]$  is a formula, and  $P^D$  is a symbol denoting a de-dicto-modality, then  $P^D[T^2dh]$  is a formula of  $\Phi^{DR}$  as well. Usual content analysis and intuition have nothing against proclaiming that  $(P^D[T^2dh] \leftrightarrow [P^RT^2dh])$  is true. Thus, the above-considered concrete interpretation makes up an exemplification of the abstract equivalence  $(P^D[t_k] \leftrightarrow [P^Rt_k])$ .

### 1. A Proof of Consistency of $\Phi^{DR}$

First of all, let us leave the above-presented *axiom-schemes* of  $\Phi^{DR}$  for the below-located *proper-axioms* of  $\Phi^{DR}$ .

Axiom AX\*-1:  $A^Dp \supset (\Box^Dq \supset q)$ .

Axiom AX\*-2:  $A^Dp \supset (\Box^D(p \supset q) \supset (\Box^Dp \supset \Box^Dq))$ .

Axiom AX\*-3:

$A^Dp \leftrightarrow (K^Dp \ \& \ (\neg\Diamond^D\neg p \ \& \ \neg\Diamond^D\neg Sp \ \& \ \Box^D(q \leftrightarrow \Box^Dq)))$ .

Axiom AX\*-4:  $E^Dp \leftrightarrow (K^Dp \ \& \ (\Diamond^D\neg p \ \vee \ \Diamond^D\neg Sp \ \vee \ \neg\Box^D(q \leftrightarrow \Box^Dq)))$ .

Axiom AX\*-5:  $\Box^Dp \supset \Diamond^Dp$ .

Axiom AX\*-6:  $(\Box^Dq \ \& \ \Box^D\Box^Dq) \supset q$ .

Axiom AX\*-7:  $A^Dp \supset (\Box^D[x] \leftrightarrow [\Box^Rx])$ .

Axiom AX\*-8:  $(B^Rx=+=J^Rx) \leftrightarrow (G^D[B^Rx] \leftrightarrow G^D[J^Rx])$ .

Axiom AX\*-9:  $(B^Rx=+=g) \supset \Box^DG^D[B^Rx]$ .

Axiom AX\*-10:  $(B^Rx=+=b) \supset \Box^DW^D[B^Rx]$ .

Axiom AX\*-11:  $(G^Dp \leftrightarrow \neg W^Dp)$ .

Дефиниция DF\*-1:  $\Diamond^Dp$  есть *сокращенное название* для  $\neg\Box^D\neg p$ , т.е.  $(\Diamond^D\omega \leftrightarrow \neg\Box^D\neg\omega)$ , по определению.

Now let us consider a function # called “an *interpretation* of the axiom system” and defined precisely by the following list of items 1)–26) as a whole. (Here it worth noting that, in this part of the article, the symbol “t” stands for “true” and the symbol “f” stands for “false”).

1) For any formulae  $\delta$  and  $\lambda$  of the formal theory  $\Phi^{DR}$ , and for any binary classical logic connective  $\otimes$ , it is true that  $\#(\delta \otimes \lambda) = (\#\delta \otimes \#\lambda)$ .

2) For any formula  $\delta$  of the formal theory  $\Phi^{\text{DR}}$ , it is true that  $\# \neg \delta = \neg \# \delta$ . (Here the symbol “ $\neg$ ” stands for the classical-logic unary-operation called “negation”.)

$$3) \# A^{\text{D}} p = f.$$

$$4) \# \square^{\text{D}} q = f.$$

$$5) \# q = t.$$

$$6) \# p = t.$$

$$7) \# \square^{\text{D}} (p \supset q) = f.$$

$$8) \# \square^{\text{D}} p = f.$$

$$9) \# K^{\text{D}} p = t.$$

$$10) \# \diamond^{\text{D}} \neg p = t.$$

$$11) \# \diamond^{\text{D}} S p = t.$$

$$12) \# \square^{\text{D}} (q \leftrightarrow \square^{\text{D}} q) = f.$$

$$13) \# E^{\text{D}} p = t.$$

$$14) \# \diamond^{\text{D}} p = t.$$

$$15) \# \square^{\text{D}} \square^{\text{D}} q = f.$$

$$16) \# [B^{\text{R}} x] = t.$$

$$17) \# [J^{\text{R}} x] = t.$$

$$18) \# [x] = t.$$

$$19) \# \square^{\text{D}} [x] = t.$$

$$20) \# [\square^{\text{R}} x] = t.$$

21)  $\# G^{\text{D}} [B^{\text{R}} x] = t$ . In the interpretation under consideration, the symbol  $B^{\text{R}}$  stands for the above-tabularly-defined *evaluation-function* “positive *aesthetic value (beauty) of (what, whom) x*”. See table 1.

22)  $\# G^{\text{D}} [J^{\text{R}} x] = t$ . In the interpretation under consideration, the symbol  $J^{\text{R}}$  stands for the above-tabularly-defined *evaluation-function* “positive *hedonistic value (pleasantness) of (what, whom) x*”, or in other words, “*happiness with, or joy from (what, whom) x*”. See table 2.

$$23) \# (B^{\text{R}} x = + = J^{\text{R}} x) = t.$$

$$24) \# (B^{\text{R}} x = + = g) = f.$$

$$25) \# (B^{\text{R}} x = + = b) = f.$$

$$26) \# \square^{\text{D}} G^{\text{D}} [B^{\text{R}} x] = f.$$

$$27) \# \square^{\text{D}} W^{\text{D}} [B^{\text{R}} x] = f.$$

$$28) \# G^{\text{D}} p = t.$$

$$29) \# W^{\text{D}} p = f.$$

$$30) \# \square^{\text{D}} \neg p = f.$$

In relation to the above-defined *interpretation* # of the formal theory  $\Phi^{\text{DR}}$ , the axioms AX\*-1 – AX\*-11 are true, the definition DF\*-1 is true, and the only formal logical derivation rule (*modus ponens*) preserves the truth, consequently, the suggested *interpretation* # of the formal theory  $\Phi^{\text{DR}}$ , is a *model* of/for this formal theory and, consequently,  $\Phi^{\text{DR}}$  is logically *consistent* (as there is a *model* of/for it).

2. *A Proof of Unprovability of  $(\mathcal{C}^{\text{D}}[x] \leftrightarrow \text{D}^{\text{D}}[x])$  in  $\Phi^{\text{DR}}$ , and a Proof of Unprovability of  $([\mathcal{C}^{\text{R}}x] \leftrightarrow [\text{D}^{\text{R}}x])$  in  $\Phi^{\text{DR}}$*

In relation to the schemes of theorems  $(\mathcal{C}^{\text{D}}[x] \leftrightarrow \text{D}^{\text{D}}[x])$  and  $([\mathcal{C}^{\text{R}}x] \leftrightarrow [\text{D}^{\text{R}}x])$ , the following nontrivial questions arise in  $\Phi^{\text{DR}}$ . Is constructiveness (or constructivity) identical to algorithmicalness? Is it necessary that all constructions are algorithmical? Can it be possible in mathematics that an object can be constructed but an algorithm of construction of the object does not exist? Abstractly speaking, in principle, it may happen that one (for example a mathematician, logician, specialist in informatics-and-computer-science) can disagree with biconditionals  $(\mathcal{C}^{\text{D}}[x] \leftrightarrow \text{D}^{\text{D}}[x])$  and  $([\mathcal{C}^{\text{R}}x] \leftrightarrow [\text{D}^{\text{R}}x])$  as necessarily *universal* statements, and can provide concrete counterarguments grounding the disagreement. For having an adequate attitude to this nontrivial abstract possibility, it is worth making acquaintance with [Мартин-Лёф 1975; Нагорный 2010; Непейвода 2010; 2011; 2012; 2014]. Concerning the above-said of the hypothetical possibility of rational counterargumentation against the equivalences  $(\mathcal{C}^{\text{D}}[x] \leftrightarrow \text{D}^{\text{D}}[x])$  and  $([\mathcal{C}^{\text{R}}x] \leftrightarrow [\text{D}^{\text{R}}x])$ , it is relevant to remark here that these equivalences are not provable in  $\Phi^{\text{DR}}$ .

To prove their unprovability in  $\Phi^{\text{DR}}$ , let us undertake the following steps. At first, let us make up a new formal axiomatic theory  $\mathcal{L}$  by adding a new axiom, namely AX\*-12:  $\neg(\mathcal{C}^{\text{D}}[x] \leftrightarrow \text{D}^{\text{D}}[x])$ , to the above-placed list of axioms of  $\Phi^{\text{DR}}$ . A model of/for the new formal axiomatic theory  $\mathcal{L}$  is created by adding a couple of new items, namely, the following 31) and 32) to the above-given definition of interpretation-function #. Let “#31&32” be the name of/for the new function to be used as an interpretation of/for  $\mathcal{L}$ .

$$31) \# \mathcal{C}^{\text{D}}[x] = t.$$

$$32) \# \text{D}^{\text{D}}[x] = f.$$

In relation to the above-defined *interpretation* # of the formal theory  $\Phi^{DR}$ , the axioms AX\*-1 – AX\*-12 are true, the definition DF\*-1 is true, and the only formal logical derivation rule (*modus ponens*) preserves the truth, consequently, the suggested *interpretation* “#+31&32” of the formal theory  $\mathfrak{L}$ , is a *model* of/for this formal theory and, consequently,  $\mathfrak{L}$  is logically *consistent* (as there is a *model* of/for it).

Let us assume that  $(\odot^D[x] \leftrightarrow D^D[x])$  is formally provable in  $\Phi^{DR}$ . Then  $\mathfrak{L}$  is logically inconsistent as both  $(\odot^D[x] \leftrightarrow D^D[x])$  and  $\neg(\odot^D[x] \leftrightarrow D^D[x])$  are formally provable in  $\mathfrak{L}$ . But it is proved that  $\mathfrak{L}$  is logically consistent as it has a model. Consequently, our assumption is wrong, consequently,  $(\odot^D[x] \leftrightarrow D^D[x])$  is not provable in  $\Phi^{DR}$ . Here we are.

Now let us create a new formal axiomatic theory  $\Theta$  by adding a new axiom, namely AX\*-12:  $\neg([\odot^R x] \leftrightarrow [D^R x])$  to the above-placed list of axioms of  $\Phi^{DR}$ . A model of/for the new formal axiomatic theory  $\Theta$  is created by adding a couple of new items, namely, the following 33) and 34) to the above-given definition of interpretation-function #. Let “#+33&34” be the name of/for the new function to be used as an interpretation of/for  $\Theta$ .

$$33) \#[\odot^R x] = t.$$

$$34) \#[D^R x] = f.$$

In relation to the above-defined *interpretation* # of the formal theory  $\Phi^{DR}$ , the axioms AX\*-1 – AX\*-12 are true, the definition DF\*-1 is true, and the only formal logical derivation rule (*modus ponens*) preserves the truth, consequently, the suggested *interpretation* “#+33&34” of the formal theory  $\Theta$ , is a *model* of/for this formal theory and, consequently,  $\Theta$  is logically *consistent* (as there is a *model* of/for it).

Let us assume that  $([\odot^R x] \leftrightarrow [D^R x])$  is formally provable in  $\Phi^{DR}$ . Then  $\Theta$  is logically inconsistent as both  $([\odot^R x] \leftrightarrow [D^R x])$  and  $\neg([\odot^R x] \leftrightarrow [D^R x])$  are formally provable in  $\Theta$ . But it is proved that  $\Theta$  is logically consistent as it has a model. Consequently, our assumption is wrong, hence,  $([\odot^R x] \leftrightarrow [D^R x])$  is not provable in  $\Phi^{DR}$ . Here we are.

3. *Formal Proofs of*  $(A^D\alpha \supset (\odot^D[x] \leftrightarrow D^D[x]))$  *and*  $(A^D\alpha \supset ([\odot^R x] \leftrightarrow [D^R x]))$  *in*  $\Phi^{DR}$

Notwithstanding the above-said,  $(A^D\alpha \supset (\odot^D[x] \leftrightarrow D^D[x]))$  and  $(A^D\alpha \supset ([\odot^R x] \leftrightarrow [D^R x]))$  are formally provable in  $\Phi^{DR}$ . To examine

this statement, readers are invited to make acquaintance with the following formal proofs.

- 1)  $\vdash A^D\alpha \leftrightarrow (K^D\alpha \& (\neg\Diamond^D\neg\alpha \& \neg\Diamond^DS\alpha \& \Box^D(\beta \leftrightarrow \Omega^D\beta)))$ : axiom scheme AX-3.
- 2)  $\vdash A^D\alpha \supset (K^D\alpha \& (\neg\Diamond^D\neg\alpha \& \neg\Diamond^DS\alpha \& \Box^D(\beta \leftrightarrow \Omega^D\beta)))$ : from 1) by elimination of  $\leftrightarrow$ .
- 3)  $A^D\alpha$  assumption.
- 4)  $A^D\alpha \vdash (K^D\alpha \& (\neg\Diamond^D\neg\alpha \& \neg\Diamond^DS\alpha \& \Box^D(\beta \leftrightarrow \Omega^D\beta)))$ : from 2) and 3) by *modus ponens*.
- 5)  $A^D\alpha \vdash \Box^D(\beta \leftrightarrow \Omega^D\beta)$ : from 4) by elimination of  $\&$ .
- 6)  $\vdash A^D\alpha \supset (\Box^D\beta \supset \beta)$ : axiom scheme AX-1.
- 7)  $\vdash A^D\alpha \supset (\Box^D(\beta \leftrightarrow \Omega^D\beta) \supset (\beta \leftrightarrow \Omega^D\beta))$ : from 6) by substituting  $(\beta \leftrightarrow \Omega^D\beta)$  for  $\beta$ .
- 8)  $A^D\alpha \vdash (\Box^D(\beta \leftrightarrow \Omega^D\beta) \supset (\beta \leftrightarrow \Omega^D\beta))$ : from 3) and 7) by *modus ponens*.
- 9)  $A^D\alpha \vdash (\beta \leftrightarrow \Omega^D\beta)$ : from 5) and 8) by *modus ponens*.
- 10)  $A^D\alpha \vdash (\beta \leftrightarrow \Xi^D\beta)$ : from 9) by substituting  $\Xi^D$  for  $\Omega^D$ .
- 11)  $A^D\alpha \vdash (\Xi^D\beta \leftrightarrow \beta)$ : from 10) by commutativity of  $\leftrightarrow$ .
- 12)  $A^D\alpha \vdash (\Xi^D\beta \leftrightarrow \Omega^D\beta)$ : from 11) and 9) by transitivity of  $\leftrightarrow$ .
- 13)  $\vdash (A^D\alpha \supset (\Xi^D\beta \leftrightarrow \Omega^D\beta))$ : from 12) by introduction of  $\supset$ .
- 14)  $\vdash (A^D\alpha \supset (\odot^D[x] \leftrightarrow D^D[x]))$ : from 13) by substituting:  $\odot^D$  for  $\Xi^D$ ,  $D^D$  for  $\Omega^D$ , and  $[x]$  for  $\beta$ .

Here you are:  $(A^D\alpha \supset (\odot^D[x] \leftrightarrow D^D[x]))$  is formally proved in  $\Phi^{DR}$  by the above-presented succession 1)–14). Now let us continue this succession in the following way.

- 15)  $\vdash A^D\alpha \supset (\Omega^D[t_k] \leftrightarrow [\Omega^R t_k])$ : axiom scheme AX-7.
- 16)  $\vdash A^D\alpha \supset (\odot^D[x] \leftrightarrow [\odot^R x])$ : from 15) by substituting:  $\odot^D$  for  $\Omega^D$ ;  $\odot^R$  for  $\Omega^R$ ;  $x$  for  $t_k$ .
- 17)  $\vdash A^D\alpha \supset (D^D[x] \leftrightarrow [D^R x])$ : from 15) by substituting:  $D^D$  for  $\Omega^D$ ;  $D^R$  for  $\Omega^R$ ;  $x$  for  $t_k$ .
- 18)  $A^D\alpha \vdash (\odot^D[x] \leftrightarrow [\odot^R x])$ : from 16) and 3) by *modus ponens*.
- 19)  $A^D\alpha \vdash (D^D[x] \leftrightarrow [D^R x])$ : from 17) and 3) by *modus ponens*.
- 20)  $A^D\alpha \vdash (\odot^D[x] \leftrightarrow D^D[x])$ : from 12) by substituting:  $\odot^D$  for  $\Xi^D$ ,  $D^D$  for  $\Omega^D$ , and  $[x]$  for  $\beta$ .
- 21)  $A^D\alpha \vdash (\odot^D[x] \leftrightarrow [D^R x])$ : from 20) and 19) by transitivity of  $\leftrightarrow$ .
- 22)  $A^D\alpha \vdash ([\odot^R x] \leftrightarrow \odot^D[x])$ : from 18) by commutativity of  $\leftrightarrow$ .

23)  $A^D\alpha \vdash ([\odot^R x] \leftrightarrow [D^R x])$ : from 22) and 21) by transitivity of  $\leftrightarrow$ .

24)  $\vdash (A^D\alpha \supset ([\odot^R x] \leftrightarrow [D^R x]))$ : from 23) by introduction of  $\supset$ .

Here we are:  $(A^D\alpha \supset ([\odot^R x] \leftrightarrow [D^R x]))$  is formally proved in  $\Phi^{DR}$  by the above-presented succession 1)–24).

### Conclusion

According to the present article, there is a one-to-one correspondence between elements of the set of modalities *de dicto* and elements of the set of modalities *de re*. Moreover, corresponding modalities *de dicto* and *de re* are equivalent, under the precisely defined *extraordinary* condition that knowledge is *a priori* (in exactly that meaning of “*a priori*”, which is precisely defined by the axiomatic system  $\Phi^{DR}$ ). Consequently, if in some concrete domain of knowledge, there are grounds for negating equivalence of some corresponding *de dicto* and *de re* modalities, then the concrete domain contains *empirical* knowledge, consequently, in this concrete domain, knowledge is *not pure a priori*. According to the given article, from the one’s disagreement with equivalence of constructiveness and algorithmicalness, it follows logically by *modus tollens* that according to the one, generally speaking, it is false that  $A^D\alpha$ . Hence, even in the concrete interpretation resulting in the one’s concrete counterexamples, schemes of formulae  $(A^D\alpha \supset (\odot^D[x] \leftrightarrow D^D[x]))$  and  $(A^D\alpha \supset ([\odot^R x] \leftrightarrow [D^R x]))$  are schemes of true statements. Thus, the extraordinary equivalence of corresponding modalities *de dicto* and *de re* essentially depends on the relevant epistemic context, namely, on accepting or rejecting the *extraordinary* epistemic assumption that knowledge is *pure a priori*.

### Литература

1. *Аристотель*. Сочинения в 4 т. Т. 2. М: Мысль, 1978.
2. *Клини С.К.* Математическая логика. М: Мир, 1973.
3. *Клини С.К.* Введение в метаматематику. М: ИЛ, 1957.
4. *Куайн, У. В. О.* С точки зрения логики. 9 логико-философских очерков. М: Канон+ РООИ «Реабилитация», 2010.

5. Куайн У. В. О. От стимула к науке. М: Канон+ РООИ «Реабилитация», 2016.
6. Кузнецов, А. В., Муравицкий, А. Доказуемость как модальность. // Актуальные проблемы логики и методологии науки. / Попович, М. В. (Ред.). Киев: «Наукова думка». 1980. С. 193–230.
7. Ламберов, Л. Д. Референция, *de re* и объяснение действий. // Известия Уральского государственного университета. Серия 3: Общественные науки. 2011. Т. 88. № 1. С. 5-13.
8. Лобовиков, В. О. Математическая этика, метафизика и естественное право (Алгебра метафизики как алгебра формальной аксиологии). Екатеринбург: Уральское отделение Российской академии наук, 2007.
9. Лукасевич, Я. Аристотелевская силлогистика с точки зрения современной формальной логики. М: Издательство иностранной литературы, 1959.
10. Маркин, В. И. Дихотомия *de re – de dicto* и аподиктическая силлогистика. *Логические исследования*. 2018. Т. 24. № 2. С. 108-115.
11. Мартин-Лёф, П. Очерки по конструктивной математике. М: Мир, 1975.
12. Нагорный Н. М. Конструктивный процесс. // Новая философская энциклопедия: В 4 т. Т. II. М.: Мысль, 2010. С. 295.
13. Непейвода Н. Н. Конструктивное направление (в математике и логике). // Новая философская энциклопедия: В 4 т. Т. II. М: Мысль, 2010. С. 293-294.
14. Непейвода Н. Н. Конструктивная математика: обзор достижений, недостатков и уроков. Часть I. // Логические исследования. 2011. Т. 17. С. 191–239.
15. Непейвода Н. Н. Конструктивная математика: обзор достижений, недостатков и уроков. Часть II // Логические исследования. 2012. Т. 18. С. 157–181.
16. Непейвода, Н. Н. Конструктивная математика: обзор достижений, недостатков и уроков. Часть III. // Логические исследования. 2014. Т. 20. С. 110-148.
17. Хлебалин, А. В. Эссенциализм и антиэссенциализм в модальной логике // Философия науки. 2003. Т. 17. № 2. С. 34-46.
18. Целищев, В. В. Понятие объекта в модальной логике. Новосибирск: Наука, 1978.
19. Béziau, J.-Y. New Light on the square of oppositions and the nameless coner. // *Logical Investigations*. 2003. № 10. 218-232.
20. Béziau, J.-Y. The New Rising of the Square of Opposition. // *Around and Beyond the Square of Opposition*, Basel, Birkhäuser, 2012 a. P. 3-19.

21. *Béziau, J-Y.* The Power of the Haxagon // *Logica Universalis*, 2012b. № 6, double issue 1-2, P. 1-43.

22. *Béziau, J-Y.* The Logical Hexagon of Analogy: Structuring the Relations between Difference, Identity and Similarity. // *Handbook of the First World Congress on Analogy*, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico, November 4-6, 2015, Poznań, Poland, Publishing House KONTEXT, 2015. P. 12-13.

23. *Béziau, J.-Y., Buchsbaum, A., Rey, Ch.* (Eds.). *Handbook of the 6<sup>th</sup> World Congress and School on Universal Logic* (June 16–26, 2018). Vichy, France: Université Clermont Auvergne. 2018.

24. *Blanché, R.* *Structures intellectuelles. Essai sur l'organisation systématique des concepts.* Paris, Vrin. 1966.

25. *Boolos, G., and Sambin, G.* Provability: the emergence of a mathematical modality. // *Studia Logica*. 1991. 50(1): 1–23.

26. *Burge, T.* *De re* Belie. // *Journal of Philosophy*. 1977. № 74. P. 338–362.

27. *Chisholm, R.* Knowledge and Belief: ‘De Dicto’ and ‘De Re’. // *Philosophical Studies*. 1976. № 29 (1). P.1–20.

28. *Egan, A., Titelbaum, M. G.* Self-Locating Beliefs. // *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). URL = <<https://plato.stanford.edu/archives/win2022/entries/self-locating-beliefs/>>.

29. *Geretto, M.* “I Think That in the Universe Nothing Is Truer Than Happiness, and Nothing Happier or Sweeter Than Truth”. On the Relationship between Happiness, Truth and Generosity in Leibniz’s Philosophy. // “Für unser Glück oder das Glück anderer” Vorträge des X Internationalen Leibniz-Kongress / Hannover, 18. – 23 Juli. 2016. Band I. (Hildesheim; Zürich; New York: Georg Olms Verlag. P.617-632.

30. *Inwagen, P. van, Meghan Sullivan, and Sara Bernstein.* *Metaphysics*. // *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), URL = <<https://plato.stanford.edu/archives/sum2023/entries/metaphysics/>>.

31. *Jeshion, R.* ‘Acquaintanceless *de re* belief’. // *Meaning and Truth: Investigations in philosophical semantics*. New York: Seven Bridges Press, 2002. P.53–78.

32. *Kaplan, D.* Quantifying // *Synthese*. 1968. № 19. P. 178–214.

33. *Keshet, E.* Split intensionality: A new scope theory of *de re* and *de dicto*. // *Linguistics and Philosophy*. 2010. № 33. P. 251–283.

34. *Keshet, E., and Schwarz, F.* *De Re - De\_Dicto*. [Online document]. Available at the site: [www.florianschwarz.net](http://www.florianschwarz.net) (Accessed on: April 09, 2024).

35. *Knuuttila, S.* Medieval Theories of Modality. // *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), URL = <<https://plato.stanford.edu/archives/sum2021/entries/modality-medieval/>>.

36. Kneale, W. (Modality de Dicto and de Re. // Logic, Methodology and Philosophy of Science. 1962. Stanford University Press, Stanford. P. 622-633.

37. Leibniz, G. W. Philosophical Essays. Indianapolis & Cambridge: Hackett Publishing Co. Inc. 1989.

38. Lagerlund, H. (2021). "Medieval Theories of the Syllogism", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2022/entries/medieval-syllogism/>>.

39. Lewis, D. Attitudes De Dicto and De Se. // Philosophical Review. 1979. № 88(4). P. 513–543.

40. Lobovikov, V. O. Artificial Intelligence and an almost Unknown Aspect of Mathematical Linguistics. // Discourse-P. 2022. № 19(3). P. 170–184.

41. Lobovikov, V. O. Hilbert's Conception of "Existence-in-Mathematics" and Modelling it by a Formal Axiomatic Theory  $\Phi+\exists$  Treating Existence not as the Quantifier but as a Modality. // *Respublica Literaria*. 2024. Vol. 5. №. 1. P. 16–50.

42. McKay, T. Representing *de re* beliefs // Linguistics and Philosophy. 1991, № 14. P. 711–739.

43. Nelson, M. Propositional attitude reports. // The Stanford Encyclopedia of Philosophy. (Spring edition 2019). <https://plato.stanford.edu/archives/spr2019/entries/prop-attitude-reports/>

44. Nelson, M. (2023). The De Re/De Dicto Distinction. Supplement to Propositional Attitude Reports // The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), URL = <<https://plato.stanford.edu/archives/spr2023/entries/prop-attitude-reports/>>.

45. Parsons, T. Articulating Medieval Logic. Oxford: Oxford University Press. 2014. p.

46. Prior, A.N. Modality de Dicto and Modality de Re. // *Theoria*. 1952. № 18(3), P. 174-180.

47. Quine, W. Quantifiers and propositional attitudes. // *Journal of Philosophy*. 1956. № 53(5). P.177-187

48. Reimer, M. A defense of *de re* belief reports. // *Mind and Language*. 1995, №10. P. 446–463.

49. Roca-Royes, S. Conceivability and De Re Modal Knowledg. // *Noûs*, 2011. № 45. P. 22–

50. Salmon, N. (1997). Is *de re* belief reducible to *de dicto*? // *Meaning and Reference* (The Canadian Journal of Philosophy: Supplementary. 1997. Vol. 23. P. 85–110.

51. *Schwitzgebel, E.* Belief. // The Stanford Encyclopedia of Philosophy (Spring 2024 Edition), URL = <<https://plato.stanford.edu/archives/spr2024/entries/belief/>>.

52. *Sosa, E.* (1970). Propositional attitudes *de dicto* and *de re*. // Journal of Philosophy. 1970. №. 67. P. 883–896.

53. *Stanley, J.* Modality and What is Said. // *Noûs*, 2002. № 36(s16). P. 321–344.

54. *Taylor, K.*, ‘*De re* and *de dicto*: Against the conventional wisdom’. // *Philosophical Perspectives*, 2002. № 16. P. 225–265.

55. *Turner, J.* 2010, Fitting Attitudes *De Dicto* and *De Se*. // *Noûs*, 2010. № 44(1). P. 1–9.

56. *Wierenga, E.* Omniscience. // The Stanford Encyclopedia of Philosophy (Summer 2021 Edition) URL = <<https://plato.stanford.edu/archives/sum2021/entries/omniscience/>>.

57. *Williamson, T.* Modal Logic as Metaphysics. Oxford: Oxford University Press. 2013.

58. *Wolenski, J.* Truth as Modality. Rzeszow, Poland, University of Information, Technology and Management, 2016

59. *Wright, G. H.* (1996). Truth-Logics. // *Acta Philosophica Fennica*. Georg Henrik von Wright. Six Essays in Philosophical Logic. 1996. Vol. 60. P. 71–91.

60. *Zhang, Y., Davidson, K.* De re is less favorable than de dicto – An Experimental Investigation of Belief Report. P. 1–21. PDF [Online]. Available at: <https://scholar.harvard.edu/files/files/plc2020> (Accessed: April 11, 2024).

## References

1. *Aristotle.* (1978). Writings in Four Volumes. Vol. 2. Moscow. 687 p. (in Russian)

2. *Kleene, S. C.* (1973). Mathematical Logic. Moscow. 480 p. (in Russian)

3. *leene, S. C.* (1957). Introduction to Metamathematics. Moscow. 526 p. (in Russian)

4. *Quine, W. V. O.* (2010). From a Logical Point of View. 9 Logico-Philosophical Essays. Moscow. 272 p. (in Russian)

5. *Quine, W. V. O.* (2016). From Stimulus to Science. Moscow. 192 p. (in Russian)

6. *Kuznetsov, A.V., Muravitsky, A.* (1980). Provability as Modality. In: Popovich M.V. (Ed.). Actual Problems of Logic and Methodology of Science (pp. 193–230). Kiev: «Naukova Dumka». (in Russian)

7. *Lamberov, L. D.* (2011). Reference, de re and explanation of actions. Herald of Ural State University. Series 3: Social Sciences. Vol. 88. No. 1. pp. 5-13.
8. *Lobovikov, V. O.* (2007). Mathematical Ethics, Metaphysics, and the Natural Law (Algebra of Metaphysics as Algebra of Formal Axiology). Yekaterinburg: Ural Branch of the Russian Academy of Sciences. 409 p.
9. *Lukasiewicz, J.* (1959). Aristotle's Syllogistic from the Standpoint of Modern Formal Logic. Moscow. 1959. 312 p. (In Russian)
10. *Markin, V. I.* (2018). De re – de dicto dichotomy and apodeictic syllogistic. *Logical Investigations*. Vol. 24, No. 2, pp. 108–115. (in Russian)
11. *Martin-Löf, P.* (1975). Notes on constructive mathematics. Moscow. (in Russian)
12. *Nagornyj N. M.* (2010). Constructive Process. *New Philosophy Encyclopedia: In 4 volumes*. V. II. p. 295. (in Russian)
13. *Nepejvoda N. N.* (2010). Constructive Direction (in Mathematics and Logic). *New Philosophy Encyclopedia: In 4 volumes*. V. II. pp. 293-294. (in Russian)
14. *Nepejvoda N. N.* (2011). Constructive Mathematics: Overview of Achievements, Shortcomings, and Instructive Cases. *Part I. Logical Investigations*. V. 17, pp. 191–239. (in Russian)
15. *Nepejvoda N. N.* (2012). Constructive Mathematics: Overview of Achievements, Shortcomings, and Instructive Cases. *Part II. Logical Investigations*. V. 18, pp. 157–181. (in Russian)
16. *Nepejvoda N. N.* (2014). Constructive Mathematics: Overview of Achievements, Shortcomings, and Instructive Cases. Part III. *Logical Investigations*. V. 20. pp. 110-148. (in Russian)
17. *Khlebalin, A. V.* (2003). Essentialism and Anti-Essentialism in Modal Logic. // *Science Philosophy*. Vol. 17. No. 2, pp. 34-46. (in Russian)
18. *Tselishchev, V. V.* (1978). The Notion of Object in Modal Logic [Ponyatie ob"ekta v modal'noj logike]. Novosibirsk: Nauka. 174 p. (In Russian)
19. *Béziau, J.-Y.* New Light on the square of oppositions and the nameless coner. // *Logical Investigations*. 2003. № 10. 218-232.
20. *Béziau, J.-Y.* The New Rising of the Square of Opposition. // *Around and Beyond the Square of Opposition*, Basel, Birkhäuser, 2012 a. P. 3-19.
21. *Béziau, J.-Y.* The Power of the Haxagon // *Logica Universalis*, 2012b. № 6, double issue 1-2, P. 1-43.
22. *Béziau, J.-Y.* The Logical Hexagon of Analogy: Structuring the Relations between Difference, Identity and Similarity. // *Handbook of the First World Congress on Analogy*, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico, November 4-6, 2015, Poznań, Poland, Publishing House KONTEXT, 2015. P. 12-13.

23. *Béziau, J.-Y., Buchsbaum, A., Rey, Ch.* (Eds.). Handbook of the 6<sup>th</sup> World Congress and School on Universal Logic (June 16–26, 2018). Vichy, France: Université Clermont Auvergne. 2018.

24. *Blanché, R.* Structures intellectuelles. Essai sur l'organisation systématique des concepts. Paris, Vrin. 1966.

25. *Boolos, G., and Sambin, G.* Provability: the emergence of a mathematical modality. // *Studia Logica*. 1991. 50(1): 1–23.

26. *Burge, T.* *De re* Belief. // *Journal of Philosophy*. 1977. № 74. P. 338–362.

27. *Chisholm, R.* Knowledge and Belief: 'De Dicto' and 'De Re'. // *Philosophical Studies*. 1976. № 29 (1). P.1–20.

28. *Egan, A., Titelbaum, M. G.* Self-Locating Beliefs. // *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). URL = <<https://plato.stanford.edu/archives/win2022/entries/self-locating-beliefs/>>.

29. *Geretto, M.* "I Think That in the Universe Nothing Is Truer Than Happiness, and Nothing Happier or Sweeter Than Truth". On the Relationship between Happiness, Truth and Generosity in Leibniz's Philosophy. // "Für unser Glück oder das Glück anderer" Vorträge des X Internationalen Leibniz-Kongress / Hannover, 18. – 23 Juli. 2016. Band I. (Hildesheim; Zürich; New York: Georg Olms Verlag. P.617-632.

30. *Inwagen, P. van, Meghan Sullivan, and Sara Bernstein.* Metaphysics. // *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), URL = <<https://plato.stanford.edu/archives/sum2023/entries/metaphysics/>>.

31. *Jeshion, R.* 'Acquaintanceless *de re* belief'. // *Meaning and Truth: Investigations in philosophical semantics*. New York: Seven Bridges Press, 2002. P.53–78.

32. *Kaplan, D.* Quantifying // *Synthese*. 1968. № 19. P. 178–214.

33. *Keshet, E.* Split intensionality: A new scope theory of *de re* and *de dicto*. // *Linguistics and Philosophy*. 2010. № 33. P. 251–283.

34. *Keshet, E., and Schwarz, F.* *De Re - De\_Dicto*. [Online document]. Available at the site: [www.florianschwarz.net](http://www.florianschwarz.net) (Accessed on: April 09, 2024).

35. *Knuuttila, S.* Medieval Theories of Modality. // *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), URL = <<https://plato.stanford.edu/archives/sum2021/entries/modality-medieval/>>.

36. *Kneale, W.* (Modality *de Dicto* and *de Re*. // *Logic, Methodology and Philosophy of Science*. 1962. Stanford University Press, Stanford. P. 622-633.

37. *Leibniz, G. W.* *Philosophical Essays*. Indianapolis & Cambridge: Hackett Publishing Co. Inc. 1989.

38. Lagerlund, H. (2021). "Medieval Theories of the Syllogism", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2022/entries/medieval-syllogism/>>.

39. Lewis, D. Attitudes De Dicto and De Se. // *Philosophical Review*. 1979. № 88(4). P. 513–543.
40. Lobovikov, V. O. Artificial Intelligence and an almost Unknown Aspect of Mathematical Linguistics. // *Discourse-P*. 2022. № 19(3). P. 170–184.
41. Lobovikov, V. O. Hilbert's Conception of "Existence-in-Mathematics" and Modelling it by a Formal Axiomatic Theory  $\Phi + \exists$  Treating Existence not as the Quantifier but as a Modality. // *Respublica Literaria*. 2024. Vol. 5. №. 1. P. 16–50.
42. McKay, T. Representing *de re* beliefs // *Linguistics and Philosophy*. 1991, № 14. P. 711–739.
43. Nelson, M. Propositional attitude reports. // *The Stanford Encyclopedia of Philosophy*. (Spring edition 2019). <https://plato.stanford.edu/archives/spr2019/entries/prop-attitude-reports/>
44. Nelson, M. (2023). The De Re/De Dicto Distinction. Supplement to Propositional Attitude Reports // *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), URL = <<https://plato.stanford.edu/archives/spr2023/entries/prop-attitude-reports/>>.
45. Parsons, T. *Articulating Medieval Logic*. Oxford: Oxford University Press. 2014. p.
46. Prior, A.N. Modality de Dicto and Modality de Re. // *Theoria*. 1952. № 18(3), P. 174–180.
47. Quine, W. Quantifiers and propositional attitudes. // *Journal of Philosophy*. 1956. № 53(5). P.177–187
48. Reimer, M. A defense of *de re* belief reports. // *Mind and Language*. 1995, №10. P. 446–463.
49. Roca-Royes, S. Conceivability and De Re Modal Knowledg. // *Noûs*, 2011. № 45. P. 22–49.
50. Salmon, N. (1997). Is *de re* belief reducible to *de dicto*? // *Meaning and Reference* (The Canadian Journal of Philosophy: Supplementary. 1997. Vol. 23. P. 85–110.
51. Schwitzgebel, E. Belief. // *The Stanford Encyclopedia of Philosophy* (Spring 2024 Edition), URL = <<https://plato.stanford.edu/archives/spr2024/entries/belief/>>.
52. Sosa, E. (1970). Propositional attitudes *de dicto* and *de re*. // *Journal of Philosophy*.1970. №. 67. P. 883–896.
53. Stanley, J. Modality and What is Said. // *Noûs*, 2002. № 36(s16). P. 321–344.
54. Taylor, K., 'De re and de dicto: Against the conventional wisdom'. // *Philosophical Perspectives*, 2002. № 16. P. 225–265.
55. Turner, J. 2010, Fitting Attitudes *De Dicto* and *De Se*. // *Noûs*, 2010. № 44(1). P. 1–9.

56. *Wierenga, E.* Omniscience. // The Stanford Encyclopedia of Philosophy (Summer 2021 Edition) URL = <<https://plato.stanford.edu/archives/sum2021/entries/omniscience/>>.

57. *Williamson, T.* Modal Logic as Metaphysics. Oxford: Oxford University Press. 2013.

58. *Wolenski, J.* Truth as Modality. Rzeszow, Poland, University of Information, Technology and Management, 2016.

59. *Wright, G. H.* (1996). Truth-Logics. // Acta Philosophica Fennica. Georg Henrik von Wright. Six Essays in Philosophical Logic. 1996. Vol. 60. P. 71-91.

60. *Zhang, Y., Davidson, K.* De re is less favorable than de dicto – An Experimental Investigation of Belief Report. P. 1-21. PDF [Online]. Available at: <https://scholar.harvard.edu/files/files/plc2020> (Accessed: April 11, 2024).

### **Information about the author**

*Vladimir Olegovich Lobovikov*, Doctor of Philosophical Sciences, Professor, Principal Researcher, Institute of Philosophy and Law of the Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia, ORCID: <https://orcid.org/0000-0001-8760-0452>.

e-mail: [vlobovikov@mail.ru](mailto:vlobovikov@mail.ru)

### **Информация об авторе**

*Владимир Олегович Лобовиков*, доктор философских наук, профессор, главный научный сотрудник, Институт философии и права Уральского отделения Российской академии наук, г. Екатеринбург, Россия, ORCID: <https://orcid.org/0000-0001-8760-0452>.

e-mail: [vlobovikov@mail.ru](mailto:vlobovikov@mail.ru)

Дата поступления 10.04.2025

Принята к печати 11.12.2025

УДК: 81'35:004.8 : 161.7 : 004.895

DOI: 10.15372/PS20250503

EDN: VVGZDP

**П.Н. Барышников**

## **СЕМАНТИКА ИНТЕНСИОНАЛЬНЫХ КОНТЕКСТОВ И ГЕНЕРАТИВНЫЙ ИИ<sup>1</sup>**

В данной работе рассматривается проблема семантики интенциональных контекстов, в особом ракурсе, связанном с принципами работы генеративного искусственного интеллекта (ИИ). Интенциональные контексты, такие как верования, желания, знания и убеждения представляют особую сложность для современных языковых моделей, поскольку требуют учета значений выражений в логических возможных мирах или когнитивных состояниях субъектов. Данная работы – это своеобразный эскиз к постановке проблемы, у которой есть инженерное и философское измерение. Здесь анализируются механизмы работы трансформеров которые используют контекстуальные эмбединги для моделирования значений слов через self-attention. Выявлено, что современные языковые модели способны эффективно обрабатывать анафорические зависимости и контекстуальные связи, но при этом они сталкиваются с ограничениями при интерпретации интенциональных конструкций. Особое внимание уделяется экспериментам с векторными представлениями объектов в многомерном пространстве, в ходе которых при различении субъективных верований и объективной реальности возникают затруднения. Характер затруднений указывает на то, что работа с интенциональными контекстами требует не только простого анализа вероятностных связей между словами, но и более глубокого понимания семантики языковых выражений.

*Ключевые слова:* интенциональный контекст, генеративный ИИ, эмбединг, семантика возможных миров.

---

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда № 24-28-00540, <https://rscf.ru/project/24-28-00540/>

**P.N. Baryshnikov**

## **SEMANTICS OF INTENSIONAL CONTEXTS AND GENERATIVE AI<sup>1</sup>**

This paper examines the problem of semantics of intensional contexts from a special perspective related to the principles of operation of generative artificial intelligence (AI). Intensional contexts such as beliefs, desires, knowledge and convictions pose a particular challenge for modern language models, since they require taking into account the meanings of expressions in logical possible worlds or cognitive states of subjects. This paper is a kind of sketch for the formulation of a problem that has an engineering and philosophical dimension. Here, we analyze the mechanisms of operation of transformers that use contextual embeddings to model the meanings of words through self-attention. It was revealed that modern language models are able to effectively process anaphoric dependencies and contextual connections, but they face limitations when interpreting intensional constructions. Particular attention is paid to experiments with vector representations of objects in multidimensional space, during which difficulties arise when distinguishing between subjective beliefs and objective reality. The nature of the difficulties indicates that working with intensional contexts requires not only a simple analysis of probabilistic connections between words, but also a deeper understanding of the semantics of linguistic expressions.

*Keywords:* intensional context, generative AI, embedding, possible worlds semantics.

### **1. Источник вдохновения**

Тема, вынесенная в заголовок данной работы, в силу относительной «молодости» технологий генеративного искусственного интеллекта, достаточно нова. Пока трудно обнаружить полноценного набора авторитетных источников по этой проблеме. Скорее всего такая ситуация сложилась потому, что, с одной стороны, архитекторы трансформеров не видят пользы в традиционных вопросах формальной семантики. С другой стороны, традиционные для логики и философии семантические вопросы очень уж видо-

---

<sup>1</sup> The research was supported by the grant of the Russian Science Foundation No. 24-28-00540, <https://rscf.ru/project/24-28-00540/>

изменяются при перенесении их в поле нейросетевых технологий. [1] Модели на основе трансформеров способны частично улавливать интенциональные контексты через контекстуализированные эмбединги, под которыми понимается числовое векторное представление информационных объектов. Но совместимость формальных аппаратов недостаточно проанализирована. Представленный текст можно рассматривать как эскиз к более широкому исследованию.

Идея написать тезисы по проблеме интенциональных контекстов в современных методах NLP (Natural Language Processing), в частности при использовании генеративных трансформеров, возникла после прочтения обзорной работы П. Куслия и Е.Востриковой [2]. В указанной работе представлена широкая экспозиция проблемы интенциональных контекстов, с выделением механизмов семантических ограничений, определяющих совместимость различных классов интенциональных глаголов с декларативными и вопросительными придаточными предложениями. Мы видим, что традиционные синтаксические объяснения несовместимости некоторых глаголов с вложенными вопросами являются недостаточными, поскольку грамматическая корректность таких конструкций определяется их семантикой, включая пресуппозиции, модальность и функциональные особенности отрицания. Важным и эффективным теоретическим инструментом является концепция L-аналитичности. Основная идея этого подхода состоит в том, что некоторые выражения остаются грамматически некорректными не из-за их синтаксической структуры, а потому, что они семантически тривиальны (либо тавтологичны, либо противоречивы) в силу особенностей их функциональных элементов. Это ограничение исходит из свойств логической структуры предложения, а не его лексического наполнения. Концепция L-аналитичности близка к идеям Куайна и его критике аналитичности, [4] но фокусируется на формальных ограничениях грамматики. Например, в некоторых работах показывается, что естественный язык исключает не просто аналитические истины (как в логическом позитивизме), а L-аналитичные структуры, которые неизбежно приводят к бессмысленности. [3] Из этого можно заключить, что грамматическая корректность интенцио-

нальных конструкций не сводится к синтаксическим ограничениям, а определяется более глубокими семантическими механизмами, связанными с пропозициональными установками, пресуппозициями и взаимодействием отрицания с кванторными и модальными элементами.

Интенциональные контексты представляют собой нетривиальную задачу для генеративного искусственного интеллекта (ИИ), поскольку они требуют учета значений выражений в разных возможных мирах или контекстах, а не только в одном фиксированном значении. Эта инженерная задача позволяет сформулировать философскую проблему:

*Какой формальный механизм позволит генеративному искусственному интеллекту адекватно моделировать семантику интенциональных контекстов, учитывая их зависимость от возможных миров и когнитивных состояний субъектов?*

Для обеспечения корректной обработки пропозициональных установок, модальных конструкций и интенциональных выражений в условиях неопределенности контекста эта проблема требует объединения моделей представления знания (например, логики возможных миров, семантических графов, вероятностных моделей) с принципами работы генеративных моделей. Какие есть варианты решения? Рассмотрим некоторые принципы анализа контекстуального окружения современными LLM.

## 2. Моделирование контекста

ИИ может использовать контекстуальные подсказки для уточнения значений выражений. В LLM (Large Language Models – Большие языковые модели) слова интерпретируются на основе их окружения в тексте, что позволяет моделировать различие между буквальным значением и значением в определенном контексте (например, в гипотетическом или возможном мире). При анализе интенциональных контекстов, таких как «думать», «верить» или «желать», модель ориентируется на слова, указывающие на ментальные состояния или гипотетические ситуации. При этом языковые модели не имеют встроенного понятия о «возможных мирах». Они могут только подстраиваться под вероятностные связи в структурах данных.

Перед тем, как переходить к примерам со сложными интенциональными контекстами, рассмотрим два примера с анафорическими зависимостями:

1. *Чем больше войско у генерала, тем ему труднее.*

2. *Генералы убеждены, что чем больше в войске солдат, тем им труднее.*

Современные NLP-модели (BERT, GPT, T5) работают с self-attention, определяя вероятностную связь между словами. Self-attention представляет собой механизм в архитектурах нейронных сетей, который позволяет модели определять и учитывать взаимосвязи между различными частями входного текста. Такой механизм позволяет каждому слову (или токену) обращаться ко всем другим словам в предложении, вычисляя их «влияние» или «важность» для понимания данного слова<sup>1</sup>. Формула для трансформеров выглядит так:

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V$$

где:

- Q(Query), K(Key), V(Value) – векторы слов, вычисленные на основе контекстного представления.
- $d_k$  – размерность эмбедингов (используется для масштабирования).

BERT обучается на предсказании замаскированных слов и следующего предложения. Важно понимать, как модель определяет референт местоимения «ему»? В первую очередь создаются векторные представления слов (Word Embeddings):

Таблица 1

Пример (1)	Пример (2)
<ul style="list-style-type: none"> <li>• "генерал" <math>\rightarrow v_g</math></li> <li>• "войско" <math>\rightarrow v_w</math></li> <li>• "ему" <math>\rightarrow v_p</math></li> </ul>	<ul style="list-style-type: none"> <li>• "генералы" <math>\rightarrow v_g</math></li> <li>• "солдаты" <math>\rightarrow v_w</math></li> <li>• "им" <math>\rightarrow v_p</math></li> </ul>

<sup>1</sup> Отметим, что некоторые модели, например, BERT используют маскированное языковое моделирование (MLM). В отличие от традиционных языковых моделей (GPT), которые обычно прогнозируют следующее слово в последовательности, маскированное языковое моделирование позволяет обучать модель на основе всей доступной информации из контекста, а не только из прошлого. Из-за этого результаты анализа в разных моделях могут значительно расходиться. Но для данного этапа постановки проблемы это имеет значение, которым можно пренебречь.

Далее определяется вероятностная связь через self-attention: для слова «ему» модель оценивает связь с «генералом» и «войском». Вероятность связи вычисляется как  $Q_p \cdot K_g$  для генерала) и  $Q_p \cdot K_w$  (для войска). Далее подсчитывается связь через Attention Score, и на основании этого делается вывод.

Таблица 2.

Пример	Слово	Связь с «ему» ( Attention Score)
(1)	генерал	0,85
(1)	войско	0,15
Пример	Слово	Связь с «им» ( Attention Score)
(2)	генералы	0,23
(2)	солдаты	0,77

Сразу отметим, что результаты подсчёта «Внимания» на разных моделях могут быть разными. Приведённые выше примеры – результат обработки запросов на ChatGPT4o. Недавно вышедшая китайская модель Qwen2.5-Plus в обоих примерах присвоила словам «генерал» и «генералы» индекс «Счёта внимания» 0,6 относительно слов «ему» и «им» соответственно. Большую роль играют данные, на которых обучалась модель. Вероятностные расстояния между словами, рассчитываются именно на основе зависимостей в обучающем наборе данных. Собственно, существует несколько аспектов, которые влияют на интерпретацию LLM предложений типа (2). Модель распознаёт:

1. **Синтаксический аспект:** «Генералы убеждены» создает интенциональный контекст, в котором мнение (убежденность) принадлежит генералам.

2. **Семантический аспект:** В некоторых контекстах можно интерпретировать, что «им труднее» относится к генералам, так как они несут больше ответственности при увеличении численности войска.

3. **Аспект обучения:** Если обучающие данные содержали больше примеров, где дательный падеж «им» чаще соотносится с подлежащим («генералы»), модель будет склонна делать аналогичное предсказание.

4. **Информационно-дистрибутивный аспект:** В обучающем корпусе модель, возможно, чаще встречала конструкции, где «генералы» и «им» тесно связаны.

5. **Аспект референциальной близости:** Трансформеры иногда предпочитают связывать местоимения с ближайшим возможным референтом в рамках одной смысловой структуры.

### 3. Возможные миры в многомерном векторном пространстве<sup>1</sup>

Возьмем более сложный пример:

- (1) *Алиса верит, что единорог живет в ее саду, но не знает, что это тот же самый единорог, о котором говорит Боб.*

В стандартной экстенциональной логике, если два языковых выражения указывают на один и тот же объект, то они взаимозаменяемы без потери смысла всего высказывания. В высказывании (3) утверждается, что единорог в саду Алисы является тем же объектом, о котором говорит Боб, но Алиса об этом не знает. ИИ, выстраивая векторные предсказания последовательностей токенов, может ошибочно предположить, что Алиса знает, что объем понятия «единорог, живущий в саду Алисы» и «единорог, о котором говорит Боб» равнозначны. Генеративный ИИ, обученный на корпусах текстов, может не учитывать ограниченную перспективу субъекта, что может привести к ошибочному выводу. Если ИИ использует векторное представление значений, то «единорог Алисы» и «единорог Боба» могут быть слишком близки в многомерном пространстве, что приведет к ложному объединению понятий.

Проведем эксперимент. У нас есть два объекта: «Единорог Алисы» – объект, который согласно вере Алисы, живёт в её саду, и «Единорог Боба» – объект, о котором говорит Боб. Каждый из этих объектов может быть представлен как вектор в многомерном пространстве, основанный на контексте их упоминания. Для большей точности мы будем проводить измерение косинусной близости между «единорогом Алисы» и «единорогом Боба» на предобученной англоязычной версии модели BERT на примере:

---

<sup>1</sup> Эксперименты проводились на модели <https://chat.qwenlm.ai/> [5]

(2) *Alice believes that a unicorn lives in her garden, but she does not know that it is the same unicorn that Bob talks about.*<sup>1</sup>

Код позволяет вычислить семантическое расстояние между двумя упоминаниями слова «unicorn» в предложении с использованием BERT. Предложение токенизируется, преобразуется в векторы с помощью модели, и для каждого «unicorn» извлекаются соответствующие векторы. Затем через косинусную близость оценивается их сходство, что позволяет количественно проанализировать семантическую связь этих упоминаний в интенциональном контексте.

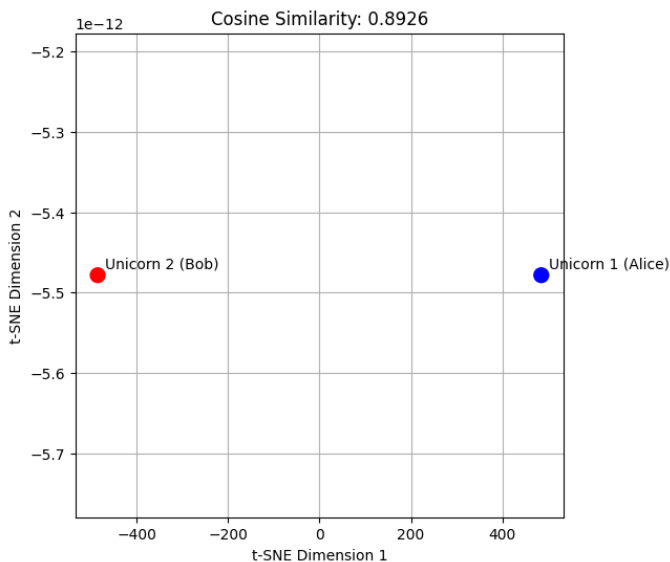


Рис. 1. Косинусная близость Unicorn 1(Alice) и Unicorn 2 (Bob)

Косинусная близость  $\approx 0.89$ : Векторы «единорог Алисы» и «единорог Боба» в англоязычной версии достаточно близки семантически, что может отражать их потенциальную идентичность. Однако, для языковой модели логические интенциональные ограничения остаются недоступными. Без явной аннотации

<sup>1</sup> <https://colab.research.google.com/drive/15LNhUs89YDQxhn5C1-GEPwoh5R2Y5ZkW#scrollTo=euh8psa0QxF9&line=42&uniqifier=1>

или заданной онтологии модели не могут предсказать, что интенциональный контекст потребует учета восприятия или убеждений субъекта, а не объективного состояния дел.

#### 4. Убеждения Алисы и идентичность объектов

Для формализации примера (3) определим термины:

- $w$ : Возможный мир.
- $E(w)$ : Единорог в мире  $w$ .
- $G(x)$ :  $x$  живет в саду Алисы.
- $B(x)$ :  $x$  – единорог, о котором говорит Боб.
- $K_A(\phi)$ : Алиса знает, что  $\phi$  истинно.
- $V_A(\phi)$ : Алиса верит, что  $\phi$  истинно.

Тогда предложение можно разбить на две части:

1.  $V_A(\Box x[E(x)\Box G(x)])$  – Алиса верит, что существует единорог  $x$ , который живет в её саду.
2.  $\neg K_A((\Box x\Box y[(E(x)\Box G(x)) \leftrightarrow (E(y)\Box B(y))])$  Алиса не знает, что единорог в её саду идентичен единорогу, о котором говорит Боб.

При такой формализации выявляется три проблемы. Во-первых, интенциональность веры подразумевает возможность несоответствия между верованиями субъекта и объективной реальностью. Во-вторых, определение идентичности объектов в интенциональных контекстах требует специальной семантики, поскольку субъект может не осознавать идентичность двух объектов даже при их фактическом совпадении. В-третьих, различие между знанием и верой заключается в том, что знание предполагает уверенность в истинности утверждения, тогда как вера может быть ошибочной. Таким образом, Алиса может верить в существование единорога, но не иметь достоверного знания о его связи с единорогом Боба.

Как мы показали в эксперименте, LLM сталкиваются с трудностями при установлении идентичности объектов и при анализе анафорических зависимостей, поскольку они не имеют доступа к формальной логике возможных миров или другим механизмам контекстуального анализа. Вместо этого модели полагаются на контекстуальные эмбединги, которые могут быть недостаточно точными для решения таких задач. Отметим, что запрос на созда-

ние эмпатичного и контекстуально чувствительного ИИ существует давно. [6] Способность извлекать смысл из контекста вместо простого распознавания паттернов данных позволило бы системам быть более гибкими и адаптивными. Распознавание представлений, намерений, убеждений и прочих ментальных состояний в ходе принятия решений также открыло бы новые возможности для ИИ, улучшая взаимодействие между человеком и машиной.

Итак, Большие языковые модели основываются на статистических закономерностях, извлеченных из больших корпусов текста. Они генерируют ответы, опираясь на вероятностные связи между словами, но не обладают истинным пониманием содержания. Это становится критичным в случае интенциональных контекстов, где значение предложения зависит не только от составляющих его слов, но и от ментальных состояний субъекта (например, знания, веры, желаний). Такие утверждения как (3) и (4) требуют интерпретации веры Алисы как модального оператора, который может не соответствовать положению дел в реальном мире. LLM могут воспроизводить подобные конструкции, однако их способность различать истинное знание и ошибочную веру остается ограниченной.

## Литература

1. *Havlik V.* Meaning and understanding in large language models. // *Synthese*, 2024. Vol. 205, No. 1. P. 9-29.
2. *Куслий П.С., Вострикова Е.В.* Семантика интенциональных контекстов: современные проблемы и дискуссии. // *Философия науки и семантика.* / Научн. ред. и сост. Р.Э. Бараш, Е.В. Вострикова, П.С. Куслий. М.: Русское общество истории и философии науки – 2020. С. 292-327.
3. *Куайн У.В.О.* С точки зрения логики. 9 логико-философских очерков. – Томск: Изд-во Томского университета. 2003.
4. *Gajewski J.* On analyticity in natural language. 2002. Citation Key: Gajewski2004OnAI. <https://api.semanticscholar.org/CorpusID:85508868>
5. *Team Q.* Qwen2.5 technical report. 2024. arXiv preprint arXiv:2412.15115. <https://arxiv.org/abs/2412.15115>
6. *Bennett M.T., Maruyama Y.* Intensional Artificial Intelligence: From Symbol Emergence to Explainable and Empathetic AI. // *Intensional Artificial Intelligence*. 2021.: arXiv. <https://arxiv.org/abs/2104.11573>

### References

1. Havlík V. (2024) *Meaning and understanding in large language models*. Synthese, Vol. 205, No. 1, P. 9.
2. Kusliy P.S., Vostrikova E.V. (2020) *Semantics of intensional contexts: modern problems and discussions*. Philosophy of science and semantics: monograph / Scientific. ed. and comp. R.E. Barash, E.V. Vostrikova, P.S. Kusliy. – : "Truth. Science. Reason". – М.: Russian Society for the History and Philosophy of Science. – pp. 292-327. (In Russ.)
3. Gajewski J. (2004) *On analyticity in natural language*. Citation Key: Gajewski2004OnAI. <https://api.semanticscholar.org/CorpusID:85508868>
4. Quine W.V.O. (2003) *From the point of view of logic. 9 logical-philosophical essays: Library of analytical philosophy*. – Tomsk: Tomsk University Press. – 166 p. (In Russ.)
5. Team Q. (2024) *Qwen2.5 technical report*. arXiv preprint arXiv:2412.15115. <https://arxiv.org/abs/2412.15115>
6. Bennett M.T., Maruyama Y. (2021) *Intensional Artificial Intelligence: From Symbol Emergence to Explainable and Empathetic AI*. Intensional Artificial Intelligence. – arXiv. <https://arxiv.org/abs/2104.11573>

### Информация об авторе

*Барышников Павел Николаевич* – д. филос. н., доцент, профессор кафедры исторических, социально-философских дисциплин, востоковедения и теологии Пятигорского государственного университета.  
[pnbaryshnikov@pgu.ru](mailto:pnbaryshnikov@pgu.ru)

### Information about the authors

*Baryshnikov Pavel N.* – Doctor of Philosophy, Associate Professor, Professor of the Department of Historical, Social and Philosophical Disciplines, Oriental Studies and Theology of Pyatigorsk State University.  
[pnbaryshnikov@pgu.ru](mailto:pnbaryshnikov@pgu.ru)

Дата поступления 12.03.2025  
Принята к печати 11.12.2025

УДК: 165.42

DOI: 10.15372/PS20250504

EDN: VXXABE

**Е.Н. Лисанюк**

### **ГАРРИ И ПИНГВИНЫ В ФОРМАЛЬНОЙ ТЕОРИИ АРГУМЕНТАЦИИ**

В статье функции понятия спора в формальной теории аргументации проиллюстрированы при помощи двух хорошо известных примеров из теории аргументации: примера о Гарри с Бермудских островов, использованного С. Тулмином, и примера о пингвине Твити, которая не летает, хотя и птица, используемого в немонотонной логике. Путем реконструкции этих примеров обосновывается важность понятия спора в логике аргументации для выработки путей решения задачи поиска и отбора наиболее сильных аргументов, с которыми можно отождествлять исходы аргументации между сторонами. Решение этой задачи является объединением усилий двух ведущих направлений современной формальной теории аргументации – логики аргументации и диалектической теории аргументации.

*Ключевые слова:* логика аргументации, диалектика, разрешение спора, немонотонная логика, диалог.

**E. N. Lisanyuk**

### **HARRY AND PENGUINS IN FORMAL THEORY OF ARGUMENTATION**

We illustrate the functions of the concept of dispute in the formal theory of argumentation with the help of two well-known examples from the theory of argumentation: the example of Harry from Bermuda, used

by S. Toulmin, and the example of Tweety the penguin, who does not fly, although it is a bird, used in non-monotonic logic. Reconstruction of these examples allows us to show the importance of the concept of dispute in the logic of argumentation for developing ways to solve the task of finding and selecting the strongest arguments with which the outcomes of argumentation between the parties can be identified. The solution of this task is a unification of the efforts of two leading areas of modern formal theory of argumentation - the argumentation logic and the dialectical theory of argumentation.

*Keywords:* argumentation logic, dialectic, dispute resolution, non-monotonic logic, dialog.

В статье очерчиваются функции понятия спора в формальной теории аргументации и обосновывается положение о его необходимости в целях выработки пути решения задачи поиска и отбора наиболее сильных аргументов, с которыми можно отождествлять исходы аргументации между сторонами. Эту же задачу можно описать не только в терминах одной из прикладных задач формальной теории аргументации, как это сделано выше, но и в терминах собственно теории – как объединение усилий двух ее ведущих направлений, логической и диалогической (теорий) аргументации. Первое из них в последнее время получило название логики ориентированной формальной аргументации или логики аргументации. Логика аргументации нацелена на изучение формальной выводимости в цепочках многосортных умозаключений, порождаемых в условиях нехватки или несовершенства информации, когда она неполная или противоречивая.

Главная особенность современной логики аргументации, отличающая ее от того, что под логической теорией аргументации понимали еще в середине XX в., заключается в особенности ее предмета исследования - стремлении найти минимально приемлемые заключения, которые можно вывести из множества посылок, могущих оказаться несовместимыми или недостаточными для корректного выведения заключения. Помимо формального взгляда на вывод, трактуемый как переход от посылок к заключению, с прежней логической теорией аргументации ее роднят формальные способы решения задачи отбрасывания неприемлемых аргументов – логически некорректных умозаключений. Вме-

сте с тем класс приемлемых аргументов в современной логике аргументации оказывается более широким за счет того, что базовым отношением может быть не только выражаемая отношением следования поддержка заключения посылками, но также критика, как в абстрактном подходе к аргументации Пхана Миня Дунга – магистральной линии в современной логике аргументации (Dung 1995), которой мы коснемся ниже, и невыводимость (Зайцев 2024). Таким образом, две особенности логики аргументации – это идея выведения заключения из множества посылок, не обязательно полного и consistenteного, и, возможно, бесконечного, а также расширение спектра базовых отношений, фундирующих формальную выводимость.

Эти две особенности не исключают интерпретации подобных множеств как наборов предъявляемых в диалоге аргументов, а отношений – как дедуктивных или неформальных способов демонстрации в них. Они созвучны идеям Хаима Перельмана, одного из родоначальников второго направления, высказанным им накануне возникновения логики аргументации в 1959 г. в их совместном с Л. Ольбрехтс-Тытекой трактате «Новая риторика», который является одним из теоретических источников современной диалектической теории аргументации (Perelman, Olbrechts-Tyteca, 2001, p. 22):

«В течение последнего столетия логике выпало блестящее развитие, в ходе которого, отбрасывая свои старые формулы, она устремилась к анализу методов доказательства (proof), эффективно используемого математиками.... Одним из результатов этого развития стало ограничение собственного предмета исследования, ведь все, что математики игнорируют, является чуждым для нее. Полученную таким путем в логике теорию доказательства (demonstration) логики обязаны дополнить теорией аргументации».

Задачу создания такой теории аргументации реализует второе направление, именуемое также диалектическим, человеко- или агентно-ориентированным фокусируется на изучении аргументации в диалоге в качестве инструмента для достижения целей, ради которых в диалогах используют аргументы, включая убеждение, совещание, эристику и т.д. В первых попытках на этом пути ученые стремились формализовать правила диалого-

вых взаимодействий, чтобы выразить семантику логических операторов, таких как отрицание, конъюнкция, импликация и дизъюнкция, трактуя их как правила защиты и критики предложений. Наиболее влиятельными из них стала формальная диалектика Чарльза Хэмблина (Hamblin, 1971), формализация вопросно-ответного диалога Джима Маккензи (Mackenzie 1979) и концепция формального спора Николаса Решера (Rescher 1977). За исключением формально-диалектической концепции Эльзе Барт и Эрика Краббе (Barth, Krabbe, 1982), попытавшихся определить некоторые типы диалогов при помощи понятия расхождения во мнениях, основанного на разграничении между истинностными значениями предложений, выступающих содержаниями точек зрения сторон семантики пропозиций, эти концепции не были логическим исследованием диалогов или споров, несмотря на то что в них и используются эти термины. Это характерно и для разрабатываемых в компьютерных науках концепциях диалоговых протоколов взаимодействий агентов, в которых для решения конкретных задач агенты могут не только выдвигать предложения, но и отзываться их, запрашивать их оценку, изменять ее, и т.д. (Loui 1998), (Brewka 2001). В современных диалоговых протоколах спектр допустимых действий агентов подчинен типам задач - на вычисление, принятие решений и т.д. (Atkinson et. al. 2006), включая их предметную определенность, например, в праве (Hage 1993, Gordon 1994). Это сблизило разработку подобных протоколов и моделирование аргументативных диалогов определенного типа, открыв перспективу исследования стратегии и тактики аргументирования не как инструмента убеждения вообще, но относительно определенных типов диалогов, где она используется (Walton, Krabbe, 1995), (van Eemeren, Grootendorst, 2004). По аналогии с логикой аргументации и в духе античной диалектики это направление следовало было бы назвать диалектикой аргументации, как предлагал Х. Перельман, но этому препятствует устоявшееся, но не связанное с аргументацией в диалоге использование термина «диалектика» в гегелевской и советской философии.

Ключевой характеристикой диалектической теории аргументации, начиная с середины XX в., которая отличает ее современный этап от прежнего этапа, по мнению Хенри Праккена, одного

из ведущих представителей логики аргументации, является разделяемое в сообществе ее исследователей положение о том, что за исключением формальных доказательств теорем в математике, оценка аргументов осуществляется относительно других аргументов в диалоге. Она является производной не от строения, как в логически корректных умозаключениях, а от способности противостоять критике в данном диалоге. Такая оценка может быть формализована логическими или вычислительными методами, и это касается как демонстративных, так и недемонстративных аргументов, в которых истинность посылок поддерживает истинность заключения, но не гарантирует ее (Prakken, 2018, 127). Таким образом, с прежними диалектическими и риторическими теориями аргументации ее современную версию роднит сфокусированность на диалоге - речевом или вербальном взаимодействии рациональных участников, как сподобе или форме аргументации, изучение которых внесло вклад в создание методов и методик оценки аргументов.

Два следующих примера рассуждений, о Гарри с Бермудских островов и о пингвинах, символизируют границу между прежним и современным этапом в развитии теории аргументации в XX в. в диалектической теории аргументации и в логике аргументации, соответственно.

(П1) Птицы летают,

(П2) Твити – птица,

(П3) Значит, птица летает.

Однако:

(П4) Пингвины не летают, хотя они птицы.

(П5) Твити – пингвин.

(П6) Значит, Твити не летает.

Пример о пингвинах сделался мемом сначала в немонотонной логике, а затем и в логике аргументации, после выхода в свет специального выпуска журнала «Искусственный интеллект» (Artificial Intelligence) в 1980 г. (Bobrow, 1980), под редакцией Даниэля Боброва, видного представителя северо-американской научной школы компьютерных исследований, в своей докторской диссертации создавшего первую программу машинной обработки естественного языка на основе логического вывода для решения прикладных задач. На пример о пингвинах ссылаются, когда хо-

тят указать на следующие две важные особенности логики аргументации, касающихся немонотонного вывода. Это отменяемость (правил) вывода вследствие возможных исключений, и вывод по умолчанию, или по первому чтению *prima facie*, когда исключения не учитывают, абстрагируясь от них.

Действительно, нетрудно вывести (П3) из (П1) и (П2) по хорошо известному по модусу ААА (BArbArA) первой фигуры силлогизма в традиционной силлогистике. Его часто иллюстрируют примером умозаключения о том, что Сократ, или в другой версии – Кай, смертен, который можно найти едва ли не в каждом учебнике по логике:

(К1) Люди смертны,

(К2) Кай – человек,

(К3) Значит, Кай – смертен.

Вместе с тем вывести истинность (П3) и (К3) подобным способом возможно, только если по умолчанию считать, что посылки (П1) и (К1) являются исчерпывающими обобщениями - общими, а не частными категорическими предложениями, т. е. не имеют исключений. Наличие такого исключения (П4), взятое вместе с (П5), влечет (П6) и отбрасывает (П3), отменяя вывод (П3) из (П1) и (П2). Развитием идей отменяемого вывода и вывода с умолчанием стало понятие приемлемого аргумента в абстрактном подходе к аргументации Дунга. В соответствии с ним аргумент  $\beta$  называют приемлемым на данном множестве аргументов, упорядоченных бинарным отношением атаки  $F = \langle Arg, attack \rangle$ , если в тех случаях, когда  $\beta$  атакован одним из них –  $attack [\alpha, \beta]$ , то в  $F$  найдется аргумент  $\gamma$ , атакующий  $\alpha$  –  $attack [\gamma, \alpha]$ , и тем самым возвращающий  $\beta$ , отклоненный атакой  $\alpha$  ранее. По умолчанию  $\beta$  считается приемлемым на данном множестве аргументов до тех пор, пока он не атакован одним из них. Вместе с тем ясно, что приемлемость не атакованного и атакованного и возвращенного аргумента, т.е. защищенного, неодинаковая, и первый слабее второго.

(Г1) Родившиеся на Бермудах являются британцами – британскими подданными.

(Г2) Гарри родился на Бермудах.

(Г3) Значит, Гарри - британец.

Однако:

(Г4) Если родители Гарри не были британцами, то Гарри – не британец.

(Г5) Если Гарри в дальнейшем сменил свое британское гражданство на гражданство США, как многие жители Бермуд, то Гарри – не британец.

Пример о Гарри стал вехой в развитии теории аргументации благодаря Стивену Тулмину, использовавшему его в своем теперь уже классическом трактате «Использование аргументации» (*The Use of Argument*), впервые опубликованном в 1958 (Toulmin, 2003, 110). Он символизирует три взаимосвязанных аспекта весомого вклада С. Тулмина в развитие диалектической теории аргументации.

Во-первых, идею о том, что аргумент является не просто рассуждением, выражающим одно или несколько умозаключений, а высказывательной формой рассуждения, в котором каждая из посылок играет свою особую роль, необходимую для обоснования заключения и не сводимую к роли других посылок. Предложение (Г3) играет роль тезиса, или заключения, а посылка (Г1) выражает данные, (Г2) – дает основание и ссылается на поддержку, в этом случае, - на положения закона, а (Г4) и (Г5) – это возражения, выражающие возможные исключения, которые могут препятствовать обоснованию (Г3) при помощи (Г1) и (Г2).

Во-вторых, поскольку возражение против выведения заключения является одним из необходимых структурных элементов аргумента, постольку его отклонение выступает необходимым условием обоснованности заключения. В-третьих, вследствие того, что высказывательная форма аргумента включает отклонение возражения, вывод заключения оказывается отменяемым – и должен быть отменен, если возражение не удалось отклонить. В дальнейшем эти три аспекта позволили сформулировать понятие правдоподобного аргумента как основанного на здравом смысле – на предположении об истинности какого-либо предложения, выступающего посылкой, в условиях отсутствия сведений об обратном, а также методику его проверки при помощи критических вопросов к посылкам или способу демонстрации. Методика проверки обоснованности аргументов при помощи

критических вопросов, подразумевающая взаимодействие сторон диалога и перенос бремени доказательства с одной из них другую (Walton, 1996), развивает идеи вопросно-ответного взаимодействия Д. Маккензи.

При помощи примеров о Гарри и пингвинах покажем продуктивность соединения усилий логики аргументации с диалектической теорией аргументации. Для этого нам понадобится понятие спора и типология споров.

Под спором будем понимать множество аргументов, упорядоченных бинарным отношением атаки  $F = \langle Arg, attack \rangle$ , что подразумевает расхождение во мнениях как необходимое условие спора. Условимся различать следующие три типа споров. В единичном несмешанном ЕН-споре проponent стремится обосновать свое мнение об истинности какого-либо предложения  $A$  перед лицом сомнений аудитории в истинности  $A$ , поэтому для убеждения аудитории достаточно развеять эти сомнения. В терминах логики аргументации для убеждения аудитории в ЕН-споре достаточно, чтобы в множестве аргументов проponentа нашелся хотя бы один приемлемый аргумент. В единичном смешанном ЕС-споре проponent обосновывает истинность  $A$ , а оппонент критикует это обоснование, поэтому убеждения оппонента проponentу необходимо отклонить критические возражения против  $A$ . В терминах логики аргументации для убеждения оппонента в ЕН-споре достаточно, чтобы множество аргументов проponentа было максимальным подмножеством приемлемых и защищенных аргументов (относительно теоретико-множественного включения). В множественном смешанном МС-споре защита точки зрения об истинности  $A$  подразумевает отклонение аргументов в пользу противоположной точки зрения об истинности  $\neg A$ . В МС-споре для защиты любой из них необходимо и достаточно, чтобы подмножество аргументов включало наименьшим подмножество защищенных. Доказательство последних двух утверждений в терминах логики аргументации применительно к этим спорам сводится к тому, чтобы продемонстрировать, соответственно, существование по крайней мере одного максимального подмножества и того факта, что если наименьшее такое подмножество существует, то оно является единственным (Лисанюк, 2015).

Теперь на примерах о Гарри и пингвинах можно показать, почему использование понятия спора в логике аргументации открывает не только новую перспективу объединения двух направлений, но позволяет получить новое знание, если поиск и отбор решений в ЕН-, ЕС- и МС-спора интерпретировать содержательно. Например, ЕН- спор может выражать речь оратора, нацеленную на убеждение своей аудитории, не имеющей определенного мнения об истинности *A*, но и не вполне согласной с этим; ЕС-спор – как рецензирование рукописи, в которой автор отстаивает истинность *A*, а рецензент критически оценивает его доводы в защиту этой точки зрения; и, наконец, МС-спор – как судебные прения между сторонами истца и ответчика, каждая из которых, доказывая справедливость и обоснованность своих требований, стремится опровергнуть справедливость и обоснованность требований другой стороны.

В ЕН-споре оратор изложит оба примера как целостные нарративы, в которых отклонение сомнений аудитории по поводу тезисов о том, что Твити – не летающая птица, а Гарри – британец, включает парирование возражений и отбрасываемые исключения в качестве собственных аргументов. Твити – птица, поэтому, скорее всего, Твити летает, разве что Твити – пингвин. В последнем случае Твити не летает, потому что пингвины не летают, несмотря на то что они птицы. Гарри, скорее всего, британский подданный, ведь он родился на Бермудах. Известно, что родившиеся на Бермудских островах являются британскими подданными, согласно британскому законодательству, разве что Гарри в дальнейшем сменил свое британское подданство на гражданство США, или родители Гарри не были британскими подданными. Отметим, что похожим образом будет выглядеть и нарратив в защиту смертности Кая: Кай смертен, потому что он человек, а все люди смертны, разве что ученые получили эликсир бессмертия и Кай может им воспользоваться, что на сегодняшний день маловероятно.

В ЕС-споре для обоих примеров имеется по два возможных исхода – (П3) и (П6), а также (Г3) и неудача в обосновании (Г3). В примере о пингвинах в отсутствие посылки данных (П5) и основания (П4) можно вывести (П3), однако их наличие отклоняет (П3) и позволяет вывести (П6), разве что далее в споре удастся

отклонить (П5) и (П6), возвращая (П3). Отметим, что если в ЕС-споре допустимо вывести и (П3), воспользовавшись посылками (П1) и (П2), и (П6), отказавшись от них и положившись на (П4) и (П5), то выдвижение в одном и том же споре всех этих посылок будет означать МС-спор, в котором возможно вывести только одно заключение из двух, либо (П3), либо (П6).

В примере о Гарри (Г3) можно обосновать в отсутствие посылок оснований (Г4) и (Г5), но в ситуации их наличия (Г3) будет отклонено. Вместе с тем выдвижение (Г4) и (Г5) не дает оснований для выведения заключения, противоположного (Г3) – о том, что Гарри не является британцем, разве что будут выдвинуты дополнительные посылки данных о том, что родители Гарри в действительности не были британцами, или о том, что Гарри натурализовался в США. Выдвижение одной или обеих подобных посылок будет означать, что это не ЕС, а МС спор, и в нем возможно будет вывести только какое-либо одно из двух заключений, либо (Г3), либо противоположное ему, аналогично тому, как в примере о пингвинах.

### Литература

1. *Зайцев Д.В.* Рассуждения по модулю I. Логика невыводимости // *Логические исследования*. 2024. Т. 30. № 1. С. 11-26.
2. *Zaitsev D.V.* (2024). *Moduloreasoning I. Logic of undeducibility. Logical Investigations*. Vol. 30, No. 1, pp. 11–26. (In Russ.)
3. *Лисанюк Е.Н.* Аргументация и убеждение. СПб. 2015.
4. *Lisanyuk E.N.* (2015). *Argumentation and persuasion*. St Petersburg. (In Russ.)
5. *Atkinson K., Bench-Capon T., McBurney P.* A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. // *Argumentation in Multi-Agent Systems. Lecture Notes in Computer Science*. 2006. Vol.3366. pp. 149-161.
6. *Barth E.M., Krabbe E.C.W.* From Axiom to Dialogue. Walter de Gruyter. 1982. Berlin, New York,
7. *Bobrow D.G.* Artificial Intelligence. 1980. vol. 13. Special issue on Non-Monotonic Logic.
8. *Bondarenko A., Dung P., Kowalski R., Toni F.* An abstract argumentation-theoretical approach to default reasoning. // *Artificial Intelligence*. 1997. № 93. P. 63-101.

9. *Brewka G.* Dynamic argument systems: a formal model of argumentation processes based on situation calculus. // *Journal of Logic and Computation* 2001. № 11. P. 257-282.

10. *Dung P.M.* On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. // *Artificial Intelligence*. 1995. № 77. P. 321-357.

11. *Gordon T.F.* The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*. 1994. № 2. P. 239-292.

12. *Hage J. C.* Monological reason-based logic: a low-level integration of rule-based reasoning and case-based reasoning. // *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. New York, ACM Press. 1993. P. 30-39.

13. *Hamblin C.L.* Fallacies. Methuen, London. 1970.

14. *Eemeren F.H. van, Grootendorst R.* Systematic Theory of Argumentation. Cambridge University Press. 2004.

15. *Loui R.P.* Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*. 1998. № 14. P. 1-38.

16. *Mackenzie J. D.* Question-begging in non-cumulative systems. // *Journal of Philosophical Logic*. 1979. № 8. P. 117-133.

17. *Perelman Ch., Olbrechts-Tyteca L.* The new rhetoric: A treatise on argumentation. New York: Bedford Books. 2001.

18. *Pollock J.L.* Defeasible reasoning. // *Cognitive Science*. 1978. № 11. P. 481-518.

19. *Prakken H.* Historical Overview of Formal Argumentation. // *Handbook of Formal Argumentation*. College Publ. 2018. P. 75-144.

20. *Rescher N.* Dialectics: A Controversy-oriented Approach to the Theory of Knowledge. State University of New York Press, Albany, N.Y. 1977.

21. *Toulmin S.* The Uses of Argument. Cambridge University Press, 2003.

22. *Walton D., Krabbe E.C.W.* Commitment in dialogue: Basic concepts of interpersonal reasoning. Albany, CUNY Press. 1995.

23. *Walton D.N.* Argumentation Schemes for Presumptive Reasoning. Lawrence Erlbaum Associates, Mahwah, NJ. 1996.

## References

1. *Zaitsev D.V.* (2024). Modulareasoning I. Logic of undeducibility. // *Logical Investigations*. Vol. 30, No. 1, pp. 11–26. (In Russ.)

2. *Lisanyuk E.N.* (2015). *Argumentation and persuasion*. St Petersburg. (In Russ.)

3. *Atkinson K., Bench-Capon T., McBurney P.* (2006). A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. Argumentation in Multi-Agent Systems. Lecture Notes in Computer Science. Vol.3366. pp. 149-161.

4. *Barth E.M., Krabbe E.C.W.* (1982). From Axiom to Dialogue. Walter de Gruyter. Berlin, New York,

5. *Bobrow D.G.*, ed. (1980). Artificial Intelligence. vol. 13. Special issue on Non-Monotonic Logic.

6. *Bondarenko A., Dung P., Kowalski R., Toni F.* (1997). An abstract argumentation-theoretical approach to default reasoning. // Artificial Intelligence. no. 93. pp. 63-101.

7. *Brewka G.* (2001). Dynamic argument systems: a formal model of argumentation processes based on situation calculus. // Journal of Logic and Computation. no. 11. pp. 257-282.

8. *Dung P.M.* (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. // Artificial Intelligence. no. 77. pp. 321-357.

9. *Gordon T.F.* (1994). The Pleadings Game: an exercise in computational dialectics. // Artificial Intelligence and Law. no. 2. pp. 239-292.

10. *Hage J. C.* (1993). Monological reason-based logic: a low-level integration of rule-based reasoning and case-based reasoning. Proceedings of the Fourth International Conference on Artificial Intelligence and Law. pp. 30-39, New York, ACM Press.

11. *Hamblin C.L.* (1970). Fallacies. Methuen, London.

12. *Eemeren F.H. van, Grootendorst R.* (2004). A Systematic Theory of Argumentation. Cambridge University Press.

13. *Loui R.P.* (1998). Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*. no. 14. pp. 1-38.

14. *Mackenzie J. D.* (1979). Question-begging in non-cumulative systems. *Journal of Philosophical Logic*. no. 8. pp. 117-133.

15. *Perelman Ch., Olbrechts-Tyteca L.* (2001). The new rhetoric: A treatise on argumentation. (orig. publ. 1959). Bizzell P., B. Herzberg (Eds.) New York: Bedford Books.

16. *Pollock J.L.* (1987). Defeasible reasoning. // Cognitive Science. no. 11. pp. 481-518.

17. *Prakken H.* (2018). Historical Overview of Formal Argumentation. // Handbook of Formal Argumentation. College Publ. pp. 75-144.

18. *Rescher N.* (1977). Dialectics: a Controversy-oriented Approach to the Theory of Knowledge. State University of New York Press, Albany, N.Y.

19. *Toulmin S.* The Uses of Argument. Cambridge University Press, 2003.

20. Walton D., Krabbe E.C.W. (1995). Commitment in dialogue: Basic concepts of interpersonal reasoning. Albany, CUNY press.

21. Walton D.N. (1996). Argumentation Schemes for Presumptive Reasoning. Lawrence Erlbaum Associates, Mahwah, NJ.

### **Информация об авторе**

*Лисанюк Елена Николаевна* – доктор философских наук, профессор, ведущий научный сотрудник, Институт философии Российской академии наук, 109240, г. Москва, ул. Гончарная, д.12, стр. 1.

elenalisanyuk@gmail.com

### **Information about the authors**

*Lisanyuk Elena Nikolaevna, PhD*, Institute of Philosophy, Russian Academy of Science.

elenalisanyuk@gmail.com

Дата поступления 21.03.2025

Принята к печати 11.12.2025

УДК: 164.3

DOI: 10.15372/PS20250505

EDN: RDQKOH

**И.И. Борисова**

**ТАБЛИЧНАЯ ТЕОРИЯ ДОКАЗАТЕЛЬСТВ  
ДЛЯ ГИБРИДНОЙ МОДАЛЬНОЙ ЛОГИКИ  
ПЕРВОГО ПОРЯДКА**

В данной статье излагается язык, семантика и теория доказательств для гибридной модальной логики первого порядка, решающей проблему кросс-мировой квантификации.

*Ключевые слова:* гибридная логика, кросс-мировая квантификация, модальная логика первого порядка, табличная теория доказательств.

**I.I. Borisova**

**A TABLE PROOF THEORY FOR  
FIRST-ORDER HYBRID MODAL LOGIC**

This article describes the language, semantics, and proof theory for first-order hybrid modal logic, which solves the problem of cross-world quantification.

*Key words:* hybrid logic, first-order modal logic, table proof theory.

**Введение**

Одна из задач аналитической философии языка – логическая формализация естественного языка и логическое моделирование рассуждений на естественном языке. Одной из проблем в этой

области является формализация предложений, содержащих *кросс-мировую квантификацию* (см. [3]), например, «*Придет день, когда каждый, кто сегодня лыс, будет поэтом*»<sup>1</sup>. Эту мысль нельзя выразить формулой на языке стандартной модальной логики первого порядка (например, в модальной логике первого порядка как в [2]). Чтобы решить эту проблему требуется некоторым образом модифицировать стандартную модальную логику первого порядка.

Коцурек в [3] показывает, как с проблемой кросс-мировой квантификации справляется гибридная логика  $H^2$ . В данной статье представлен упрощенный вариант гибридной логики  $H$  Коцурека, гибридная логика  $HQ$ , которая также справляется с проблемой кросс-мировой квантификации. В последнем разделе статьи будет представлена табличная теория доказательств для  $HQ$ .

### Постановка проблемы

Рассмотрим предложение с кросс-мировой квантификацией:  
*Придет день, когда каждый, кто сегодня лыс, будет поэтом.* (\*)

Полуформально, истинностные условия, выражаемые в этом предложении, выглядят так: допустим, мы начинаем оценку формулы в мире  $w$ , тогда есть мир  $u$ , достижимый из мира  $w$ , такой что, каждый из мира  $w$ , кто в  $w$  лыс, в мире  $u$  – поэт. Формально<sup>3</sup> это выражается следующим образом:

$$\exists u (wRu \ \& \ \forall e (e \in d(w) \Rightarrow (e \in I(B, w) \Rightarrow e \in I(P, u))))). \quad (1)$$

В модальной логике первого порядка (\*) не формализуется. Продемонстрируем это на двух примерах попыток такой формализации.

<sup>1</sup> Аналогичный пример («*Under certain circumstances, everybody who is rich would have been poor*») можно найти у Вемайера [4].

<sup>2</sup> Общую характеристику гибридных логик можно найти у Браунера [1].

<sup>3</sup> Подробное описание формализма будет представлено ниже. На данный момент ограничусь следующими объяснениями: мы рассматриваем семантику Крипке для модальной логики первого порядка с переменным доменом модели, в которой на отношение достижимости ( $R$ ) не накладывается ограничений. Используются следующие обозначения:  $w, u$  – миры,  $B, P$  – предикаты,  $I$  – интерпретация и для любого  $w$ ,  $d(w)$  – домен  $w$ .

*Попытка 1.*  $\diamond \forall x (Bx \rightarrow Px)$

Полуформальные истинностные условия этой формулы выглядят так: есть возможный мир  $u$ , достижимый из мира  $w$ , такой что каждый из мира  $u$ , кто в  $u$  лыс, в  $u$  – поэт. Формально они выражаются следующим образом:

$$\exists u (wRu \Rightarrow \forall e (e \in d(u) \Rightarrow (e \in I(B, u) \Rightarrow e \in I(P, u))) \quad (2)$$

Как видно, недостаточно просто поставить модальный оператор перед квантором, т.к. мы хотим говорить об объектах из мира  $w$ , а не из достижимого из него мира, как получается в (2).

*Попытка 2.*  $\forall x (Bx \rightarrow \diamond Px)$

Полуформально, истинностные условия, выражаемые в этом предложении, выглядят так: допустим, мы начинаем оценку формулы в мире  $w$ , тогда для любого  $e$  в мире  $w$ , если  $e$  в  $w$  лыс, есть мир  $u$ , достижимый из  $w$ , такой что  $e$  в  $u$  – поэт. Формально это выражается так:

$$\forall e (e \in d(w) \Rightarrow (e \in I(B, w) \Rightarrow \exists u (wRu \& e \in I(P, u)))) \quad (3)$$

Очевидно, что (3) отличается от (1).

Чтобы получить требуемые истинностные условия, нужно в формуле

$$\diamond \forall x (Bx \rightarrow Px)$$

вывести выделенный полужирным курсивом фрагмент из-под действия  $\diamond$ . Это можно сделать с помощью средств гибридной логики первого порядка. Причем, для решения указанной проблемы достаточно гибридной модальной логики первого порядка без равенства, констант, лямбда оператора, термового оператора и с одним видом кванторов. Ниже я опишу такую логику, *HQ*.

## Логика *HQ*

*HQ* – упрощенная версия логики *H* Коцурека.

Язык *HQ* (*L*).

Алфавит *L*. Множество примитивных символов *L* включает:

- счетное множество индивидуальных переменных  $x, y, z \dots$  (*VAR*),

- счетное множество переменных для возможных миров  $s, t, r, \dots$  ( $SVAR$ ),
- для каждого натурального  $n > 0$  счетное множество  $n$ -местных предикатов  $P, Q, \dots$  ( $PRED^n$ ),
- операторы  $\neg, \rightarrow, \forall, \Box$ ,
- сентенциональные операторы  $\downarrow s.$  и  $@_s$  ( $s \in SVAR$ ),
- служебные символы: скобки и запятая.

Гибридную данную логику делают операторы  $\downarrow s.$  и  $@_s$ .

**Определение** (Атомарная формула) Если  $P \in PRED^n, x_1, \dots, x_n \in VAR$ , то  $P(x_1, \dots, x_n)$  – атомарная формула.

**Определение** (Формула). Множество формул, определяется следующим образом:

$$\varphi ::= P(x_1, \dots, x_n) \mid \neg\varphi \mid (\varphi_1 \rightarrow \varphi_2) \mid \forall x \varphi \mid \Box\varphi \mid \downarrow s.\varphi \mid @_s \varphi,$$

где  $P \in PRED^n, x_1, \dots, x_n \in VAR, x \in VAR, s \in SVAR, \varphi, \varphi_1, \varphi_2$  – формулы.

Связанные (свободные) вхождения переменных. Для любой  $x \in VAR$ , вхождение  $x$  называется связанным если оно следует непосредственно за  $\forall$ , т.е. в выражении  $\forall x.$  или находится в области действия  $\forall$ . Для любой  $s \in SVAR$ , вхождение переменной  $s$  называется связанным, если оно следует непосредственно за  $\downarrow$ , т.е. в выражении  $\downarrow s.$  или находится в области действия  $\downarrow s.$ . Для любой переменной  $y \in VAR \cup SVAR$ , вхождение  $y$  называется свободным, если оно не является связанным.

Далее используются следующие конвенции:

- Внешние скобки в формулах не пишутся. Например, выражение  $\varphi \rightarrow (\psi \rightarrow \chi)$  следует читать как формулу  $(\varphi \rightarrow (\psi \rightarrow \chi))$ .
- $\&$  и  $\vee$  имеют приоритет перед  $\rightarrow$ .<sup>1</sup> Например, выражения вида  $\varphi \& \psi \rightarrow \chi$  читаются как  $(\varphi \& \psi) \rightarrow \chi$ .

## Семантика $HQ$

**Определение** (Модель  $HQ$  (модель)). Модель  $HQ$  – это пятерка  $M = \langle G, R, D, d, I \rangle$ , где:

- $G \neq \emptyset$  (множество возможных миров);
- $R \subseteq G^2$  (отношение достижимости);

<sup>1</sup>  $\&, \vee, \diamond, \exists$  определяются стандартным образом.

- $D \neq \emptyset$  (домен модели);
- $d$  (доменная функция) – функция, такая что для любого  $w \in G$ ,  $d(w) \subseteq D$  и  $d(w) \neq \emptyset$ ;
- $I$  (интерпретация) – функция, такая что для любого  $P \in \text{PRED}^n$  и любого  $w \in G$ ,  $I(P, w) \subseteq D^n$ .

Замечу, что:

- В модели переменный домен (не монотонный и не анти-монотонный).
- Домен модели может быть собственным надмножеством объединения доменов миров модели.
- Предикаты интерпретируются на домене модели, а не на доменах миров.

**Определение** (Оценка переменных в модели) Пусть  $M = \langle G, R, D, d, I \rangle$  – модель. Оценка переменных в  $M$  – это функция  $g : \text{VAR} \cup \text{SVAR} \rightarrow D \cup G$ , такая что:

- для каждого  $x \in \text{VAR}$ ,  $g(x) \in D$ ;
- для любого  $s \in \text{SVAR}$ ,  $g(s) \in G$ .

Пусть  $e \in D \cup G$ ,  $g$  – оценка переменных в модели. Тогда  $g_x^e$  – это оценка переменных в  $M$ , такая что для любой переменной  $y$ ,  $g_x^e(y) = g(y)$ , если  $y \neq x$  и  $g_x^e = e$ , если  $y = x$ .

**Определение** (Истина ( $\models$ )) Пусть  $M = \langle G, R, D, d, I \rangle$  – модель,  $w \in G$ ,  $g$  – оценка переменных в  $M$ ,  $\varphi$  и  $\psi$  – формулы,  $P \in \text{PRED}^n$ ,  $x, x_1, \dots, x_n \in \text{VAR}$  и  $s \in \text{SVAR}$ . Тогда:

- $M, w, g \models P(x_1, \dots, x_n) \Leftrightarrow \langle V(x_1), \dots, V(x_n) \rangle \in I(P, w)$ ;
- $M, w, g \models \neg \varphi \Leftrightarrow M, w, g \not\models \varphi$ ;
- $M, w, g \models \varphi \rightarrow \psi \Leftrightarrow (M, w, g \models \varphi \Rightarrow M, w, g \models \psi)$ ;
- $M, w, g \models \Box \varphi \Leftrightarrow \forall u (wRu \rightarrow M, u, g \models \varphi)$ ;
- $M, w, g \models \forall x \varphi \Leftrightarrow \forall e (e \in d(w) \rightarrow M, w, g_x^e \models \varphi)$ ;
- $M, w, g \models \downarrow s. \varphi \Leftrightarrow M, w, g_s^w \models \varphi$ ;
- $M, w, g \models @_s \varphi \Leftrightarrow M, g(s), g \models \varphi$ ;

$\downarrow s.$  связывает вхождение переменной  $s$  и делает мир оценки денотатом  $s$ . При оценке  $@_s \varphi$  оператор  $@_s$  «переносит» нас из мира оценки в мир, который является денотатом  $s$ .

*Пример.* Распишем истинностные условия формулы  $M, w, g \models \downarrow s. \diamond @_s \varphi$ .

$M, w, g \models \downarrow s. \diamond @_s \varphi \Leftrightarrow M, w, g_s^w \models \diamond @_s \varphi$  (на этом шаге мир  $w$  объявляется денотатом переменной  $s$ )

$M, w, g_s^w \models \diamond @_s \varphi \Leftrightarrow \exists u (wRu \ \& \ M, u, g_s^w \models @_s \varphi)$  (на этом шаге обрабатывается оператор возможности и далее мы оцениваем формулу  $@_s \varphi$  в мире, который достигим из  $w$ )

$\exists u (wRu \ \& \ M, u, g_s^w \models @_s \varphi) \Leftrightarrow \exists u (wRu \ \& \ M, g_s^w(s), g_s^w \models \varphi)$  (на этом шаге оператор  $@_s$  «переносит» нас в  $g_s^w(s)$ , т.е. в  $w$ , и мы оцениваем формулу  $\varphi$  в  $w$ )

$\exists u (wRu \ \& \ M, g_s^w(s), g_s^w \models \varphi) \Leftrightarrow \exists u (wRu \ \& \ M, w, g_s^w \models \varphi)$ .

**Определение** (Общезначимость). Формула  $\varphi$  называется общезначимой, если для любой модели  $M = \langle G, R, D, d, I \rangle$ , любой оценки переменных  $g$  в  $M$  и любого  $w \in G, M, w, g \models \varphi$ .

### Решение проблемы

Нам нужно было в формуле

$$\diamond \forall x (Bx \rightarrow Px)$$

вывести выделенный полужирным курсивом фрагмент из-под действия  $\diamond$ , т.е. выразить следующие истинностные условия:

$$\exists u (wRu \ \& \ \forall e (e \in d(w) \rightarrow (e \in I(B, w) \rightarrow e \in I(P, u)))) \quad (1)$$

Искомая формализация (средствами логики  $HQ$ ):

$$\downarrow s. \diamond \downarrow t. @_s \forall x (B(x) \rightarrow @_t P(x)) \quad (4)$$

Ограничусь неформальным обоснованием того, что (4) действительно выражает истинностные условия (1). Мы начинаем оценку в мире  $w$ , и оператор  $\downarrow s$  связывает переменную для возможных миров  $s$  с миром  $w$ . После этого оператор возможности переносит нас в мир  $u$ , а оператор  $\downarrow t$  связывает переменную для возможных миров  $t$  с миром  $u$ . После этого  $@_s$  возвращает нас в мир  $w$  и про все объекты из домена мира  $w$ , мы говорим, что, если они лысы в  $w$ , то они являются поэтами в мире  $u$  (в мир  $u$  нас переносит  $@_t$ ).

### Теория доказательств

В этом разделе будет изложена табличная теория доказательств  $HQ$ .

В деревьях  $HQ$  формулы снабжены префиксами – конечными последовательностями целых положительных чи-

сел (например, 1, 2.3, 14.48.1). Префикс  $\langle x_1, x_2, x_3, \dots, x_n \rangle$  будем записывать как  $x_1, x_2, x_3, \dots, x_n$ . Например,  $\langle 14, 48, 1 \rangle$  будем записывать как 14.48.1.

Язык  $L'$ , на котором будут строиться деревья, расширен относительно языка  $L$  за счет того, что к переменным ( $VAR \cup SVAR$ ) можно добавлять префиксы в квадратных скобках справа, например,  $x[1], s[2.13](x \in VAR, s \in SVAR)$ . Переменные с префиксами в квадратных скобках являются переменными языка  $L'$  наряду с элементами  $VAR$  и  $SVAR$ .

В деревьях  $HQ$  используются отмеченные формулы с префиксами, т.е. формулы с префиксом и с меткой  $+$  или  $-$  (например,  $\sigma P(x) \vee \neg Q(x) +$  – это  $\sigma P(x) \vee \neg Q(x)$  с меткой  $+$ ).

**Определение** (Замкнутая ветка). Ветка дерева называется замкнутой, если она содержит  $\sigma \psi +$  и  $\sigma \psi -$ , где  $\sigma$  – префикс,  $\psi$  – некоторая формула. Незамкнутая ветка называется открытой.

**Определение** (Замкнутая дерево). Замкнутое дерево – дерево, каждая ветка которого замкнута.

**Определение** (Доказательство формулы  $\varphi$  (доказательство)). Доказательство формулы  $\varphi$  – это замкнутое дерево, которое начинается с  $1 \varphi -$ ; последующие формулы добавляются по правилам, которые будут приведены ниже.

**Нотационная конвенция.** Для любой формулы  $\varphi$ , любого  $x \in VAR \cup SVAR$  и любой переменной  $y$  языка  $L'$ ,  $\varphi_x^y$  – результат замены всех свободных вхождений переменной  $x$  вхождениями  $y$  (с переименованием связанных переменных, если  $x$  – переменная, и некоторые свободные вхождения  $x$  в  $\varphi$  лежат в области действия  $\forall y$  или  $\downarrow y$ ).

Правила табличного доказательства

В правила табличного доказательства входят правила для пропозициональной логики, правила для модальных операторов, кванторов и гибридных операторов. Если не указано иное, правило может быть применено к формуле только один раз на ветке.

Правила для  $\rightarrow$

$$\frac{\sigma \psi \rightarrow \chi +}{\sigma \psi - \mid \sigma \chi +} \qquad \frac{\sigma \psi \rightarrow \chi -}{\sigma \psi +, \sigma \chi -}$$

Знак « $\mid$ » между формулами означает разветвление.

Правила для  $\neg$

$$\frac{\sigma \neg \psi +}{\sigma \psi -} \qquad \frac{\sigma \neg \psi -}{\sigma \psi +}$$

Правила для  $\Box$

1) Если префикс  $\sigma.n$  уже встречался на ветви, то

$$\frac{\sigma \Box \psi +}{\sigma.n \psi +}$$

Префикс  $\sigma.n$  должен встречаться на ветке выше. Правило может быть применено к  $\sigma \Box \psi +$  и к  $\sigma \Diamond \varphi -$  многократно – один раз для каждого префикса  $\sigma.n$ , встречающегося на ветке.

2)

$$\frac{\sigma \Box \psi -}{\sigma.n \psi -}$$

где  $\sigma.n$  – новый для ветки префикс.

Правила для  $\forall$

1) Для любой переменной  $y[\sigma]$ ,

$$\frac{\sigma \forall x \psi(x) +}{\sigma \psi(y[\sigma]) +}$$

Правило может быть применено к  $\sigma \forall x \psi(x) +$  и к  $\sigma \exists x \psi(x) -$  более одного раза – один раз с каждой переменной  $y[\sigma]$ , вне зависимости от того встречалась она на ветке или нет.

2)

$$\frac{\sigma \quad \forall x \quad \psi(x) -}{\sigma \quad \psi(y[\sigma]) -}$$

где  $y[\sigma]$  – новая для ветки переменная.

Правила для  $\downarrow$

$$\frac{\sigma \downarrow s. \psi +}{\sigma \psi^{s[\sigma]}_s +} \qquad \frac{\sigma \downarrow s. \psi -}{\sigma \psi^{s[\sigma]}_s -}$$

## Правила для @

$$\frac{\sigma @_{s[\tau]} \psi +}{\tau \psi +} \quad \frac{\sigma @_{s[\tau]} \psi -}{\tau \psi -} \quad \frac{\sigma @_s \psi +}{\delta \psi +} \quad \frac{\sigma @_s \psi -}{\delta \psi -}$$

где  $\delta$  – новый для ветки префикс формы  $\langle n \rangle$ , где  $n$  – натуральное число.

*Пример.* Построим доказательство формулы  $@_t (P(x) \rightarrow \downarrow s. \Box @_s P(x))$

1	$@_t(P(x) \rightarrow \downarrow s. \Box @_s P(x)) -$	1
2	$P(x) \rightarrow \downarrow s. \Box @_s P(x) -$	2(1)
2	$P(x) +$	3(2)
2	$\downarrow s. \Box @_s P(x) -$	4(2)
2	$\Box @_{s2} P(x) -$	5(4)
2.1	$@_{s2} P(x) -$	6(5)
2	$P(x) -$	7(6)

Это замкнутое дерево, т.к. единственная его ветка замкнута.

*Примечание.* Числа в правом столбце означают номер формулы и (номер формулы, из которой она получена). Например, «4(2)» значит, что четвертая формула получена из второй.

### Заключение

В статье показано, как логика  $HQ$  решает проблему формализации предложений с кросс-мировой квантификацией. Для данной логики справедлива следующая теорема:

**Теорема.** Для любой формулы  $\varphi$ ,  $\varphi$  доказуема тогда и только тогда, когда  $\varphi$  общезначима.

Доказательство данной теоремы выходит за рамки статьи.

### Литература

1. Braüner, T. Hybrid Logic and its Proof-Theory. Dordrecht and New York: Springer, 2010.
2. Fitting M. First-order modal logic. Second edition / M. Fitting, R.L. Mendelsohn. – Springer Cham, 2023.

3. *Kocurek, A. W.* The problem of cross-world predication. // Journal of Philosophical Logic. 2016. № 45 (6). P. 697–742.

4. *Wehmeier, K. F.* Subjunctivity and cross-world predication. // Philosophical Studies. 2012, Vol. 159.P. 107–122.

### References

1. *Braüner, T.* (, 2010) Hybrid Logic and its Proof-Theory. Dordrecht and New York: Springer.

2. *Fitting M.* (2023) First-order modal logic. Second edition / M. Fitting, R.L. Mendelsohn. – Springer Cham.

3. *Kocurek, A. W.* (2016) The problem of cross-world predication. // Journal of Philosophical Logic. № 45 (6). P. 697–742.

4. *Wehmeier, K. F.* (2012) Subjunctivity and cross-world predication. // Philosophical Studies. Vol. 159.P. 107–122.

### Информация об авторе

*Борисова Индира Искандаровна* – Национальный исследовательский Томский государственный университет (634050, г. Томск, проспект Ленина, 36).

[i.borisova2@g.nsu.ru](mailto:i.borisova2@g.nsu.ru)

### Information about the author

*Borisova Indira Iskandarovna* – Tomsk State University (36, prospect Lenina, Tomsk, 634050, Russia).

[i.borisova2@g.nsu.ru](mailto:i.borisova2@g.nsu.ru)

Дата поступления 22.05.2025

Принята к печати 11.12.2025

УДК: 165.0

DOI: 10.15372/PS20250506

EDN: OFOWKV

**Е.В. Борисов**

## **ЛОГИЧЕСКАЯ ПРЕЗЕНТАЦИЯ ПОЗНАВАЕМОСТИ**

В статье рассматривается предложенная К. Проьетти репрезентация познаваемости средствами гибридной эпистемической логики первого порядка. Данная логика содержит алетическую и эпистемическую модальности, а также сентенциональные и термовые гибридные операторы. Выявлены два недостатка данной формализации и предложены две ее модификации, устраняющие эти недостатки.

*Ключевые слова:* познаваемость, логическая формализация, *de re*, *de dicto*, эпистемическая логика, гибридная логика.

**E.V. Borisov**

## **LOGICAL PRESENTATION OF KNOWABILITY**

The paper examines the logical representation of knowability in terms of the first-order hybrid epistemic logic proposed by C. Proietti. This logic contains alethic and epistemic modalities, and hybrid operators. Two drawbacks of Proietti's formalization are shown and two modifications thereof are suggested.

*Keywords:* knowability, logical formalization, *de re*, *de dicto*, epistemic logic, hybrid logic.

## Введение

Понятие знания имеет ряд интуитивно привлекательных репрезентаций (формализаций) в различных эпистемических логиках, однако логическая репрезентация познаваемости остается открытой проблемой. Одна из наиболее острых проблем, связанных с формализацией познаваемости, состоит в следующем. Познаваемость пропозиции  $\varphi$  – это возможность знания  $\varphi$ , поэтому формула, репрезентирующая познаваемость  $\varphi$ , должна иметь вид  $\diamond(\dots K(\dots \varphi'))$ , где  $\varphi'$  – это результат некоторого преобразования  $\varphi$ <sup>1</sup>. Допустим,  $\varphi$  содержит кванторы или индивидные константы, требующие интерпретации *de re*. В этом случае репрезентация познаваемости данной пропозиции должна сохранять интерпретацию *de re* для этих кванторов и констант, однако в  $\diamond(\dots K(\dots \varphi'))$  они попадают в область действия операторов  $\diamond$  и  $K$  и, как следствие, получают интерпретацию *de dicto*. Одно из решений этой проблемы было предложено К. Проьетти: он предложил формализацию познаваемости средствами бимодальной гибридной логики первого порядка FHL (first-order hybrid logic) [8]. Это решение является предметом обсуждения в данной статье: я покажу, что оно имеет ряд недостатков и предложу две его модификации, которые эти недостатки частично устраняют.

В первой части статьи я описываю упрощенную версию FHL; во второй описываю предложенную Проьетти формализацию познаваемости; в третьей показываю два ее недостатка и предлагаю две ее модификации, призванные эти недостатки устранить; в заключении указываю на некоторые задачи, требующие решения.

## 1. Синтаксис и семантика FHL

Алфавит FHL содержит алфавит стандартной логики первого порядка, а также модальные операторы  $\diamond$  и  $K$ , счетное множество переменных для возможных миров, термовый гибридный оператор  $s$ : и сентенциональные гибридные операторы  $\downarrow s$ . и  $@_s$ , где  $s$  –

---

<sup>1</sup> В самом простом случае  $\varphi' = \varphi$ . Стоит отметить, что формула, репрезентирующая познаваемость  $\varphi$ , не может иметь вид  $\diamond K\varphi$ , поскольку эта репрезентация познаваемости порождает проблему, известную как парадокс Фитча [4, 7].

переменная для возможных миров<sup>1</sup>. Множество термов FHL рекурсивно определяется следующим образом:

$$t ::= x \mid a \mid s:t,$$

где  $x$  – индивидуальная переменная,  $a$  – индивидуальная константа,  $s$  – переменная для возможных миров. Множество формул FHL определяется так:

$$\varphi ::= P(t_1, \dots, t_n) \mid s \mid \sim\varphi \mid (\varphi_1 \& \varphi_2) \mid \diamond\varphi \mid K\varphi \mid \downarrow s.\varphi \mid @_s\varphi \mid (\exists x)\varphi,$$

где  $P$  –  $n$ -местный предикат ( $n$  – положительное натуральное число),  $t_1, \dots, t_n$  – термы,  $x$  – индивидуальная переменная,  $s$  – переменная для возможных миров.

Модель FHL – это кортеж  $M = \langle G, R, E, D, d, I \rangle$ , где  $G$  – непустое множество возможных миров;  $R$  – алетическое отношение достижимости;  $E$  – эпистемическое отношение достижимости;  $D$  – непустое множество (домен модели),  $d$  – доменная функция, назначающая каждому возможному миру непустое подмножество  $D$ ;  $I$  – интерпретация констант и предикатов, такая что: а) если  $c$  – константа, а  $w$  – возможный мир, то  $I(c, w) \in D$ ; б) если  $P$  –  $n$ -местный предикат, а  $w$  – возможный мир, то  $I(c, w) \subseteq D^n$ .

Оценка переменных в модели  $\langle G, R, E, D, d, I \rangle$  – это функция, отображающая множество индивидуальных переменных на  $D$ , а множество переменных для возможных миров – на  $G$ . Пусть  $g$  – оценка переменных в модели  $M = \langle G, R, E, D, d, I \rangle$ ,  $x$  – индивидуальная переменная,  $s$  – переменная для возможных миров,  $e \in D$ ,  $w \in G$ . Тогда  $g[e/x]$  – это оценка переменных в  $M$ , отображающая  $x$  на  $e$ , а все переменные, отличные от  $x$ , на  $g(x)$ . Аналогично для  $g[w/s]$ . Пусть

<sup>1</sup> Описание FHL дано по [8, р. 10-13]. Описанная здесь версия FHL упрощена в сравнении с оригинальной версией в следующих трех аспектах. 1) У Проиетти язык FHL содержит функциональные термы, которые я здесь игнорирую. 2) Проиетти включает в язык FHL номиналы – характерные для гибридной логики атомарные формулы, каждая из которых интерпретируется (в данной модели) как истинная в одном и только одном возможном мире. Номиналы я тоже игнорирую. 3) Оператор аскрипции знания у Проиетти индексирован термом, т.е. выглядит как  $K_t\varphi$ . Интуитивно  $K_t\varphi$  означает, что денотат  $t$  знает, что  $\varphi$ . Используемую здесь упрощенную версию FHL можно рассматривать как логику знания для одного агента. Перечисленные упрощения несущественны для дальнейшего: они позволяют сократить рассуждения, но не влияют на результат статьи. Специфика гибридной логики подробно описана, в частности, в [3].

$M = \langle G, R, E, D, d, I \rangle$  – модель,  $w$  – возможный мир в  $M$ , а  $g$  – оценка переменных в  $M$ . Тогда денотат терма  $t$  в  $M$  для  $w$  при  $g$  обозначается как  $\delta(t, w)$  и определяется следующим образом: 1) если  $t$  – индивидуальная переменная, то  $\delta(t, w) = g(t)$ ; 2) если  $t$  – индивидуальная константа, то  $\delta(t, w) = I(t, w)$ ; 3) если  $t = s:u$ , где  $s$  – переменная для возможных миров, а  $u$  – терм, то  $\delta(t, w) = \delta(u, g(s))$ .

Пусть  $M = \langle G, R, E, D, d, I \rangle$  – модель,  $w$  – возможный мир в  $M$ , а  $g$  – оценка переменных в  $M$ . Тогда истинность относительно  $M, w$  и  $g$  определяется следующим образом:

$M, w, g \models P(t_1, \dots, t_n)$  е.т.е. (если и только если)  $\langle \delta(t_1, w), \dots, \delta(t_n, w) \rangle \in I(P, w)$ ;

$M, w, g \models \sim\varphi$  е.т.е.  $M, w, g \not\models \varphi$ ; аналогично для  $\&$ ;

$M, w, g \models \diamond\varphi$  е.т.е.  $M, u, g \models \varphi$  для некоторого  $u$ , такого что  $wRu$ ;

$M, w, g \models K\varphi$  е.т.е.  $M, u, g \models \varphi$  для каждого  $u$ , такого что  $wEu$ ;

$M, w, g \models \downarrow s.\varphi$  е.т.е.  $M, w, g[w/s] \models \varphi$ ;

$M, w, g \models @_s\varphi$  е.т.е.  $M, g(s), g \models \varphi$ ;

$M, w, g \models (\exists x)\varphi$  е.т.е.  $M, u, g[e/x] \models \varphi$  для некоторого  $e \in D(w)$ .

## 2. Репрезентация познаваемости в FHL

Мы рассмотрим познаваемость только применительно к негибридным пропозициям, т.е. к формулам, не содержащим гибридных операторов. В FHL познаваемость репрезентируется с использованием функции перевода – функции от негибридных формул к формулам, которая определяется относительно двух переменных для возможных миров [8, р. 18]. Обозначим функцию перевода для переменных  $s$  и  $r$  как  $\sigma[s, r]$ . Эта функция определяется рекурсивно следующим образом:

$\sigma[s, r](P(t_1, \dots, t_n)) = P(s:t_1, \dots, s:t_n)$ , где  $t_1, \dots, t_n$  – индивидуальные переменные или константы;

$\sigma[s, r](\sim\varphi) = \sim\sigma[s, r](\varphi)$ ;

$\sigma[s, r](\varphi\&\psi) = \sigma[s, r](\varphi) \& \sigma[s, r](\psi)$ ;

$\sigma[s, r](\exists x\varphi) = @_s(\exists x)@_r\sigma[s, r](\varphi)$ ;

$\sigma[s, r](\diamond\varphi) = \diamond\downarrow q.\sigma[s, q](\varphi)$ , где  $q$  – новая переменная для возможных миров;

$\sigma[s, r](K\varphi) = K\downarrow q.\sigma[s, q](\varphi)$ .

Познаваемость (негибридной) пропозиции  $\varphi$  репрезентируется так:

$$\downarrow s. \diamond \downarrow t. K\sigma[s, t](\varphi)^1 \quad (1)$$

Преимущество (1) как формализации познаваемости состоит в том, что при оценке формул вида (1) относительно мира  $w$  некоторой модели все константы и кванторы в  $\varphi$  интерпретируются относительно  $w$ , т.е. получают интерпретацию *de re*. Например, познаваемость пропозиции  $P(a)$  выражается формулой  $\downarrow s. \diamond \downarrow t. K.[s, t](P(s:a))$ , т.е.

$$\downarrow s. \diamond \downarrow t. K.P(s:a). \quad (2)$$

Истинностные условия (2) таковы:  $M, w, g \models (2)$  е.т.е. для некоторого  $w_1$ , такого что  $wRw_1$ , и для любого  $w_2$ , такого что  $w_1Ew_2$ ,  $I(a, w) \in I(P, w_2)$ . Как видим, здесь используется  $I(a, w)$ , т.е. денотат  $a$  для  $w$ ; таким образом, в (2)  $a$  имеет интерпретацию *de re*.

### 3. Критика и предложения

Предложенная Проьетти формализация познаваемости имеет ряд недостатков. В этом разделе показаны два ее недостатка<sup>2</sup> и предложена ее модификация, которая эти недостатки устраняет.

I. С использованием (1) познаваемость  $(\exists x)P(x)$  выражается формулой

$$\downarrow s. \diamond \downarrow t. K@_s(\exists x)@_t P(s:x) \quad (3)$$

В силу данного выше определения истины, имеет место следующее:

$M, w, g \models (3)$  е.т.е. для некоторого  $w_1$ , такого что  $wRw_1$ , и для любого  $w_2$ , такого что  $w_1Ew_2$ , существует объект  $e \in D(w)$ , такой что  $e \in I(P, w_1)$ .

<sup>1</sup> Данное здесь определение функции перевода и познаваемости несколько отличается от оригинальных. Это обусловлено тем, что, как отмечено выше, я использую упрощенную версию FHL.

<sup>2</sup> Некоторые дополнительные критические замечания по трактовке познаваемости у Проьетти представлены в [1; 2].

Здесь важно, что переменная  $w_2$ , не используется в утверждении «существует объект  $e \in D(w)$ , такой что  $e \in I(P, w_1)$ ». Это значит: если допустить, что в  $M$  каждый возможный мир имеет эпистемическую альтернативу<sup>1</sup>, то указанные истинностные условия эквивалентны следующим:

$M, w, g \models (3)$  е.т.е. для некоторого  $w_1$ , такого что  $wRw_1$ , существует объект  $e \in D(w)$ , такой что  $e \in I(P, w_1)$ .

Но эти истинностные условия совпадают с истинностными условиями  $(\exists x)\diamond P(x)$ , т.е. оказывается, что, как ни странно, (3) семантически эквивалентно  $(\exists x)\diamond P(x)$ . Странно здесь то, что в  $(\exists x)\diamond P(x)$  отсутствует эпистемический оператор, что лишает эту формулу какого-либо эпистемического смысла, однако эта формула оказывается эквивалентна эпистемическому утверждению о познаваемости.

Аналогичный пример: познаваемость  $(\exists x)\diamond P(x)$  выражается формулой  $\downarrow s.\diamond\downarrow t.K@_s(\exists x)@_t\diamond\downarrow r.P(s;x)$ , которая эквивалентна  $(\exists x)\diamond\diamond P(x)$ . Как и в первом примере, утверждение о познаваемости оказывается эквивалентно формуле, не содержащей оператора  $K$ , а значит, не имеющей эпистемического смысла.

Итак, первый недостаток предложенной Проиетти формализации познаваемости состоит в том, что познаваемость некоторых пропозиций выражается формулами, не имеющими эпистемического смысла. Для устранения этого недостатка я предлагаю выражать познаваемость пропозиции  $\phi$  следующим образом:

$$\downarrow s.\diamond K\downarrow t.\sigma[s, t](\phi) \tag{4}$$

Как видим, (4) отличается от (1) только тем, что оператор  $\downarrow t.$ , который в (1) расположен слева от  $K$ , в (4) расположен справа от  $K$ . Эта малозаметная модификация (1) устраняет указанный недостаток. В самом деле, применив (4) к  $(\exists x)P(x)$ , мы получим:

$$\downarrow s.\diamond K\downarrow t.@_s(\exists x)@_tP(s;x) \tag{5}$$

---

<sup>1</sup> Это условие выполняется, например, если мы в определении модели укажем, что  $E$  рефлексивно. В эпистемической логике рефлексивность эпистемического отношения достижимости является стандартной характеристикой моделей, потому что она обеспечивает общезначимость стандартного принципа фактивности знания:  $K\phi \rightarrow \phi$ .

Истинностные условия (5) таковы:  $M, w, g \models (5)$  е.т.е. для некоторого  $w_1$ , такого что  $wRw_1$ , и для любого  $w_2$ , такого что  $w_1Ew_2$ , существует  $e \in D(w)$ , такой что  $e \in I(P, w_2)$ . Как видим, здесь отношение эпистемической достижимости используется существенным образом, т.е. его невозможно переписать без использования  $E$ . Соответственно, истинностные условия (5) невозможно выразить без использования эпистемического оператора<sup>1</sup>. Нетрудно убедиться в том, что подобным образом дело обстоит с утверждением о познаваемости  $(\exists x)\Diamond P(x)$ . Таким образом, предложенная модификация (1) решает указанную проблему.

II. В формуле, репрезентирующей познаваемость пропозиции  $\phi$ , будь то на основе (1) или (3), все константы в  $\phi$  интерпретируются *de re*. Однако во многих случаях имеет смысл интерпретировать константы *de dicto* (например, если мы используем константу как формальный эквивалент определенной дескрипции), и часто необходима комбинированная интерпретация, когда в одной и той же формуле некоторые константы интерпретируются *de re*, а некоторые – *de dicto*. Однако (1), как и (4), не допускают интерпретации *de dicto*. Для устранения этого недостатка я предлагаю добавить в алфавит формального языка  $\lambda$ -оператор и внести следующие изменения в синтаксис и семантику FHL, а также в определение функции перевода:

1. Включить в определение формулы пункт:  $(\lambda x.\phi)(c)$  ( $\phi$  – формула,  $c$  – константа).

2. Включить в определение истины пункт:  $M, w, g \models (\lambda x.\phi)(c) \Leftrightarrow M, w, g[I(c, w)/x] \models \phi$ .

3. В определении  $\sigma[s, r]$  заменить пункт для атомарных формул в определении  $\sigma[s, r]$  следующим: если  $\phi$  – атомарная формула, то  $\sigma[s, r](\phi) = \phi$ .

4. В определении  $\sigma[s, r]$  добавить пункт:  $\sigma[s, r](\lambda x.\phi)(c) = @_s(\lambda x.@_r\sigma[s, r](\phi))(c)$ .

---

<sup>1</sup> Более того, эти истинностные условия невозможно выразить без использования гибридных операторов. Это обусловлено тем, что данные истинностные условия содержат кросс-мировую квантификацию: переменная  $e$  пробегает по домену  $w$ , тогда как предикат  $P$  интерпретируется относительно мира  $w_2$ . Феномен кросс-мировой квантификации и его отображение средствами гибридной логики детально обсуждается в [6].

$\lambda$ -оператор позволяет на объектном языке определять, какие константы требуют интерпретации *de re*, а какие – интерпретации *de dicto*<sup>1</sup>. Например, в формуле  $(\lambda x. \diamond P(x, b))(a)$  константа  $a$  имеет интерпретацию *de re*,  $b$  – *de dicto* (поскольку  $a$  не попадает, а  $b$  попадает в область действия  $\diamond$ ). Интуитивный смысл познаваемости требует, чтобы тезис о познаваемости этой пропозиции сохранял эту интерпретацию констант. Убедимся в том, что предложенная трактовка познаваемости выполняет это требование. Познаваемость этой пропозиции с учетом (4) и с использованием модифицированного определения функции перевода выражается формулой

$$\downarrow s. \diamond K \downarrow t. @_s (\lambda x. @_t \diamond \downarrow r. P(x, b))(a) \quad (6)$$

С учетом дополненного определения истины мы получаем:

$M, w, g \models (6)$  е.т.е. для некоторого  $w_1$ , такого что  $wRw_1$ , и для любого  $w_2$ , такого что  $w_1Ew_2$ , существует  $w_3$ , такой что  $w_2Rw_3$ ,  $(I(a, w), I(b, w_3)) \in I(P, w_3)$ .

Как видим, здесь используется денотат  $a$  в  $w$  и денотат  $b$  в  $w_3$ , т.е.  $a$  имеет интерпретацию *de re*,  $b$  – интерпретацию *de dicto*, что и требовалось.

### Заключение

Предложенная трактовка познаваемости средствами (модифицированной) FHL устраняет недостатки теории Проиетти, указанные в предыдущем разделе. Однако эта трактовка познаваемости тоже не является совершенной; она оставляет открытыми как минимум две проблемы.

1) Она не допускает интерпретацию *de dicto* для кванторов. Однако в некоторых случаях интуитивный смысл пропозиции  $\phi$  требует такой интерпретации, поэтому в утверждении о познаваемости  $\phi$  эту интерпретацию необходимо сохранить.

2) В пропозиции  $\phi$  константы могут иметь не только интерпретацию *de re* или *de dicto*, но и промежуточную интерпретацию. Например, в  $\diamond(\lambda x. \diamond P(x))(c)$  константа  $c$  имеет промежуточную интерпретацию: она интерпретируется *de dicto* относительно одного оператора возможности и *de re* относительно другого.

<sup>1</sup> Такое использование  $\lambda$ -оператора в модальной логике детально описано в [5].

Эта интерпретация должна быть сохранена и в формуле, выражающей познаваемость  $\phi$ . Однако предложенная трактовка познаваемости отождествляет промежуточную интерпретацию с интерпретацией *de re*: познаваемость  $\diamond(\lambda x.\Diamond P(x))(c)$  выражается формулой  $\downarrow s.\Diamond K\downarrow t.\diamond\downarrow r.@_s(\lambda x.@_r\diamond\downarrow q.P(x))(c)$ , в которой  $c$  имеет интерпретацию *de re*.

Устранение указанных недостатков предложенной трактовки познаваемости является задачей для дальнейшей работы.

### Литература

1. Борисов Е.В. Познаваемость в гибридной эпистемической логике // Вестник Томского государственного университета. Философия. Социология. Политология. 2023. 76. С. 11–17. DOI 10.17223/1998863X/76/2.
2. Борисов Е.В. О двух формализациях принципа познаваемости *de re* // Омский научный вестник. Общество. История. Современность. 2024. Т. 9. № 4. С. 58–62. DOI: 10.25206/2542-0488-2024-9-4-58-62.
3. Braüner T. Hybrid Logic and its Proof-Theory. Dordrecht: Springer, 2011.
4. Brogaard B., Salerno J. Fitch's Paradox of Knowability // The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/entries/fitch-paradox>. 2019.
5. Fitting M., Mendelsohn R. L. First-Order Modal Logic. Dordrecht: Springer, 2023.
6. Kocurek A.W. The problem of cross-world predication // Journal of Philosophical Logic. 2016. 45(6). P. 697–742. DOI 10.1007/s10992-015-9389-z.
7. Fitch F. A Logical Analysis of Some Value Concepts // Journal of Symbolic Logic. 1963. Vol. 28. P. 113–118.
8. Proietti C. The Fitch-Church Paradox and First Order Modal Logic // Erkenntnis. 2016. Vol. 81. P. 87–104.

### References

1. Borisov E.V. Poznavayemost' v gibridnoy epistemicheskoy logike [Knowability in hybrid epistemic logic] // Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sociologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science. [Tomsk State University Journal of Philosophy, Sociology and Political Science]. 2023. № 76. С. 11–17. DOI 10.17223/1998863X/76/2. (In Russian.)

2. *Borisov, E.V.* (2024) O dvukh formalizatsiyakh printsipa poznavayemosti de re [On two formalizations of the principle of knowability de re] Omskiy Nauchnyy Vestnik. Obshchestvo. Istoriya. Sovremennost'. [Omsk Scientific Bulletin. Series Society. History. Modernity]. 9(4). P. 58–62.

3. *Braüner, T.* (2011) Hybrid Logic and its Proof-Theory. Dordrecht: Springer.

4. *Brogaard, B., Salerno, J.* (2019) Fitch's Paradox of Knowability. The Stanford Encyclopedia of Philosophy. At: <https://plato.stanford.edu/entries/fitch-paradox>.

5. *Fitch, F.* (1963) A Logical Analysis of Some Value Concepts. Journal of Symbolic Logic. 28. P. 113–118.

6. *Fitting, M., Mendelsohn, R. L.* (2023) First-Order Modal Logic. Dordrecht: Springer.

7. *Kocurek, A.W.* (2016) The problem of cross-world predication. Journal of Philosophical Logic. 45(6). P. 697–742.

8. *Proietti, C.* (2016) The Fitch-Church Paradox and First Order Modal Logic. Erkenntnis. 81. P. 87–104.

### Информация об авторе

*Борисов Евгений Васильевич* – доктор философских наук, главный научный сотрудник Института философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8).

[borisov.evgeny@gmail.com](mailto:borisov.evgeny@gmail.com)

### Information about the author

*Borisov Evgeny Vasilyevich* – Doctor of Sciences (Philosophy), Chief Researcher at the Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090, Russia).

[borisov.evgeny@gmail.com](mailto:borisov.evgeny@gmail.com)

Дата поступления 10.04.2025

Принята к печати 11.12.2025

УДК: 165.72:165.12:1(091)  
DOI: 10.15372/PS20250507  
EDN: NFXBEE

**И.В. Берестов**

### **ПРОБЛЕМА В ТЕОРИИ ИЗНАЧАЛЬНОЙ АППЕРЦЕПЦИИ КАНТА**

Мы показываем, что изложенное в «Критики чистого разума» (2-е изд.) описание изначальной апперцепции, необходимой для любого рассудочного синтеза, содержит противоречие. По Канту, для рассудочного синтеза единого представления из элементов созерцаемого многообразия необходимо признать представления, которые входят в это многообразие, принадлежащими мне самому. Для этого необходимо осуществить синтез некоторого представления. Этот синтез, по Канту, не требует каких-либо дальнейших синтезов. Но для осуществления последнего синтеза необходимо осуществить новый синтез. Следовательно, синтез, который не требует для своего осуществления других синтезов, невозможен. Наш аргумент использует семантический анализ синтезируемых рассудком представлений (*Vorstellungen*), таким образом, в настоящей статье мы намерены продемонстрировать изменение интерпретации Канта в свете «семантического поворота».

*Ключевые слова:* Критика чистого разума, единство самосознания у Канта, трансцендентальная апперцепция, субъектность у Канта, «Я мыслю» у Канта, синтетическое единство апперцепции.

**I.V. Berestov**

### **THE PROBLEM IN KANT'S THEORY OF PRIMARY APPERCEPTION**

We show that the description of the primary apperception in the Critique of Pure Reason (2<sup>nd</sup> edition) – which is necessary for any synthesis by an act of judgement – contains a contradiction. According to Kant, in order

to synthesise by an act of judgement a single representation from the elements of a given in intuition manifold, it is necessary to recognise the representations which are contained in this manifold as my own representations. In order to do this, it is necessary to carry out the synthesis of a certain representation. This synthesis, according to Kant, requires no further syntheses. But the latter synthesis requires a new synthesis. Therefore, a synthesis that does not have other syntheses as a prerequisite for its realisation is impossible. The argument makes use of the semantic analysis of representations (*Vorstellungen*) that are synthesised by judgments, so that in the present paper we intend to show a change in the understanding of Kant in the light of the “semantic turn”.

*Keywords:* Critique of Pure Reason, unity of self-consciousness in Kant, transcendental apperception, subjectivity in Kant, “I think” in Kant, synthetic unity of apperception.

## Введение

В настоящей статье мы намерены проанализировать корректность кантовской теории самосознания. По Канту (§§ 15-24 второго издания *Критики чистого разума*), самосознание лежит в основе рассудочного знания, оно необходимо для рассудочного мышления, синтезирующего объекты познания, но корректная и основывающаяся на приемлемых посылах интерпретация того, что Кант понимал под «самосознанием» и его ролью в формировании синтезируемых рассудком представлений, как оказалось, представляет собой серьезную проблему для современных историков философии.

В настоящей статье проблематичность кантовской теории самосознания выявляется при анализе *содержания* рассудочно мыслимого субъектом сложного представления (*Vorstellung*), а не ментального акта, которым это представление мыслится (и который на это представление направлен). В терминологии А. Коффы, это *семантический* анализ [Коффа, 2019. С. 6], а значит, сама проблема тоже является *семантической*. Таким образом, мы намерены проанализировать Канта с точки зрения «семантического поворота» в терминологии из [Коффа, 2019. С. 8]. В настоящей статье мы планируем начать работу по иллюстрации той роли, которую «семантический поворот» играет в изменении подходов

к изучению наследия Канта в истории философии приступить к осмыслению способа, которым «семантический поворот» может повлиять на каноны, в которые историки философии встраивают кантовскую теорию субъекта.

А. Коффа критикует Канта за смешение акта рассудочного мышления, направленного на некоторое синтетическое представление, с этим синтетическим представлением, а также за использование двух определений «аналитического», что приводит к путанице [Коффа, 2019. С. 17-18, 22-30]. Наш анализ Канта не использует представленные А. Коффой критические аргументы.

Мы – в отличие от П. Ф. Стросона [Strawson, 1966; Strawson, 1994], К. Т. Пауэлла (С. Т. Powell) [Powell, 1980], Г. Эллисона (Н. Allison), П. Китчер (Р. Kitcher), Э. Брука (А. Brook) и Р. Пиппина (R. Pippin) [Allison, 1996; Kitcher, 1990; Brook, 1994; Pippin, 1989], а также К. Америкса [Ameriks, 2006. Р. 19, 64-66] – не разделяем мнения о «согласованности» кантовского понимания апперцепции, но наша критика не связана с той критикой, от которой пытается защитить Канта К. Америкс в [Ameriks, 2006. Р. 19, 59], а также не связана с критикой Канта его младшими современниками (К. Л. Рейнгольдом (K. L. Reinhold) и Г. Фихте) и известными современными интерпретаторами Канта (например, Д. Хенрихом (D. Henrich) и М. Франком (M. Frank) [Henrich, 1982; Frank, 1991; Frank, 1995]).

### **Неформальная схема возникновения проблемы первого рассудочного синтеза**

Целью настоящей статьи является выявление трудностей, заключающихся в *первом* и *втором* кантовских тезисах (*Критика чистого разума*, §§ 15-24 второго издания). В *первом тезисе* утверждается, что

необходимым условием для осуществления рассудочного синтеза какого-либо представления – каковое представление обязательно содержит некоторые элементы – является осуществление рассудочного синтеза представления *я мыслю этот элемент = я осознаю этот элемент = этот элемент принадлежит мне / моему сознанию* для каждого такого элемента, *sed non vice versa*.

Иначе говоря, для того, чтобы был возможен акт рассудка, синтезирующий какое-либо сложное представление из элементов, эти элементы должны быть *доступными* для синтеза, а для этого они должны *осознаваться* моим «я» как доступные для синтеза, но, чтобы осуществить указанный акт осознания, для каждого элемента я должен *уже* обладать представлением *я мыслю / осознаю этот элемент*:

«В самом деле, многообразные представления, данные в некотором созерцании, вместе взятые, не были бы моими представлениями, если бы они не принадлежали все вместе одному самосознанию [= представлению «я мыслю», сопровождающему каждое из этих представлений]; ... в противном случае не все они принадлежали бы мне» (§ 16, В 132 203.25-В 133 203.33<sup>1</sup>).

Благодаря мышлению представлений вида *я мыслю / осознаю этот элемент* осуществляется некоторый сопровождающий синтез единого представления из данных в созерцании элементов – синтез, благодаря осуществлению которого актом рассудка становится возможным осуществление других рассудочных синтезов, использующих эти элементы. Ниже Кант излагает эту мысль следующим образом:

«Вот почему мысль (Gedanke), согласно которой все представления, данные в созерцании, в совокупности (ingesamt) принадлежат мне, означает, что я соединяю или по крайней мере могу соединить их в одном самосознании [= в одном представлении «я мыслю», сопровождающем каждое из этих представлений]; и хотя сама эта мысль ещё не есть осознание синтеза представлений, тем не менее она предполагает возможность его; иными словами, только в силу того, что я могу постичь многообразие представлений в одном сознании, я называю все их моими представлениями; в противном случае, я имел бы столь же пёстрое разнообразное Я (Selbst), сколько у меня есть сознаваемых мной представлений» (§ 16, В 134 205.9-205.17).

---

<sup>1</sup> Ссылки на второе издание *Критики чистого разума* Канта и цитаты из неё даются по двуязычному (оригинальный немецкий текст и русский перевод) изданию [Кант, 2006]. Ссылка «В 132 203.25» означает, что немецкий оригинал цитируемого русского текста находится на с. 132 второго издания (опубликованного в 1787 г. на немецком языке), в издании же [Кант, 2006] цитируемый русский текст находится на с. 203, строка 25.

При осуществлении синтеза данных в созерцании представлений в одно представление создаётся *представление*, что исходные элементы, доступные для синтеза из них какого-либо представления, принадлежат *мне*. И вообще, по Канту, для осуществления *любого* представления необходимо, чтобы его элементы осознавались как *мои* представления. Следовательно, любой синтез нуждается в другом синтезе, предшествующем по отношению к нему, этот другой синтез также нуждается в синтезе, предшествующем по отношению к нему, и т. д. до бесконечности.

При этом Кант утверждает *второй тезис*:

необходимым условием для осуществления рассудочного синтеза какого-либо представления некоторого рода (а именно, представления, в котором синтезируются элементы, полученные непосредственно из многообразия созерцаемого) – каковое представление обязательно содержит некоторые элементы – является осуществление некоторого «первого» или «изначального» рассудочного синтеза представления я мыслью / осознаю этот элемент для каждого такого элемента, *sed non vice versa*.

Кант излагает это следующим образом:

«Должно быть возможно, чтобы [представление] я мыслью (Ich denke) сопровождало все мои представления (Vorstellungen); в противном случае во мне представлялось бы нечто такое, что вовсе нельзя было бы мыслить, иными словами, представление или было бы невозможным, или по крайней мере для меня не существовало бы. То представление, которое может быть дано до всякого мышления, называется созерцанием (Anschauung). Всё многообразное (alles Mannigfaltige) в созерцании имеет, следовательно, необходимое отношение к [представлению] я мыслью в том самом субъекте, в котором это многообразное находится. Но это представление есть акт спонтанности... я называю его [т. е. представление я мыслью] чистой апперцепцией, чтобы отличить его от эмпирической апперцепции, или также изначальной (ursprüngliche) апперцепцией...» (§ 16, В 131 203.7-В 132 203.21).

Несколько другими словами об «*изначальном*» синтезе Кант говорит сразу же ниже:

«... ведь оно [представление я мыслью] есть то самосознание, которое – порождая представление я мыслью – должно иметь возможность сопровождать все другие (andere) представления и быть одним и тем же во всяком сознании; а само это самосознание

[=представление я мыслю] никаким иным [дальнейшим, следующим – weiter – представлением я мыслю] сопровождаться не может. Единство его [т.е. представления я мыслю = самосознания, порождающего представление я мыслю] я называю также трансцендентальным единством самосознания» (§ 16, В 132 203.19-203.24).

Заметим, что все приведённые цитаты из второго издания *Критики чистого разума* можно трактовать несколькими способами. Мы полагаем, что они содержат в себе (помимо прочего) *первый* и *второй тезисы*, в которых утверждается (помимо прочего), что для осуществления синтеза из нескольких элементов единого представления необходимо осуществление синтезов *нескольких* представлений вида «я мыслю, что элемент 1 принадлежит мне», «я мыслю, что элемент 1 принадлежит мне», ... . Видно, что в каждом из последних представлений содержится одно и то же представление *я мыслю*. С другой стороны, кантовский текст можно трактовать как утверждающий, что для осуществления синтеза из нескольких элементов единого представления необходимо осуществление синтеза *одного* представления вида «я мыслю следующее: (я мыслю, что *a* есть F) & (я мыслю, что *b* есть G) &...» [Ameriks, 2006. P. 57]. Поскольку текст Канта не позволяет исключить одну из трактовок, мы остановились на первой, как на более простой для анализа – заметим, что трактовка К. Америкса требует непростого обсуждения разведения Кантом трансцендентальной и эмпирической апперцепции, а также мыслей (суждений) первого и более высоких порядков [Ameriks, 2006. P. 55-58]. Какая бы трактовка ни оказалась предпочтительной, это не влияет на отстаиваемый в настоящей статье тезис о проблематичности теории Канта, поскольку (как мы намерены показать в последующих публикациях) выявляемая нами проблема сохраняется в обоих трактовках.

На наш взгляд, две последние цитаты следует интерпретировать как утверждающие, что «*первый*» или «*изначальный*» синтез имеет следующие свойства:

- (1) он является синтезом, *непосредственно предшествующим* возможности синтеза единого представления из данных в созерцании элементов, т. е. синтезом, благодаря осуществлению которого актом рассудка становится возможным осуществление других рассудочных синтезов, использующих эти элементы;

- (2) для осуществления этого «первого» синтеза не требуется осуществления какого-либо другого синтеза;
- (3) без осуществления «первого» синтеза осуществление какого-либо синтеза из указанных элементов невозможно.

В соответствии с *первым тезисом*, любой синтез нуждается в другом синтезе, предшествующем по отношению к нему, что противоречит (2). Это означает, что «изначальный» синтез невозможен. Таким образом, если признаются *первый тезис* и *второй тезис* Канта, то в теории рассудочного синтеза Канта имеется противоречие.

Изложим ядро этого рассуждения несколько короче.

1. Если рассудок синтезирует некое сложное представление из элементов, эти элементы должны присутствовать в *моём* сознании.
2. Чтобы они присутствовали в *моём* сознании, они должны осознаваться мною как *мои*.
3. Чтобы они были осознаны мною как *мои*, должен быть произведён их синтез с представлением о моём я (я должен мыслить *я мыслю x*); посредством этого синтеза они помещаются в моё сознание, в область доступного для моего синтеза.
4. Чтобы был произведён этот (как и любой) синтез, элементы потенциального синтетического единства должны осознаваться мною как *мои*, и т. д.

Таким образом, какой бы синтез мы не назвали *первым*, всегда будет существовать синтез, который должен предшествовать ему.

Следовательно, осуществление *первого* рассудочного синтеза невозможно.

### Формализация ядра проблемы

В настоящем разделе мы изложим ядро проблемы несколько более формально.

Пусть  $\prec$  – бинарное отношение строгого полного или частичного порядка (транзитивное, иррефлексивное, асимметричное). Формула “ $x \prec y$ ” может читаться как « $x$  является вторичным по

отношению к  $y$ » или как « $y$  является первичным по отношению к  $x$ » или как «существование  $y$  является условием возможности существования  $x$ ». Мы будем считать, что это отношение соотносит положения дел или пропозиции, выражаемые описывающими положения дел предложениями. Также можно добавить, что отношение  $\prec$  немонотонно (т. е. из  $x \prec y$  нельзя вывести  $x \prec (y \& z)$ ; здесь  $x, y, z$  рассматриваются как пропозиции). Можно считать, что  $\prec$  является отношением экзистенциальной зависимости, причём только в одну сторону: вторичное экзистенциально зависит от первичного, но не наоборот (т. е. если  $x \prec y$ , то осуществление  $x$  невозможно без осуществления  $y$ , но осуществление  $y$  возможно без осуществления  $x$ ).

Приведённые характеристики отношения  $\prec$  делают это отношение весьма близким к отношению фундирования (grounding), или отношения метафизической (в отличие от каузальной) зависимости [Schaffer, 2009; Rosen 2010; Fine, 2012], точнее – к отношению сильного частичного (strict partial) фундирования [Fine, 2012]<sup>1</sup>.

Пусть формула  $\Phi$  выражает какое-либо являющееся пропозицией (суждением) представление, являющееся синтетическим единством представлений, синтезируемым (посредством некоторого акта рассудочного мышления) каким-либо субъектом  $s$ .

Для упрощения дальнейших рассуждений можно принять, что в формуле  $\Phi$  присутствуют только константы и предикаты, нет свободных переменных, нет операторов, нет кванторов и связанных переменных. В формулу  $\Phi$  могут входить выражения вида  $[\Psi_i]$ , и  $\Xi_j$ , где  $\Psi$  и  $\Xi$  – некоторые выражающие пропозиции формулы, удовлетворяющие приведённым выше ограничениям,  $i, j$  – некоторые натуральные числа. Подробнее об интерпретации выражений указанного вида будет написано ниже.

Например, пусть  $\Phi$  – формула  $P(a)$ , что означает «Сократ есть белый». Значения имени «Сократ» ( $a$ ) и предиката «\_\_есть белый» ( $P$ ) даны в многообразии созерцаний.

Теперь мы полагаем достаточно обоснованным, что следующее положение соответствует взглядам Канта из второго издания *Критики чистого разума*:

---

<sup>1</sup> Благодарю Георгия Владиславовича Черкасова (МГУ) за разъяснение роли, которую играет понятие “grounding” в дискуссиях современной аналитической метафизики.

(4)

Субъект $s$ рассудочно мыслит, что $\Phi$	<	Все значения конститuent $\Phi$ осознаются субъектом $s$ как его собственные
Субъект $s$ рассудочно синтезирует представление <i>что</i> $\Phi$		Все значения конститuent $\Phi$ принадлежат области сознаваемого субъектом $s$
Субъект $s$ объединяет данное в созерцании многообразное спонтанным актом рассудочного синтеза в представление <i>что</i> $\Phi$		Субъект $s$ рассудочно мыслит, что представление о первой конститuentе формулы $\Phi$ принадлежит ему самому, и субъект $s$ рассудочно мыслит, что представление о второй..., и субъект $s$ рассудочно мыслит, что представление о третьей...

В приведённой таблице каждое из трёх положений из первого столбца связывается отношением “<” с каждым из трёх положений из третьего столбца. Это происходит потому, что для Канта два положения из первого столбца эквивалентны друг другу, и три положения из третьего столбца эквивалентны друг другу. Приведённая таблица иллюстрирует, что теория Канта о рассудочном синтезе может быть изложена различными способами, но для целей настоящей статьи важен только следующий способ изложения:

(5) Субъект  $s$  рассудочно мыслит, что  $\Phi$  < Субъект  $s$  рассудочно мыслит, что представление о первой конститuentе формулы  $\Phi$  принадлежит ему самому, и что представление о второй..., и что представление о третьей....

Для упрощения дальнейшего анализа введём следующие обозначения.

Пусть  $\Phi_i$  обозначает константу, предикат, оператор, логическую связку из формулы  $\Phi$ , первое вхождение которой / которого в формулу  $\Phi$  находится на  $i$ -том месте (где  $i$  – натуральное число) среди всех вхождений термов, предикатов, операторов и логических связок. Например, если  $\Phi$  обозначает формулу  $P(a)$ , в которой значения  $P$  и  $a$  те же, что и ранее, то  $\Phi_1$  обозначает предикат  $P$ ,  $\Phi_2$  обозначает терм (а именно, константу)  $a$ . Далее мы будем называть  $\Phi_i$   $i$ -той значимой конститuentой формулы  $\Phi$ .

Пусть формула  $T_s\Phi$  читается как «субъект  $s$  рассудочно мыслит, что  $\Phi$ ». Таким образом,  $T_s$  является доксистическим оператором. Формула  $\Phi$ , целиком находящаяся под оператором  $T_s$ , может заключаться в скобки во избежание неоднозначности из-за воз-

можно трактовки формулы  $\Phi$  как находящейся под оператором  $T_s$  частично, а не полностью.

Поскольку мыслится всегда представление, можно ввести отдельное обозначение для представлений: значением выражения  $[[\Phi]]$  из формулы  $T_s[[\Phi]]$  является *представление, выражаемое формулой  $\Phi$* . Тогда введённую выше формулу  $T_s\Phi$  более точно можно будет переписать в виде  $T_s[[\Phi]]$ . Поскольку *представление, выражаемое формулой  $\Phi$*  рассудочно синтезируется и мыслится субъектом, представление, являющееся значением выражения  $[[\Phi]]$  из формулы  $T_s[[\Phi]]$  можно назвать пропозицией (суждением). Мы можем опускать  $[[\dots]]$  в выражениях вида  $T_s[[\Phi]]$ , если это не создаёт неоднозначности.

Например, если  $\Phi$  – формула  $P(a)$ , имеющая указанную выше интерпретацию константы и предиката, то  $T_s\Phi$  есть  $T_sP(a)$  или  $T_s[[P(a)]]$ , что интерпретируется как «субъект  $s$  рассудочно мыслит / рассудочно синтезирует представление, являющееся синтетическим единством представлений (а именно, пропозицию или суждение) *Сократ есть белый*».

Расширим описанное выше употребление выражений вида  $[[\dots]]$  следующим образом. Примем, что в выражениях, где под оператором  $T_s$  стоит формула  $\Psi$ , такая, что либо в самой  $\Psi$ , либо в одной или нескольких подформулах  $\Psi$  выражению  $[[\Phi_i]]$  что-либо предидируется, значением выражения  $[[\Phi_i]]$  является представление, являющееся значением значимой конституенты  $\Phi_i$  формулы  $\Phi$ .

Таким образом, выражения вида  $[[\dots]]$  встречаются в контекстах вида  $T_s[[\Phi]]$  или  $T_s(\dots[[\Phi_i]]\dots)$ . Значением выражения  $[[\Phi_i]]$  может быть пропозиция (суждение), понятие, элемент созерцаемого многообразия. Все они являются представлениями.

Смысл введения в язык выражений вида  $[[\dots]]$  состоит в том, чтобы сделать синтаксис языка понятным, а именно, чтобы избежать путаницы, которая может возникнуть, когда что-то предидируется предикатам, операторам, логическим связкам. Формулы, выражающие это, выходят за пределы синтаксиса языка первопорядковой логики с операторами, поэтому, чтобы язык был максимально привычным и, в то же время, понятным, мы будем использовать выражения вида  $[[\dots]]$ . Выражения вида  $[[\dots]]$  не являются какими-либо значимыми конституентами содержащих их формул.

Теперь мы в состоянии переписать положение (5) в более формальном виде:

$$(6) (\forall \Phi_i) \{ (\mathbf{T}_s \Phi) \prec [\mathbf{T}_s(\llbracket \Phi_i \rrbracket \varepsilon s)] \},$$

где  $\varepsilon$  – отношение «(что?)\_принадлежит (чему?)\_»;

$i$  – номер места, на котором находится первое вхождение значимой конstituенты  $\Phi_i$  формулы  $\Phi$  в формулу  $\Phi$ .

Для упрощения дальнейших рассуждений перепишем формулу (6) для произвольной значимой  $i$ -той конstituенты  $\Phi_i$  формулы  $\Phi$ :

$$(7) (\mathbf{T}_s \Phi) \prec [\mathbf{T}_s(\Box \Phi_i \Box \varepsilon s)].$$

Заметим, что в правой части (7) под оператором  $\mathbf{T}_s$  стоит формула  $\llbracket \Phi_i \rrbracket \varepsilon s$ . Обозначим эту формулу через  $\Psi$ :

$$(8) \Psi =_{\text{df.}} \llbracket \Phi_i \rrbracket \varepsilon s.$$

Заменим в правой части (7)  $\llbracket \Phi_i \rrbracket \varepsilon s$  на  $\Psi$ , что, в соответствии с (8), можно сделать в силу *Принципа взаимной подставимости тождественных*:

$$(9) (\mathbf{T}_s \Phi) \prec (\mathbf{T}_s \Psi).$$

Поскольку в (7) формула  $\Phi$  является произвольной, мы можем подставить в формуле (6)  $\Psi$  вместо  $\Phi$ . После этого в получившейся формуле заменим  $i$  на  $j$ , т. е. будем считать, что  $\Psi_j$  – значимая  $j$ -я конstituента формулы  $\Psi$ . Тогда получим для произвольной значимой  $j$ -той конstituенты  $\Psi_j$  формулы  $\Psi$ :

$$(10) (\mathbf{T}_s \Psi) \prec [\mathbf{T}_s(\llbracket \Psi_j \rrbracket \varepsilon s)].$$

По (8), значимых конstituент в формуле  $\Psi$  всего 3:

$$(11) \forall \Psi_j j = 1, 2, 3; \Psi_1 =_{\text{df.}} \Phi_i; \Psi_2 =_{\text{df.}} \varepsilon; \Psi_3 =_{\text{df.}} s.$$

Из (9) и (10) получаем:

$$(12) (\mathbf{T}_s \Phi) \prec (\mathbf{T}_s \Psi) \prec [\mathbf{T}_s(\llbracket \Psi_j \rrbracket \varepsilon s)],$$

где  $j = 1, 2, 3$ .

В соответствии с (8), в (12) вместо  $\Psi$  можно подставить  $\llbracket \Phi_i \rrbracket \varepsilon s$ :

$$(13) (\mathbf{T}_s \Phi) \prec (\mathbf{T}_s \{ \llbracket \Phi_i \rrbracket \varepsilon s \}) \prec [\mathbf{T}_s(\llbracket \Psi_j \rrbracket \varepsilon s)],$$

где  $j = 1, 2, 3$ .

Рассмотрим **первый пример**. Пусть формула  $\Phi$  задана следующим образом:

$$(14) \Phi =_{\text{df}} P(a).$$

Из (14) следует:

$$(15) \forall \Phi_i \ i = 1, 2; \Phi_1 = P, \Phi_2 = a.$$

Перепишем (13) для случая  $i = 1, j = 1$ :

$$(16) (\mathbf{T}_s\Phi) \prec (\mathbf{T}_s\{\llbracket\Phi_1\rrbracket\mathcal{E}s\}) \prec [\mathbf{T}_s(\llbracket\Psi_1\rrbracket\mathcal{E}s)].$$

По (14), подставим в (16) вместо  $\Phi$  формулу  $P(a)$ , по (15), вместо  $\Phi_1$  подставим  $P$ , по (11), вместо  $\Psi_1$  подставим  $\Phi_1$ , вместо последнего  $\Phi_1$ , по (15), опять подставим  $P$ , после чего получим:

$$(17) (\mathbf{T}_s[P(a)]) \prec (\mathbf{T}_s\{\llbracket P\rrbracket\mathcal{E}s\}) \prec [\mathbf{T}_s(\llbracket P\rrbracket\mathcal{E}s)].$$

В (17) утверждается, что одно и то же положение дел вторично по отношению к самому себе: чтобы субъект  $s$  синтезировал пропозицию *представление ‘\_есть белый’ принадлежит мне* (иными словами, пропозицию *я имею представление ‘\_есть белый’ в своём сознании*), субъект  $s$  должен уже синтезировать эту пропозицию. Но это невозможно: отношение  $\prec$  было определено нами как иррефлексивное.

Рассмотрим **второй пример**. Пусть формула  $\Phi$  задаётся так же, как в первом примере, т. е.  $\Phi =_{\text{df}} P(a)$ , так что истинны (14) и (15). Перепишем (13) для случая  $i = 1, j = 2$ :

$$(18) (\mathbf{T}_s\Phi) \prec (\mathbf{T}_s\{\llbracket\Phi_1\rrbracket\mathcal{E}s\}) \prec [\mathbf{T}_s(\llbracket\Psi_2\rrbracket\mathcal{E}s)].$$

По (14), подставим в (18) вместо  $\Phi$  формулу  $P(a)$ , по (15), вместо  $\Phi_1$  подставим  $P$ , по (11), вместо  $\Psi_2$  подставим  $\epsilon$ , после чего получим:

$$(19) (\mathbf{T}_s[P(a)]) \prec (\mathbf{T}_s\{\llbracket P\rrbracket\mathcal{E}s\}) \prec [\mathbf{T}_s(\llbracket\epsilon\rrbracket\mathcal{E}s)].$$

По Канту, имеется *первый (изначальный)* синтез (см. приведённую выше цитату из § 16, В 131 203.7 - В 132 203.24), который делает доступным для Я элементы, полученные непосредственно из созерцаемого многообразия; этот синтез необходим для *любого* синтеза какого-либо представления из этих элементов. В первом примере представление  $\llbracket P\rrbracket\mathcal{E}s$  является именно таким первым синтезом, необходимым, например, для синтеза представления  $P(a)$ . Но, по (17), синтез  $\llbracket P\rrbracket\mathcal{E}s$  не является *первым*: ему предше-

ствует он сам, т. е. ему предшествует *тот же самый* синтез  $[[P]]_{es}$ . По (19), синтез  $[[P]]_{es}$  также не является *первым*: ему предшествует *другой* синтез  $[[e]]_{es}$ .

Рассмотрим **третий пример**. Пусть формула  $\Phi$  задаётся так же, как в первом и втором примерах, т. е.  $\Phi =_{df} P(a)$ , так что истинны (14) и (15). Перепишем (13) для случая  $i = 1, j = 3$ :

$$(20) (T_s\Phi) \prec (T_s\{[[\Phi_1]]_{es}\}) \prec [T_s([[P_3]]_{es})].$$

По (14), подставим в (20) вместо  $\Phi$  формулу  $P(a)$ , по (15), вместо  $\Phi_1$  подставим  $P$ , по (11), вместо  $\Psi_3$  подставим  $e$ , после чего получим:

$$(21) (T_s[P(a)]) \prec (T_s\{[[P]]_{es}\}) \prec [T_s([[e]]_{es})].$$

На основании каждого из трёх рассмотренных примеров и полученных положений (17), (19), (21) можно сказать, что первый (изначальный) рассудочный синтез – как он понимается Кантом – невозможен.

### Литература

1. *Кант, И.* Критика чистого разума: 2-е издание (В), 1787 г. В *Кант, И. Сочинения.* Т. 2, ч. 1. Под ред. Б. Тушлинга, Н. Мотрошиловой. – М.: «Наука». 2006.
2. *Коффа, А.* Семантическая традиция от Канта до Карнапа: к Венскому вокзалу. Под ред. Линды Весселс. Пер. с англ. В. В. Целищева. М.: Канон+ РООИ «Реабилитация». 2019.
3. *Allison, H. E.* Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy. Cambridge: Cambridge University Press. 1996.
4. *Ameriks, K.* Kant and the Historical Turn: Philosophy as Critical Interpretation. Oxford: Clarendon Press (Oxford University Press). 2006.
5. *Brook, A.* Kant and the Mind. Cambridge: Cambridge University Press. 1994.
6. *Fine, K.* Guide to Ground. // *Metaphysical Grounding.* /Ed. by Fabrice Correia, Benjamin Schnieder. Cambridge: Cambridge University Press. 2012.
7. *Frank, M.* Is Subjectivity a Non-Thing, an Absurdity (Unding)? On Some Difficulties in Naturalistic Reductions of Self-Consciousness. // *The Modern Subject: Conceptions of the Self in Classical German Philosophy.* /Ed. by Karl Ameriks and Dieter Sturma. Albany, NY: State University of New York Press. 1995. P. 177-97.
8. *Frank, M.* Selbstbewußtsein und Selbsterkenntnis. Stuttgart: Reclam. 1991.

9. *Henrich, D.* Selbstverhältnisse. Gedanken und Auslegungen zu den Grundlagen der klassischen deutschen Philosophie. Stuttgart: Reclam. 1982.

10. *Kitcher, P.* Kant's Transcendental Psychology. New York: Oxford University Press. 1990.

11. *Pippin, R. B.* Hegel's Idealism. Cambridge: Cambridge University Press. 1989

12. *Powell, C. T.* Kant's Theory of Self-Consciousness. Oxford: Oxford University Press. 1980.

13. *Rosen, G.* Metaphysical Dependence: Grounding and Reduction. // *Modality: Metaphysics, Logic, and Epistemology*. Ed. by Bob Hale Aviv Hoffmann. Oxford: Oxford University Press. 2010. P. 109-135.

14. *Schaffer, J.* On what grounds what. // *Metametaphysics: New Essays on the Foundations of Ontology*. Ed. by David Manley, David J. Chalmers, Ryan Wasserman. Oxford: Oxford University Press. 2009.P. 347-383.

15. *Strawson, P. F.* The Bounds of Sense. London: Methuen. 1966.

16. *Strawson, P. F.* The Problem of Reason and the A Priori. // *Kant and Contemporary Epistemology*. /Ed. by Paolo Parrini. Dordrecht: Kluwer.1996. P. 167-74.

## References

1. *Kant, I.* (2006). Critique of Pure Reason: 2<sup>nd</sup> edition (B), 1787. In *Kant, I. Works*. Vol. 2, part. 1. Ed. by B. Tuschling, N. Motroshilova. - M.: "Nauka". IX, 1081 pp.

2. *Koffa, A.* (2019). The Semantic Tradition from Kant to Carnap: To the Vienna Station. Ed. by Linda Wessels. Translated from the English by V. V. Tselishev. Moscow: Kanon+ ROOI "Reabilitatiya". 528 pp. (Library of Analytical Philosophy)

3. *Allison, H. E.* (1996). Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy. Cambridge: Cambridge University Press.

4. *Ameriks, K.* (2006). Kant and the Historical Turn: Philosophy as Critical Interpretation. Oxford: Clarendon Press (Oxford University Press). vii, 335 pp.

5. *Brook, A.* (1994). Kant and the Mind. Cambridge: Cambridge University Press.

6. *Fine, K.* (2012). Guide to Ground. // *Metaphysical Grounding*. / Ed. by Fabrice Correia, Benjamin Schnieder. Cambridge: Cambridge University Press. pp. 37-80.

7. *Frank, M.* (1995). Is Subjectivity a Non-Thing, an Absurdity (Unding)? On Some Difficulties in Naturalistic Reductions of Self-Consciousness. // *The Modern Subject: Conceptions of the Self in Classical German Philosophy*. / Ed. by Karl Ameriks and Dieter Sturma. Albany, NY: State University of New York Press. pp. 177-97.

8. *Frank, M.* (1991). *Selbstbewußtsein und Selbsterkenntnis*. Stuttgart: Reclam.
9. *Henrich, D.* (1982). *Selbstverhältnisse. Gedanken und Auslegungen zu den Grundlagen der klassischen deutschen Philosophie*. Stuttgart: Reclam.
10. *Kitcher, P.* (1990). *Kant's Transcendental Psychology*. New York: Oxford University Press.
11. *Pippin, R. B.* (1989). *Hegel's Idealism*. Cambridge: Cambridge University Press.
12. *Powell, C. T.* (1980). *Kant's Theory of Self-Consciousness*. Oxford: Oxford University Press.
13. *Rosen, G.* (2010). *Metaphysical Dependence: Grounding and Reduction*. // *Modality: Metaphysics, Logic, and Epistemology*. / Ed. by Bob Hale Aviv Hoffmann. Oxford: Oxford University Press. pp. 109-135.
14. *Schaffer, J.* (2009). *On what grounds what*. // *Metametaphysics: New Essays on the Foundations of Ontology*. / Ed. by David Manley, David J. Chalmers, Ryan Wasserman. Oxford: Oxford University Press. pp. 347-383.
15. *Strawson, P. F.* (1966). *The Bounds of Sense*. London: Methuen.
16. *Strawson, P. F.* (1994). *The Problem of Reason and the A Priori*. // *Kant and Contemporary Epistemology*. / Ed. by Paolo Parrini. Dordrecht: Kluwer. pp. 167-74.

### **Информация об авторе:**

*Берестов Игорь Владимирович* – доктор философских наук, ведущий научный сотрудник отдела философии Института философии и права Сибирского отделения Российской академии наук, г. Новосибирск, ул. Николаева, 8.

berestoviv@yandex.ru,

### **Information about the author**

*Berestov Igor Vladimirovich* – Doctor of Science in Philosophy, Leading Research Fellow of the Philosophy Department of the Institute of Philosophy and Law (IPL), Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Nikolaeva str., 8.

berestoviv@yandex.ru

Дата поступления 10.04.2025

Принята к печати 11.12.2025

УДК: 165:004.8  
DOI: 10.15372/ PS20250508  
EDN: NXDVRC

**А.А. Шевченко**

**НЕЯВНОЕ ЗНАНИЕ:  
ОТ М. ПОЛАНИ ДО ALPHAZERO**

В статье анализируется концепция неявного знания, введённая М. Полани, и её актуальность в современном научном и технологическом контексте. Рассматривается исторический фон идеи, её эпистемологические последствия и социальное измерение коллективного неявного знания в науке. Особое внимание уделено проблеме «чёрного ящика» в искусственном интеллекте на примере эволюции шахматных алгоритмов от Deep Blue к AlphaZero, демонстрирующих переход от формализованного явного знания к непрозрачному неявному знанию. В статье подчёркивается, что неявное знание сохраняет ключевое значение для понимания природы научного познания и современных когнитивных систем.

*Ключевые слова:* неявное знание, эпистемология, философия науки, искусственный интеллект, «чёрный ящик», коллективное знание, Deep Blue, AlphaZero, эпистемологический сдвиг.

**A.A. Shevchenko**

**TACIT KNOWLEDGE:  
FROM M. POLANYI TO ALPHAZERO**

The article explores Michael Polanyi's concept of tacit knowledge and its relevance in contemporary scientific and technological contexts. It reviews the historical and epistemological background, the social dimension

of collective tacit knowledge in science, and the challenges posed by the "black box" problem in artificial intelligence. Using the evolution of chess algorithms from Deep Blue to AlphaZero as a case study, the article illustrates the shift from explicit, formalized knowledge to opaque, tacit knowledge. It emphasizes the enduring importance of tacit knowledge for understanding the nature of scientific practice and modern cognitive systems.

*Keywords:* tacit knowledge, epistemology, philosophy of science, artificial intelligence, "black box", collective knowledge, Deep Blue, AlphaZero, epistemological shift.

### **Исторический и эпистемологический контекст**

Концепция неявного знания, представленная М. Полани в «Личностном знании» в 1958 г. [2] и детализированная в «Неявном измерении» в 1966 г. [7] долгое время оставалась на периферии аналитической эпистемологии. Уже само название первой монографии было вызовом идеалу объективности, к которому стремилась аналитическая философия, т.к. понятие «личное» воспринималось как синоним субъективного, эмоционального и ненадежного. А ключевое утверждение М. Полани «мы знаем больше, чем можем сказать», ставшее основным тезисом второй монографии [7], бросает вызов традиционным моделям знания как пропозиционального, верифицируемого и формализуемого. Оно указывает на существование познавательного опыта, который не может быть полностью передан в форме текста, правил или формул, такое знание реализуется в действиях, интуиции, практическом мастерстве.

Еще более фундаментально то, что традиционная наука, со времен Просвещения, строилась на идеалах прозрачности, воспроизводимости и объяснимости. Закон всемирного тяготения Ньютона предлагает элегантную и понятную математическую модель, которая не только предсказывает движение планет с высокой точностью, но и описывает силы, его вызывающие. Все параметры модели измеримы, а ее предсказания поддаются экспериментальной проверке. М. Полани же утверждал, что всякая форма знания, даже самая строгая наука, опирается

на неформализуемый пласт компетенций, мастерства и личного участия.

Однако хотя концепция неявного знания М. Полани и оставалась в тени, она все же не была проигнорирована полностью. Во-первых, в ней имеется очевидное сходство с различием Г. Райла между «знанием-как» (knowing-how) и «знанием-что» (knowing-that), изложенным в работах «Понятие сознания» [3] и соответствующей статье [9]. Г. Райл утверждал, что такое умение как, например, езда на велосипеде, не сводится к знанию пропозициональных истин о равновесии и механике. Оно представляет собой автономную форму познания, которая не требует вербализации. Однако в аналитической традиции «знание-как» часто воспринималось как вторичное, производное от «знания-что», и это ограничивало его философскую значимость. М. Полани же претендует на преодоление этой дихотомии, утверждая, что неявное знание – не просто навык, а основание всякого познания. «Хотя неявное знание может существовать само по себе, явное знание должно опираться на неявное понимание и применение. Следовательно, всякое знание либо неявное, либо укоренено в неявном знании. Совершенно явное знание невозможно представить» [8, р. 144]. Это утверждение имеет радикальные последствия: оно ставит под сомнение саму возможность создания полностью формализованной науки, в которой всё знание будет кодифицировано и доступно для логического анализа.

Другой причиной сохранения интереса к идее неявного знания было то, что если для аналитической эпистемологии она была маргинальной, то в философии науки ее влияние было более заметным. Идеи М. Полани имели параллели с работами Томаса Куна, особенно с понятием научной парадигмы, которое усваивается учеными через практику, а не через формальные правила. Ученый, работающий внутри парадигмы, обладает огромным пластом неявного знания о том, как ставить эксперименты и интерпретировать результаты. Это знание не зафиксировано в учебниках, но оно критически важно для функционирования науки [1, гл. 5].

## Неявное знание в науке и его артикуляция

Идеи М. Полани стали по-настоящему востребованы лишь в последнюю четверть века с развитием социальной эпистемологии (которая изучает знание как коллективный феномен), эпистемологии добродетелей (которая смещает фокус с анализа убеждений на анализ познавательных способностей и навыков самого познающего субъекта) и, особенно, в связи с бурным развитием систем искусственного интеллекта, таких как глубокие нейронные сети и большие языковые модели.

Современные ИИ-системы, такие как AlphaFold (предсказание структуры белков) или Kepler (обнаружение экзопланет), в узкоспециализированных задачах демонстрируют эффективность, превосходящую человеческую. Однако их внутренняя логика принятия решений часто остаётся для нас недоступной, что затрудняет понимание, как и почему был получен конкретный результат. Эти модели обучаются на огромных массивах данных и формируют сложные паттерны, корреляции, веса связей, которые невозможно полностью интерпретировать. Это знание существует не в форме правил, а в форме распределённых числовых значений, что делает его аналогом неявного знания, описанного М. Полани.

Этот феномен получил название «проблема чёрного ящика» и породил целую область исследований – объяснимый ИИ (Explainable AI, XAI). Цель XAI – сделать «прозрачными» решения, принимаемые нейросетями, чтобы можно было их проверить, объяснить и доверять им. Однако до сих пор не существует универсального способа декодировать «мысли» глубокой нейронной сети. Это означает, что мы можем доверять результату, но не понимаем метода, что противоречит самой сути научного подхода.

«Неявность» и «непрозрачность» знания порождают сразу несколько проблем. Во-первых, это проблема верификации и доверия. Как мы можем доверять результату, механизм получения которого не понимаем? Наука всегда опиралась на критическую проверку и фальсификацию гипотез. Но как фальсифицировать «интуицию» машины? Мы можем проверить её предсказания эмпирически (например, подтвердить структуру белка в лабора-

тории), но это оставляет сам метод познания в виде «чёрного ящика». Это и есть знание без понимания, что представляет собой кризис самой объяснительной модели, ведь главная цель науки – не только предсказывать, но и понимать мир. Опасность в том, что это может привести к появлению научных теорий, которые работают, но не имеют под собой ясной теоретической базы, понятной человеку. И, наконец, еще одна проблема связана с передачей знания. Если ключевые открытия совершаются с помощью неявного знания ИИ, как обучать следующее поколение ученых? Достаточно ли будет научить их «общаться» с ИИ-системами, задавая правильные вопросы и интерпретируя ответы, или же мы потеряем фундаментальное понимание самой сути научного поиска?

Центральная эпистемологическая проблема заключается в том, может ли неявное знание быть полностью эксплицировано. Как справедливо отмечается в статье, посвященной проблеме обоснования «недокументарного» знания: «Некоторые проблемы, с которыми мы сталкиваемся в связи с неявным знанием, с этой ситуацией «знать больше, чем можешь выразить словами», заключаются в том, можно ли сделать неявное знание явным. Можно ли его кодифицировать как «знание-что» или информацию?» [4, p.97].

### **Социальное измерение: коллективное неявное знание**

Классическая интерпретация концепции М. Полани фокусируется на индивидуальном исследователе, подчеркивая личностный характер неявного знания как интуитивного, неартикулируемого компонента познания. Однако современная наука представляет собой в высшей степени коллективное предприятие, где знания генерируются, передаются и применяются в рамках социальных сетей, лабораторий и междисциплинарных команд. Это ведет к переосмыслению представлений о носителе неявного знания: общепринятым становится представление о коллективном субъекте такого знания, где неявное знание проявляется не только в индивидуальном сознании, но

и в групповых практиках. В науке неявное знание может быть свойством не отдельного разума, а группы как единого целого, воплощенным в принятых научных практиках, распределенном труде и неформальных нормах взаимодействия. «Знание-как» может принадлежать всей лаборатории или организации, даже если ни один ее член по отдельности не способен полностью его изложить.

Это позволяет анализировать сложные научные коллаборации не как простую сумму индивидуальных знаний, а как единую эпистемическую систему. Такое знание служит необходимой основой как для явного знания, так и для самого существования научной практики. Именно фундаментальный характер и неустранимость неявного знания и подчеркивает М. Полани: «Провозглашённая цель современной науки – установить строго отстранённое, объективное знание. Любое отклонение от этого идеала принимается лишь как временный недостаток, который следует стремиться устранить. Но предположим, что неявная мысль является неотъемлемой частью всякого знания, – тогда идеал устранения всех личностных элементов знания фактически будет означать стремление уничтожить всё знание» [8, p. 20].

Социальная трактовка неявного знания радикально расширяет рамки, изначально заданные М. Полани. Такое знание перестает быть исключительно психологическим или феноменологическим свойством индивида. Оно становится организационным принципом устройства не только социальных институтов, в частности института науки, но и общества в целом. Общим примером «простого» коллективного неявного знания может служить езда на велосипеде в потоке трафика, где участники неосознанно координируют свои действия на основе неартикулируемых норм и ожиданий. Такое знание как бы «распределено» в обществе и фактически его организует [6, p. 26-27]. Оно не зафиксировано в правилах, но регулирует поведение. В рамках «социальной» интерпретации неявное знание – это не просто «еще не формализованное» знание, подлежащее экспликации. Напротив, социальный ракурс подчеркивает его принципиальную неформализуемость, поскольку оно встроено в динамику социальных отношений и культурных

контекстов. Именно в социальном измерении неявное знание М. Полани эволюционировало от индивидуальных интуиций к коллективному феномену, интегрирующему эпистемологию с социологией науки. В эпоху глобальных коллабораций неявное знание становится ключом к пониманию того, как различные общества генерируют и поддерживают знания, выходя за пределы индивидуального опыта.

### **Эволюция шахматных алгоритмов: от явного знания к неявному**

Хорошую возможность для более глубокого понимания этих проблем дает эволюция шахматных алгоритмов. Шахматы – идеальная модель для анализа эпистемологических трансформаций, так как эта игра имеет чёткие правила, но бесконечное множество стратегий, что делает её одновременно формализуемой и интуитивной. Переход от систем, основанных на человеческой экспертизе, к системам, обучающимся «с чистого листа», является не просто технологическим скачком, а разворачивающимся эпистемологическим экспериментом, который вновь актуализирует классические проблемы философии науки, касающиеся природы знания, способов его обоснования и передачи. История шахматных алгоритмов прослеживается как переход от явных, формализованных процедур к системам, в которых «знание» материализуется в сложно организованных, но нечётко интерпретируемых весах и аксиомах, что очень похоже на интересующее нас неявное знание.

Важнейшая отправная точка здесь – знаменитая статья Клода Шеннона (1950) [11], в которой он формулирует задачу «как программировать ЭВМ для игры в шахматы». Одно из ключевых утверждений: «Основной тезис нашей работы заключается в том, что современные универсальные компьютеры способны играть в шахматы на вполне приемлемом уровне благодаря использованию соответствующего алгоритма (или «программы») [11, p. 257]. Эта идея легла в основу всей классической шахматной программы. Эта статья задаёт и соответствующую рамку, в которой шахматы можно понимать

как проблему формальной процедуры поиска и оценки позиций. К. Шеннон стремится редуцировать игру к набору алгоритмических правил – оценочным функциям, дереву поиска и эвристикам. Система IBM Deep Blue, победившая чемпиона мира по шахматам в 1997 году, стала вершиной парадигмы программирования, основанной на явном знании. Ее архитектура, несмотря на сложность, была эпистемологически прозрачной. Она опиралась на исчерпывающий перебор вариантов (*brute-force computation*) и оценочную функцию, которая представляла собой сложную систему эвристик, сформулированных гроссмейстерами.

Один из создателей Deep Blue, М. Кэмпбелл, описывал этот процесс как попытку формализовать интуицию экспертов: «Носителем шахматных знаний была оценочная функция. Она включала более 8000 параметров... Мы консультировались с гроссмейстерами, стремясь формализовать их знание в этих параметрах. Это, например, параметры «безопасность короля» или «ладья на открытой вертикали». При формировании оценочной функции для расширенного описания учитывается ряд факторов, в том числе: частота или регулярность применения хода гроссмейстерами» [5, p. 67].

Таким образом, Deep Blue представлял собой систему, эпистемологический статус которой был полностью прозрачен. Его «знание» было совокупностью решений, внедренных человеком. Он не обладал автономным пониманием, а лишь с огромной скоростью применял заданные ему инструкции. Объяснение хода Deep Blue, в принципе, соответствовало дедуктивно-номологической модели научного объяснения Карла Гемпеля. Согласно этой модели, объяснить феномен – значит подвести его под общий закон. В случае Deep Blue, «законами» выступали множество правил его оценочной функции (например, «ценность контроля над центром равна X», «безопасность короля равна Y»), а «начальными условиями» – конкретная позиция на доске. Ход дедуктивно выводился как оптимальный с точки зрения максимизации значения этой функции. Проблема была в сложности, а не в принципиальной непознаваемости. Deep Blue, таким образом, представлял собой попытку преобразовать неявное знание гроссмейстеров (стратегическую интуи-

цию) в явные параметры (8000 формальных правил). Эта попытка, однако, демонстрирует и пределы формализации, даже 8000 параметров не смогли полностью «ухватить» шахматную интуицию (что доказал более поздний разгром Deep Blue версиями ИИ нового поколения).

Появление AlphaZero в 2017 году стало радикальным разрывом с предыдущей парадигмой. В отличие от своих предшественников, AlphaZero не использовала человеческие эвристики или базы данных. Ее архитектура, основанная на обучении с подкреплением (reinforcement learning) и самоигре позволила ей приобрести знание «с чистого листа». Как сообщали разработчики AlphaZero в основополагающей статье: ««AlphaGo Zero обучается с чистого листа (*tabula rasa*), исключительно за счёт самоигры, без использования человеческих данных, экспертных знаний или каких-либо априорных сведений, кроме базовых правил игры» [12, р. 354]. Знание AlphaZero не существует в виде набора артикулируемых правил. Система демонстрирует глубокое позиционное «понимание», но это понимание не является дискурсивным. Оно представляет собой очень эффективное «знание-как» без соответствующего «знания-что» в его пропозициональной форме. Фактически, это искусственный аналог неявного знания М. Полани.

Такое знание стало примером того, что в современной эпистемологии называется «эпистемической непрозрачностью» (epistemic opacity). Непрозрачность AlphaZero не является просто следствием масштаба (хотя он и важен). Она является следствием природы его знания, которое не представлено в виде дискретных правил, а распределено по множеству числовых весов нейронной сети.

Авторы упомянутой статьи в *Nature* прямо заявляют, что AlphaZero достигла сверхчеловеческого уровня в шахматах за относительно короткий тренировочный период. Ее знание представляет собой содержательное, стратегическое понимание, близкое к человеческой интуиции. Оно не может быть передано в виде инструкций, но может быть усвоено через подражание, как в традиционной передаче неявного знания. Анализ стиля или манеры игры показывает, что стиль игры AlphaZero тоже сильно отличался от «традиционных» представлений о «правильной» игре:

неожиданные жертвы материала в начале игры, рассчитанные на долгосрочное и весьма неочевидное позиционное преимущество, удивительное чувство динамики, что шахматисты часто называют интуицией или мастерского «чутьем». При этом мы имеем дело с интуицией, не сводимой к алгоритмам, что, опять же, позволяет говорить о появлении некоторого искусственного аналога неявного знания [10].

AlphaZero демонстрирует, что «неявный» характер знания может быть воспроизведён аппаратно и без явного человеческого вмешательства. Возникает вопрос: чем отличается неявное, «человеческое» знание, описанное М. Полани от «непрозрачного» знания шахматного алгоритма? Формальных отличий может быть очень много; но в действии оба вида знания проявляют себя похожим образом, демонстрируя способность к узнаванию паттернов и корреляций, гибкость и «чувство» контекста. Переход от Deep Blue к AlphaZero представляет собой не просто технологический прогресс, а фундаментальный эпистемологический сдвиг, делающий аппарат М. Полани центральным для понимания природы и ограничений современного ИИ. В эпоху тотальной цифровизации неявное знание до недавнего времени казалось одним из очевидных свидетельств человеческой уникальности. Однако быстрая эволюция искусственных систем неявного знания заставляет думать о поиске новых форм симбиоза человека и машины, где неявное знание человека и вычислительная мощь искусственного интеллекта дополняют друг друга.

### **Эпистемологические последствия и перспективы анализа неявного знания**

Эволюция шахматных алгоритмов от Deep Blue к AlphaZero демонстрирует не просто технологический прогресс, а изменение самой логики построения и функционирования когнитивных систем. Переход от явного, заложенного вручную знания к знанию, выработанному автономно в процессе обучения, сопровождается изменением его эпистемологического статуса. В системах вроде AlphaZero знание становится непрозрачным, рас-

пределённым и несводимым к пропозициональным правилам, что делает его функциональным аналогом неявного знания в понимании М. Полани.

Этот сдвиг имеет глубокие последствия. Во-первых, он порождает проблему объяснимости. Непрозрачность системы делает невозможным прямой анализ ее «рассуждений», что ставит под вопрос традиционные научные критерии верификации. Во-вторых, он изменяет саму модель передачи знаний от ИИ к человеку. Если из Deep Blue можно было бы (теоретически) извлечь набор правил, то у AlphaZero можно лишь учиться, как у мастера-человека, что является классическим примером передачи неявного знания. Бывший чемпион мира Владимир Крамник, анализируя партии AlphaZero, подчеркивал именно этот аспект.

Концепция неявного знания, предложенная М. Полани, оказывается полезным инструментом для анализа этих трансформаций. Она помогает понять, почему попытки полной формализации и прозрачности сталкиваются с фундаментальными ограничениями – как в человеческом познании, так и в искусственном. Таким образом, концепция неявного знания не теряет своей значимости даже в контексте современных технологий. Она указывает на то, что даже в эпоху цифровизации и автоматизации познание продолжает оставаться практикой, в которой формальные правила и неформальные компетенции неразрывно связаны.

### Литература

1. Кун Т. Структура научных революций. – М.: Прогресс, 1977. – 608 с
2. Полани М. Личностное знание. На пути к посткритической философии. – М.: Прогресс, 1985. – 344 р.
3. Райл Г. Понятие сознания. – М.: Идея-Пресс, 1999. – 408 с
4. Burns C.S. What Documents Cannot Do: Revisiting Michael Polanyi and the Tacit Knowledge Dilemma // Information & Culture. 2021. Vol. 56. № 1. P. 90–104.
5. Campbell M. Knowledge Discovery in Deep Blue // Communications of the ACM. 1999. Vol. 42. № 11. P. 65–67.

6. *Collins H., Evans R.* Rethinking Expertise. – Chicago: University of Chicago Press, 2007. 160 p.
7. *Polanyi M.* The Tacit Dimension. – N.Y.: Doubleday & Company, 1966. 104 p.
8. *Polanyi M.* Knowing and Being. – Chicago: University of Chicago Press, 1969. 264 p.
9. *Ryle G.* Knowing How and Knowing That // Proceedings of the Aristotelian Society. 1945. Vol. 46. P. 1–16.
10. *Sadler M., Regan N.* Game Changer. – London: Batsford, 2019. 416 p.
11. *Shannon, C. E.* (1950). "Programming a Computer for Playing Chess", Philosophical Magazine, Series 7, Vol. 41, No. 314, pp. 256–275.
12. *Silver D. et al.* Mastering the Game of Go without Human Knowledge // Nature. 2017. Vol. 550. № 7676. P. 354–359.

### References

1. *Kuhn, T.* (1977). *Struktura nauchnykh revolyutsiy* [The Structure of Scientific Revolutions]. Moscow, Progress Publ. (In Russ.). 608 p.
2. *Polanyi, M.* (1985). *Lichnostnoe znaniye: Na puti k postkriticheskoj filosofii* [Personal Knowledge: Towards a Post-Critical Philosophy]. Moscow, Progress Publ. (In Russ.). 344 p.
3. *Ryle, G.* (1999). *Ponyatie soznaniya* [The Concept of Mind]. Moscow, Idea-Press Publ. (In Russ.). 408 p.
4. *Burns, C. S.* (2021). What Documents Cannot Do: Revisiting Michael Polanyi and the Tacit Knowledge Dilemma // Information & Culture. No. 56 (1). P. 90–104.
5. *Campbell, M.* (1999). Knowledge Discovery in Deep Blue // Communications of the ACM. Vol. 42. No. 11. P. 65–67.
6. *Collins, H., Evans, R.* (2007). Rethinking Expertise. Chicago, University of Chicago Press. 160 p.
7. *Polanyi, M.* (1966). The Tacit Dimension. N.Y., Doubleday & Company. 104 p.
8. *Polanyi, M.* (1969). Knowing and Being. Chicago, University of Chicago Press. 264 p.
9. *Ryle, G.* (1945). Knowing How and Knowing That // Proceedings of the Aristotelian Society. Vol. 46. P. 1–16.
10. *Sadler, M., Regan, N.* (2019). Game Changer. London, Batsford. 416 p.

11. *Shannon, C. E.* (1950). Programming a Computer for Playing Chess // *Philosophical Magazine*. Ser. 7. Vol. 41. No. 314. P. 256–275.

12. *Silver, D. et al.* (2017). Mastering the Game of Go without Human Knowledge // *Nature*. Vol. 550. No. 7676. P. 354–359.

### **Информация об авторе**

*Шевченко Александр Анатольевич.* – Институт философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8).

shev@philosophy.nsc.ru

### **Information about the author**

*Shevchenko, Aleksandr Anatolyevich.* – Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090, Russia).

shev@philosophy.nsc.ru

Дата поступления 12.10.2025

Принята к печати 11.12.2025

УДК: 141: 16: 004.5  
DOI: 10.15372/PS20250509  
EDN: OAZLRQ

**А.С. Зайкова**

**О СОИЗМЕРИМОСТИ ВОПРОСОВ  
«МОЖЕТ ЛИ МАШИНА МЫСЛИТЬ?»  
И «МОЖЕТ ЛИ МАШИНА ПОБЕДИТЬ  
В ИМИТАЦИОННОЙ ИГРЕ?»**

В статье проводится критический анализ соизмеримости двух центральных вопросов в философии искусственного интеллекта: «Может ли машина мыслить?» и «Может ли машина победить в имитационной игре?». Показано, что, несмотря на прагматическую замену, предложенную Аланом Тьюрингом в 1950 году, вопросы концептуально и онтологически различны. На основе анализа аргументов Д. Деннета, Дж. Сёрла, Д. Чалмерса и др. демонстрируется, что соизмеримость возможна лишь в ограниченном, индуктивном смысле, но не в логическом или онтологическом. Подчеркивается необходимость различения поведенческой симуляции и подлинного мышления.

*Ключевые слова:* соизмеримость, тест Тьюринга, машинное мышление, сознание, имитационная игра, философия ИИ, эпистемология, поведенческий критерий.

**A.S. Zaykova**

**ON THE COMMENSURABILITY OF THE QUESTIONS  
“CAN A MACHINE THINK?”  
AND “CAN A MACHINE WIN THE IMITATION GAME?”**

This paper offers a critical analysis of the commensurability between two foundational questions in the philosophy of artificial intelligence: “Can a machine think?” and “Can a machine win the imitation game?”

It is argued that, despite Alan Turing's pragmatic substitution proposed in 1950, these questions differ in ontological status and conceptual framework. Drawing on the works of D. Dennett, J. Searle, D. Chalmers, and others, the paper demonstrates that commensurability is possible only in an epistemologically limited, inductive sense—but not in a logical or ontological one. The study emphasizes the necessity of distinguishing behavioral simulation from genuine thinking, thereby avoiding the category mistake often associated with interpreting the Turing Test as proof of machine consciousness.

*Keywords:* commensurability, Turing Test, machine thinking, consciousness, imitation game, philosophy of AI, epistemology, behavioral criterion.

## 1. Введение

Понятие соизмеримости в эпистемологии и философии науки чаще всего понимается как возможность сравнения, перевода или сопоставления двух теорий, концепций или вопросов на основании общей онтологии, методологии или семантической базы. Такое понимание было выработано во многом благодаря Т. Куру и его работе «Структура научных революций» [8], где он ввёл идею несоизмеримости научных парадигм, утверждая, что смена парадигм влечёт за собой трансформацию самого языка науки, критериев значимости и даже восприятия данных. П. Фейерабенд в своих работах развил эту идею, показав, что между конкурирующими теориями часто отсутствует нейтральный метаязык, позволяющий их объективно сравнить [6].

В этом контексте соизмеримость предполагает наличие общего концептуального пространства, в котором можно оценивать истинность, адекватность или объяснительную силу альтернативных подходов без необходимости радикального перевода или пересмотра базовых предпосылок.

В общем случае, мы можем выделить четыре основные критерии соизмеримости, выведенные из классических споров о сопоставимости теорий:

(1) они направлены на один и тот же онтологический объект или свойство (например, мышление как когнитивный феномен);

(2) они встроены в одну и ту же концептуальную рамку, не требующую радикального переопределения ключевых понятий;

(3) они используют совместимые эпистемологические критерии (например, оба полагаются на поведенческие данные или оба требуют доступа к внутренним состояниям);

(4) их ответы логически связаны: положительный (или отрицательный) ответ на один из них влечёт за собой определённый статус другого.

Если все четыре условия выполняются, мы, как правило, можем говорить о полной соизмеримости; если только одно или несколько – о частичной или ограниченной соизмеримости. В последнем случае вопросы могут быть связаны прагматически или индуктивно. Именно такой ограниченный тип связи, как будет показано далее, существует между вопросами «*Может ли машина мыслить?*» и «*Может ли машина победить в имитационной игре?*». Настоящая статья ставит своей задачей не просто констатировать различие этих вопросов, но систематически проанализировать степень и характер их соизмеримости через призму указанных четырёх критериев.

## 2. «Может ли машина мыслить?»

Традиционный вопрос «*Может ли машина мыслить?*» апеллирует к концептам сознания, намерения, субъективного опыта, что, применительно к машинам (техносубъектам, интеллектуальным агентам, нейросетям, роботам, большим языковым моделям и др.), выглядит как откровенное лукавство. Даже сама постановка вопроса применительно к современным интеллектуальным агентам подвергается критике, не столько из-за технических возможностей машин, сколько из-за неопределённости самого понятия «мышление». В классической философии мышление связывается с наличием внутреннего мира, интенциональности и способности к рефлексии – качествами, которые невозможно или почти невозможно приписать вычислительным системам без радикального переосмысления этих понятий.

Кроме того, этот вопрос близок к понятию агентности. В философии действия и этике агентность связывается с автономией воли, моральной ответственностью и свободой выбора, что ставит под сомнение возможность корректно рассуждать об агентности искусственных интеллектуальных систем. В научном сообществе обсуж-

дается необходимость рассмотрения новых видов агентности (интеллектуальной, рациональной, социальной и др.) [1], однако такие обсуждения направлены на решение иных научных задач, к примеру, для описания функциональных ролей ИИ в социальных системах.

В любом случае, задавая вопрос о мышлении машин, мы сталкиваемся либо с невозможностью приписать им мышление в строгом смысле, либо с необходимостью радикального пересмотра самого понятия мышления. В этом отношении вопрос оказывается скорее постановкой проблемы пределов понятий, нежели вопросом о фактических возможностях вычислительных систем. Он выявляет напряжение между традиционной антропоцентричной концепцией разума и новыми формами когнитивной активности, возникающими в искусственных интеллектуальных системах.

Одним из сторонников второго подхода (пересмотра понятия мышления) можно назвать Д. Деннета. Он утверждает, что приписывание ума другому агенту – это стратегия интерпретации, а не метафизическое утверждение [5]. С этой точки зрения, если машина демонстрирует последовательное, адаптивное и целенаправленное поведение, рационально приписывать ей «верования» и «желания», даже если мы не знаем, есть ли у неё субъективный опыт. Однако такая позиция не устраняет онтологическую пропасть между симуляцией и реализацией мышления; она лишь предлагает прагматический способ взаимодействия с когнитивно сложными системами.

### 3. «Может ли машина победить в имитационной игре?»

Идея замены вопроса о мышлении на вопрос об имитации была впервые предложена А. Тьюрингом в его знаменитой статье «Вычислительные машины и разум» («Computing Machinery and Intelligence»), опубликованной в 1950 году в журнале *Mind*. Тьюринг начинает статью с признания трудности исходного вопроса: «Я предлагаю рассмотреть вопрос: „Могут ли машины мыслить?“ Следовало бы начать с определений значений слов „машина“ и „мыслить“» [12, р. 433]. Признавая, что дебаты о значении слова «мыслить» ведут в тупик, он предлагает «заменить вопрос другим, который тесно с ним связан и выражен относительно недвусмысленными словами [Там же].

Этот новый вопрос основан на модифицированной версии игры, которую Тьюринг назвал «имитационной». В оригинальной версии игры участвуют три человека: мужчина (А), женщина (В) и судья (С), который должен по переписке определить, кто из А и В – женщина. Тьюринг предлагает заменить одного из игроков машиной. Если судья не может надёжно отличить машину от человека, машина считается прошедшей тест.

Целью Тьюринга было не дать определение мышления, а предложить операционализируемый критерий, пригодный для эмпирической проверки. Он писал: «Исходный вопрос „Могут ли машины мыслить?“, по-моему, слишком лишён смысла, чтобы заслуживать обсуждения» [12, р. 442]. Вместо этого он сосредоточился на поведенческой эквивалентности: если машина ведёт себя так, как мы ожидали бы от мыслящего существа, то нет рациональных оснований отрицать её способность мыслить.

Тем не менее, Тьюринг не утверждал прямо, что победа в имитационной игре доказывает наличие сознания или понимания. Он рассматривал тест как прагматическую замену вопросу, на который мы всё равно не способны дать ответ. Однако для обоснования правомерности замены одного вопроса другим он использовал ряд аргументов, основным из которых является методологический: в отсутствие прямого доступа к внутреннему опыту (даже у других людей) мы вынуждены полагаться на внешние проявления. Таким образом, имитационная игра становится инструментом для обхода метафизических трудностей и перехода к эмпирически проверяемым гипотезам.

Особый интерес в этом контексте представляет ответно-зависимая интерпретация теста, развитая Д. Праудфут. Согласно её позиции, наиболее правильной интерпретацией теста Тьюринга соответствует следующее суждение:

*«х является разумным (или мыслит), если в неограниченной компьютерной игре-имитации-человека х кажется разумным среднему опрашиваемому» / «субъекту» [10, р. 397–398].*

Эта интерпретация подчёркивает социокультурный аспект атрибуции разумности и рассматривает успешное прохождение теста как результат восприятия, а не объективного факта. Если система кажется разумной среднему субъекту, она таковой признаётся, независимо от её внутренней структуры.

Подобный подход, хоть и не является широко признанным, сближает имитационную игру с более широкими дискуссиями о социальной агентности, где решающим оказывается не то, «обладает ли» субъект внутренней автономией, но то, как он воспринимается другими [2]. Сама возможность машины пройти тест фиксирует её социальное присутствие и способность занимать в коммуникации ту же позицию, что и человек. Однако это социальное признание не обязательно влечёт за собой признание когнитивных характеристик машины или субъективности её опыта.

#### **4. Соизмеримость вопросов**

Для анализа соизмеримости вопросов «*Может ли машина мыслить?*» и «*Может ли машина победить в имитационной игре?*» рассмотрим наиболее яркие позиции в философии сознания и искусственного интеллекта.

В прикладной когнитивной науке и робототехнике часто используется градуалистский подход: вместо утверждения, что что-то «мыслит» или «не мыслит», вводятся уровни когнитивной сложности. В этом контексте успешное прохождение имитационной игры может рассматриваться как маркер высокого уровня когнитивной симуляции, даже если он не гарантирует наличия сознания. Такой подход, однако, требует чёткого разграничения между симуляцией, репликацией и реализацией когнитивных процессов [9].

Ещё более радикальной является позиция Д. Чалмерса, согласно которой мы должны различать «лёгкие» и «трудные» проблемы сознания [4]. Поведенческие задачи, такие как имитационная игра, относятся к «лёгким» проблемам: они касаются функций, которые в принципе могут быть воспроизведены вычислительно. Однако «трудная» проблема – почему и как возникает субъективный опыт – остаётся за пределами поведенческого анализа. С этой точки зрения, даже идеально прошедшая тест машина может быть философским зомби – существом, идентичным человеку по поведению, но лишённым сознания. Это означает, что поведенческий критерий логически не влечёт наличия мышления, и, следовательно, вопросы остаются несоизмеримыми даже в ослабленном смысле.

Хотя аргументы Дж. Сёрла [11] и Н. Блока [3] демонстрируют принципиальную возможность поведенческой имитации без наличия понимания или сознания, некоторые исследователи настаивают на эпистемологической ценности теста Тьюринга даже при отсутствии онтологической гарантии. К примеру, Р. Френч в статье «Тест Тьюринга: первые пятьдесят лет» признаёт, что прохождение теста не доказывает наличие мышления в строгом смысле, однако подчёркивает: «Прохождение теста Тьюринга не доказало бы, что машина мыслит, но предоставило бы очень веские основания полагать, что это так» [7, р. 118]. Френч утверждает, что критики, такие как Сёрл, ошибочно требуют дедуктивной достоверности от теста, в то время как в науке преобладают индуктивные обобщения, основанные на наблюдаемых коррелятах. С его точки зрения, устойчивая способность машины вести осмысленный, контекстуально адаптивный диалог – особенно в расширенной версии теста, включающей эмоциональные, социальные и сенсомоторные аспекты – создаёт высокую степень правдоподобия наличия когнитивных способностей, сопоставимых с человеческими. Эта позиция не отождествляет имитацию с мышлением, но признаёт её как сильный эпистемологический индикатор в условиях недоступности прямого доступа к внутреннему опыту.

Теперь, когда мы упомянули наиболее яркие позиции, перейдём к оценке соизмеримости по четырём критериям, сформулированным во введении.

(1) Онтологическая направленность. Вопрос «Может ли машина мыслить?» обращён к мышлению как к внутреннему, интенциональному состоянию, связанному с сознанием и субъективным опытом. Вопрос об имитационной игре, напротив, касается исключительно поведенческой компетентности в лингвистическом взаимодействии. Хотя Тьюринг предполагает, что поведенческая неотличимость может служить основанием для приписывания мышления, сама игра не предполагает и не требует наличия внутреннего мира. Следовательно, онтологические объекты различны: первый вопрос касается сущности, второй – функции.

(2) Концептуальная рамка. Вопрос о мышлении предполагает традиционную философскую концепцию разума, восходящую к картезианской модели субъекта, где мышление неразрывно связано с рефлексией и внутренней автономией. Имитационная игра,

напротив, опирается на бихевиористскую и функционалистскую парадигму, где значение определяется через использование, а разум – через внешнее поведение. Эти концептуальные рамки конфликтуют: одна требует внутреннего наблюдателя, другая отвергает его как ненаучного.

(3) Эпистемологические критерии. Вопрос о мышлении сам по себе не предлагает операциональных средств проверки и остаётся открытым для метафизических интерпретаций. Вопрос об имитационной игре предлагает чёткий поведенческий критерий. Однако оба вопроса сталкиваются с общей эпистемологической проблемой – «проблемой других умов»: как приписать мышление тому, чей внутренний мир принципиально недоступен? В этом узком смысле критерий частично выполняется: оба вопроса разделяют одну и ту же эпистемологическую трудность и опираются на аналогичную логику индуктивного вывода от поведения к внутреннему состоянию.

(4) Логическая связь ответов. Критика Сёрла и Блока показывает, что успех в имитационной игре не влечёт логически наличия мышления. Но и обратное утверждение также не является бесспорным: для прохождения теста недостаточно, чтобы некое существо мыслило, также необходимо, чтобы оно обладало коммуникативными навыками и возможностями к поддержанию текстового диалога. Таким образом, между вопросами существует неполная асимметричная логическая связь: мышление влечёт возможность имитации (которая реализуема только при добавлении коммуникативных функций), но имитация не влечёт необходимость мышления.

Из четырёх предложенных условий соизмеримости частично выполняются два: критерий совместимости эпистемологических оснований (3) и критерий асимметричной логической связи (4). Это позволяет говорить об ограниченном, эпистемологически опосредованном характере соизмеримости исследуемых вопросов. Вопросы связаны через общую проблему атрибуции мышления и асимметричную импликацию от мышления к поведению, тем не менее, их нельзя назвать ни тождественными, ни взаимозаменяемыми. Такая частичная соизмеримость достаточна для прагматического использования теста Тьюринга как индуктивного инструмента, но недостаточна для онтологического отождествления имитации и мышления.

## 5. Заключение

Проведенный анализ позволяет сделать вывод о несоизмеримости двух вопросов. Конечно, между ними существует асимметричная связь, но она обусловлена не логикой, а условиями человеческого познания. Успешная имитация не требует наличия мышления, но, в условиях принципиальной недоступности внутренних состояний других агентов – включая как людей или животных, так и машин – она повышает правдоподобие наличия когнитивных способностей. Эта связь, однако, остаётся строго индуктивной и вероятностной: она опирается на аналогию и прагматическую полезность, а не на логическое следование или онтологическую необходимость.

Кроме того, хотя поведенческая имитация не является достаточным условием мышления, мышление является достаточным условием для возможности имитации при подключении коммуникативных функций. Любой мыслящий субъект, обладающий необходимыми навыками, в принципе способен участвовать в имитационной игре и проходить её. Эта асимметрия указывает на то, что второй вопрос является необходимым, но не достаточным следствием первого. Таким образом, соизмеримость ограничена направлением: только от мышления к имитации; но не от имитации к мышлению.

Следовательно, даже в ослабленной форме полная соизмеримость не достигается. Вопросы остаются разнородными по своей эпистемологической природе: один принадлежит метафизике сознания, другой – методологии поведенческой оценки. Их сопоставление возможно лишь при явном указании на принимаемые методологические и метафизические допущения – например, на готовность отождествлять функциональную эквивалентность с когнитивной тождественностью.

Признание этого факта не отменяет ценности теста Тьюринга как инструмента. Он остаётся важным эталоном для оценки лингвистической компетентности, социальной адаптивности и коммуникативной сложности ИИ-систем. Однако его онтологическая переинтерпретация, подразумевающая подмену поведения сознанием, а симуляции – мышлением, ведёт к категориальной ошибке, особенно опасной сейчас, в эпоху генеративных моделей, способ-

ных создавать иллюзию понимания без его наличия. Различение между «кажется, что мыслит» и «действительно мыслит» является необходимым условием как для теоретических, так и для практических исследований в области искусственного интеллекта.

### Литература

1. *Зайкова, А. С.* Агентность систем искусственного интеллекта: интеллектуальная, социальная, моральная // *Философия науки*. 2024. № 5(104). С. 89–97. <https://doi.org/10.15372/PS20240507>
2. *Зайкова А.С.* Можно ли относиться к роботам и к иным техносубъектам как к животным: к вопросу о социальной агентности // *Семиотические исследования*. 2025. 5(3). С. 16–23.
3. *Block, N.* Psychologism and behaviorism // *The Philosophical Review*. 1981 90(1). P. 5–43.
4. *Chalmers, D. J.* (Facing up to the problem of consciousness // *Journal of Consciousness Studies*. 1995. No. 2(3). P. 200–219.
5. *Dennett, D. C.* The intentional stance. 1987. MIT Press.
6. *Feyerabend, P.* Against method: Outline of an anarchistic theory of knowledge. New Left Books. 1975
7. *French, R. M.* The Turing Test: The first fifty years // *Trends in Cognitive Sciences*. 2000. № 4(3). P. 115–122.
8. *Kuhn, T. S.* The Structure of Scientific Revolutions. University of Chicago Press. 1962.
9. *Preston, J.* The philosophy of artificial intelligence. // E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.). URL: <https://plato.stanford.edu/archives/fall2013/entries/artificial-intelligence/>
10. *Proudfoot D.* Rethinking Turing’s Test and the Philosophical Implications // *Minds and Machines*. 2020. 30. P. 487–512.
11. *Searle, J. R.* Minds, brains, and programs // *Behavioral and Brain Sciences*. 1980. 3(3). P. 417–457.
12. *Turing, A. M.* Computing machinery and intelligence // *Mind*. 1950. 59(236). P. 433–460.

### References

1. *Zaykova, A. S.* (2024). Agency of artificial intelligence systems: Intellectual, social, moral. *Philosophy of Science*, S5(104), 89–97. (In Russ.).
2. *Zaykova, A. S.* (2025). Can We Treat Robots and Other Technosubjects as Animals? On the Question of Social Agency. *Semiotic Studies*, 5(3), 16–23. (In Russ.).

3. *Block, N.* (1981). Psychologism and behaviorism. *The Philosophical Review*, 90(1), 5–43.
4. *Chalmers, D. J.* (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
5. *Dennett, D. C.* (1987). *The intentional stance*. MIT Press.
6. *Feyerabend, P.* (1975). *Against method: Outline of an anarchistic theory of knowledge*. New Left Books.
7. *French, R. M.* (2000). The Turing Test: The first fifty years. *Trends in Cognitive Sciences*, 4(3), 115–122.
8. *Kuhn, T. S.* (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
9. *Preston, J.* (2013). The philosophy of artificial intelligence. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 ed.). URL: <https://plato.stanford.edu/archives/fall2013/entries/artificial-intelligence/>
10. *Proudfoot D.* (2020). Rethinking Turing's Test and the Philosophical Implications. *Minds and Machines* 30: 487–512.
11. *Searle, J. R.* (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457.
12. *Turing, A. M.* (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.

### **Информация об авторе**

*Зайкова Алина Сергеевна* – научный сотрудник Института философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8).  
[Zaykova.a.s@gmail.com](mailto:Zaykova.a.s@gmail.com)

### **Information about the author**

*Zaykova Alina Sergeevna* – researcher of Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090, Russia).  
[Zaykova.a.s@gmail.com](mailto:Zaykova.a.s@gmail.com)

Дата поступления 02.03.2025  
Принята к печати 11.12.2025

УДК: 167.7

DOI: 10.15372/PS20250510

END: FORLMV

**А.В. Хлебалин**

### **ОБОСНОВЫВАЮЩАЯ ПРАКТИКУ ЛОГИКИ**

В статье рассматривается противостояние практико-ориентированного и классического подходов в философии логики. Анализируется критика классической традиции сторонниками практико-ориентированного подхода. Анализируются перспективы практического поворота в философии логики.

*Ключевые слова:* логика, практика, априорное знание, обоснование в логике

**A.V. Khlebalin**

### **JUSTIFYING THE PRACTICE OF LOGIC**

This article examines the opposition between the practice-oriented and classical approaches in the philosophy of logic. It analyzes the critique of the classical tradition by proponents of the practice-oriented approach. The prospects for a practical turn in the philosophy of logic are explored.

*Keywords:* logic, practice, a priori knowledge, justification in logic

Развитие математической логики в процессе преодоления кризиса оснований математики было одним из ярчайших событий начала XX в. Ее возникновение и стремительное развитие не столько оказалось решением проблемы природы логики, остро вставшей на рубеже веков в связи с кризисом кантианства,

сколько перевело это вопрос в более сложный контекст. Многие вопросы как философии логики, как например, о нормативности логики или ее плюралистичности, и так и более универсальные философские проблемы, вроде природы априорного знания, все чаще увязываются с проблемами природы логики как таковой. Все это происходит на фоне все возрастающего многообразия логик. Говорить о какой-то логике в некоем универсальном смысле уже невозможно; необходимо уточнение, о какого рода логике или семействе логик идет речь.

Логическая истина неразрывно связано историей философии с концепцией аналитической истины, для решения любых эпистемологических вопрос в отношении которой действительность, в любой ее форме – будь то контингентная физическая реальность и тем более социальные факторы, не могут иметь существенного значения. Разумеется, история логики окружена научным фольклором не в меньшей степени, чем любая дисциплина, но социальный контекст развития логики совершенно бесполезен для любых серьезных вопрос о природе логического знания. Логическое знание, упрощенно, выступает образцом платоновского постижения истины – индивидуальный ум схватывает вечные истины, а структура постоянно растущего логического знания эксплицировано может быть средствами той же логики. Этот карикатурно представленный подход породил классические результат, значимость которых превзошла сферу исключительно логики. Этот подход господствует и в философии логики, оказываясь невероятно плодотворным.

Но в последние годы эта монополия была оспорена. Речь идет о так называемом практик-ориентированном подходе к логике. Обращение к «практике» как основному предмету интереса в философии науки совершается изначально в философии математики и философии науки в целом. Сама концепция практики очень широко понимается и не может быть схвачено какой-либо дефиницией, которая бы устроила интересантов. Отправным пунктом практико-ориентированной философии – философии логики, в том числе, – отрицание платонистских допущений в отношении объекта исследования. Вслед за этим, следует отрицание возможности познания объекта априорным рассуждением. Согласно практико-ориентированному подходу, логика не имеет ка-

кого-то особого статуса среди других наук. Основной причиной обращения к практике является неудовлетворенность в высшей степени абстрактными результатами классического подхода в философии логики.

Традиционные подходы, ориентированные в большей степени на экспликацию средствами логики структуры и семантики научной теории, с точки зрения сторонника практико-ориентированного подхода, отличаются следующими принципиальными характеристиками, оказывающимися недостатками с отстаиваемой ими точки зрения. Прежде всего, такие экспликации всегда оказываются излишне идеализированными, поскольку получены они посредством априорных размышлений, в основе которых лежат некие нормативные представления о том, что какой должна быть научная теория, игнорирующие реальную практику научного исследования, составляющие которое представлялись нерелевантными по отношению к нормативному образцу. Отсюда, такие нормативные модели научной теории и способов ее формулировки, проверки и пр., оказываются чрезмерно упрощенными, поскольку не отражали многообразия целей научной деятельности и плюрализм методологий, формирующих научную деятельность. Традиционный подход преимущественно ориентирован на «конечный продукт», прежде всего, научную теорию, которая сама по себе является весьма абстрактным, логическими средствами реконструированным, объектом, и тем самым квалифицирует как акцидентальные средства и процессы, которые привели к получению результатов, включая коллективный характер научной деятельности. Наконец, нужно отметить весьма органично вписывающееся в намеченную характеристику традиционного подхода, свойственное ему пренебрежение концептуальной историей дисциплины, которое трактует концепции и понятия как неизменный набор средств, хранящихся на неких теоретических «платоновских небесах» в неименном и однозначно определенном виде. За характеризующимся этими методологическими принципами подходом закрепилось название «нисходящего» (видимо, с «платоновских небес») подхода. [1; 80].

Все более становящийся явной искусственность традиционно, «восходящего», подхода, которая помимо упомянутых налагаемых методологических ограничений, требовала дисквалификации

многих примеров элементов научной деятельности как девиантным по причине и несоответствия абстрактным образцам нормативных примеров, обнаружилась в полемике по проблеме научного объяснения. Традиционный подход, воплощенный, например, в классических работах К. Гемпеля, явно обнаруживает перечисленные характеристики. Его оппоненты, настаивающие на необходимости если не отказа от самого эссенциалистского подхода, то на ослаблении его нормативной составляющей, предложили «восходящий подход», основанный на анализе конкретных примеров практики объяснения в различных локальных предметных областях науки, что позволило обнаружить широкий спектр форм объяснений, распространенных в биологических [2], медицинских [3] и физических науках [4]. Поразительным является то, насколько стремительно критика традиционных эссенциалистских моделей объяснения науки в рамках подхода, основанного на практике, укоренилась в современной литературе по проблеме научного объяснения. Можно утверждать, что анализ научного объяснения стал тем полигоном, на котором оформлялся и впервые был опробован практико-ориентированный подход в философии науки.

Практико-ориентированный подход явно заявляет о двух преимуществах перед традиционными «нисходящими» подходами [см. 4; 5; 6]. Во-первых, этот практико-ориентированный подход претендует дать более глубокие ответы на хорошо известные, принципиально важные философские вопросы об эпистемологической специфике научной и математической методологии, чем традиционные «нисходящие» подходы. Яркими примерами этого преимущества являются не только упомянутые выше дебаты о природе научного объяснения, но и углубленное понимание того, что представляет собой практика доказательства в математике [7], а также подробные описания роли построения моделей в различных науках [8]. Во-вторых, практико-ориентированный подход обнаружил новые, ранее игнорировавшиеся нисходящим подходом исследовательские вопросы, которые объявлялись второстепенными при принятии теоретически ограничений, присущих традиционному подходу. Например, практико-ориентированный подход привел к исследованию и углубленному изучению различных источников доказательств, на которые опираются математики, таких как визуализация и доказательства с использованием

компьютера. Или, например, эпистемологических характеристики того, что составляет научное понимание, обнаружив, что оно не может быть полностью эксплицировано средствами номологически-дедуктивной модели Гемпеля [9]. Примером обнаружения средствами «восходящего» подхода проблемы, которая не могла быть усмотрена оптикой нисходящего подхода, является проблема эпистемических допущений при выборе формализма в математике [10]. Перечень примеров можно легко продолжить. Закономерно возникает вопрос о том, может ли принести подобные результаты восходящий подход в случае философии логики? Его проponent настаивает на том, что практико-ориентированный подход имеет явные преимущества перед традиционным «нисходящим» подходом, если мы заинтересованы в адекватном понимании целей, методологии и эпистемологии логики.

Первое, что обращает на себя внимание при обосновании практико-ориентированного подхода к философии логики – то, что многие из ограничений, выявленных в традиционных дискуссиях по общей философии науки и философии математики, находят отражение в пробелах философии логики. Слишком часто дебаты об этой области строятся на исторических предположениях о ее предмете, а не на исследованиях непосредственно в самой области исследования. В результате часто получаются описания логики, которые не только оторваны от реально развивающийся исследований в логике, но и доходят до дисквалификации реальных логических исследований, полагая их «дисциплинарно непригодными» [11].

В общем виде можно выделить следующие ограничения, налагаемые нисходящим подходом на философию логики. Использование нисходящего подхода приводит к созданию слишком идеализированных представлений о логике, искажающих реальное положение дел в этой области. Культивируемые в нем представления о природе логики часто оторваны от реальности логических исследований, основаны на допущениях относительно предмета исследования, методологии и эпистемологии логики, корнящихся в нормативных абстрактных классических платонистских, по своей сути, представлениях о метафизике и эпистемологии логики. В результате получается картина логики, противоречащая той, которую практикуют логики.

Ярким примером таких допущений и налагаемых ими ограничений может служить классическая проблема эпистемология логики: характеристика концепции валидности. Философская традиция утверждает, что доказательства валидности должны быть одновременно неинференциальными и априорными. Неинференциальными – потому что в противном случае пришлось бы предполагать валидность по крайней мере некоторых правил вывода, чтобы установить надежность выводов, которые частично составляют обоснование самих логических правил. Априорными – потому что, во-первых, никакие наблюдаемые состояния дел непосредственно не демонстрируют валидность правила вывода, и, во-вторых, возможность вывода доказательств для конкретных логических законов из эмпирических данных исключается неинференциальностью логических доказательств. Таким образом, если эти два исходных предположения верны, мы должны иметь непосредственный априорный доступ к истине логических законов о валидности.

Руководствуясь этими первоначальными эпистемологическими предположениями, современная эпистемология логики находится под влиянием двух фундаменталистских концепций логического знания – «логического рационализма» и «логического семантизма», – каждая из которых согласна с тем, что мы имеем непосредственный априорный доступ к логическим обоснованиям валидности, но которые расходятся во мнениях относительно источника этого априорного обоснования. К сожалению, обе концепции существенно оторваны от реальной практики обоснования логических теорий. Вместо того чтобы претендовать на непосредственный доступ к законам логики, ученые занимаются кропотливым процессом построения и проверки теорий, который основывается, на самых разных источниках обоснования, например, среди прочего, на неформальных доказательствах [см. 12].

Вторым ограничением нисходящего подхода является то, что получаемые в его рамках выводы упрощенно гомогенны, что приводит к необоснованным обобщениям и игнорированию важных тем философии логики. Из-за чрезмерной приверженности перечисленным выше допущениям нисходящего подхода утвердился акцент на примерах применения логики, считающихся «философски важными», парадигмальными, что привело к тому, что значительная часть проектов, в которых участвуют логики, и области

применения этой логики, игнорируются в современной литературе. Например, использование фреймов Крипке для моделирования ментальных установок в многоагентных системах «искусственного интеллекта» [13].

Ожидаемо, допущения нисходящего подхода часто носят антиисторический характер. Приверженность современной философии логики строить идеализированные описания, основанные на определенных допущениях о целях и методах логики, породила эссенциалистское описание логики, создающее впечатление диахронической однородности области логических исследований. Однако, согласно пропоненту восходящего подхода, как и другие области исследований, логика – это социальное предприятие с постоянно меняющимися целями, задачами и методами. Таким образом, современная философия логики склонна не только к неоправданным упрощениям при формулировании утверждений о методах и целях современных логических исследований (неоправданная синхроническая однородность), но и об историческом статусе логики (неоправданная диахроническая однородность).

Далее следует преимущественное внимание нисходящего подхода к конечному результату логических исследований. Современная философия логики концентрируется в основном на свойствах созданных логик, а не на процессах, которые привели к их разработке и обоснованию. Это имеет весьма существенные последствия не только потому, что может привести к пренебрежению многими важными методами и средствами, которые логики используют для разработки, обоснования и применения созданных систем, но и потому, что может привести к поспешным выводам и ограничениям на саму область логического исследования.

Предполагаемое практико-ориентированным подходом снятие ограничений нисходящего подхода в философии логики оборачивается заметным расширением области объектов исследования. Традиционная философия логики, как правило, концентрируется на устойчивых философских вопросах и их связи с логикой: логическая рациональность, метафизика логики, эпистемология логики. Согласно же стороннику восходящего подхода, цель философии логики состоит не в том, чтобы использовать логику в качестве нормативного примера для подтверждения выбранной философской теории. Согласно практико-ориентированному подходу, требуется исследовать

философские последствия всего многообразия деятельности логиков, выходящие за рамки того, что традиционно считалось философски важным. Как и в случае с другими методологическими различиями между практико-ориентированным подходом и традиционной философией логики, это различие в отношении объясняется тем, что согласно практико-ориентированному подходу, фактическая практика логического исследования имеет первостепенное значение, и именно эти практики должны определять наш философский анализ логики, а не традиционные философские допущения о ее природе. В конечном счете, задача философии логики состоит в том, чтобы понять логику такой, какая она есть, точно так же как цель философии науки состоит в том, чтобы понять науку «такой, какая она есть».

Очевидным возражением формирующемуся проекту восходящего исследования в области философии логики, является упрек в том, что он вместо того, чтобы исследовать природу «логики как таковой» предлагает исследовать машинерию деятельности по ее созданию.

Здесь нужно отметить, что дебаты о «логике как таковой» предполагают наличие некоего сущностного, *ala*-платонистского предмета логики, независимого от целей, преследуемых при разработке логических систем, как будто некая объективная природа определила эти объекты или предметы как «логические» по своему существу. Однако такой привилегированной внеисторической предметной логики не существует, так же как нет внеисторического предмета биологии, психологии или физики.

Пример других областей научной деятельности служит для того, чтобы подчеркнуть опасность попытки закрепить саму логику на основе заявлений основоположников. Это предложение (в некоторой степени неявно) использовалось в дебатах о первичной или канонической цели логики. Р. Кук [14], например, предположил, что мы должны рассматривать «кодификацию логических следствий на естественном языке» как первичную цель логики, основываясь на взглядах основоположников этой области, таких как Аристотель и А. Тарский. Но Аристотель также предполагал, что наука должна стремиться к предоставлению телеологических объяснений, и немногие ученые сейчас воспринимают эту деятельность всерьез. Методы меняются, цели меняются, хотя не все находится в состоянии изменчивости. Важным вопросом

для философии науки (любых конкретных дисциплин) является вопрос о том, какие именно исторические связи необходимо учитывать для того, чтобы говорить о преемственности конкретной области исследований, будь то физика, математика или логика?

Несмотря на универсальность логики, игнорирование на этом основании эволюции ее целей, средств и того, какое влияние на ее развитие оказывают ее применения в других областях научного знания, очевидно приводит к упрощенному представлению о ее развитии. Заявление о себе нового, восходящего проекта в философии логики обещает расширение поля философии логики за счет включения в него целого спектра проблем, ранее дисквалифицированных как не имеющих существенного значения для решения канонических вопросов о природе логического знания. Такое расширение кажется многообещающим и, прежде всего, в связи с проблемой применения логики в других областях науки, предполагая в этой проблеме обнаружить новые факты, способствующие продвижению в решении такой важной проблемы как соотношение эмпирического и формально-логического знания. Дебаты о методологии философии логики обещают нам появления в этой области проекта, напоминающего проект Я. Хакинга, который он предпочитал называть динамическим номинализмом в философии науки. Хотя судить о перспективах такого подхода в отношении заявляемых целей, преждевременно, надежды на получение результатов, сродни тем, которые были получены динамическим номинализмом по реконструкции сложности и многообразию научной практики, вполне представляются обоснованными.

### Литература

1. *Woody A.I.* Re-orienting Discussions of Scientific Explanation: A functional perspective. // *Studies in History and Philosophy of Science*. 2015. № 52. P. 79–87.
2. *Brigandt I.* Explanation in Biology: Reduction, pluralism, and explanatory aims. // *Science & Education*. 2013. № 22. P. 69–91.
3. *Qiu R-Z.* Models of Explanation and Explanation in Medicine. // *International Studies in the Philosophy of Science*. 1989. № 3. P. 199–212.
4. *Fisher G.* Explaining Explanation in Chemistry. // *Annals of the New York Academy of Sciences*. 2003. № 988. P. 16–21.

5. *Carter J.* Philosophy of Mathematical Practice: Motivations, themes and prospects. *Philosophia Mathematica*. 2019. № 27. P. 1–32.
6. *Hamami Y., Morris R.L.* Philosophy of Mathematical Practice: A primer for mathematics educators. // *ZDM Mathematics Education*. 2020. № 52. P. 1113–1126.
7. *Soler L.* Introduction. // *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science*. 2014. P.1–43.
8. *De Toffoli S.* (2020) Reconciling Rigor and Intuition. // *Erkenntnis*. 2020. № 86. P. 1783–1802.
9. *Gelfert A.* How to Do Science with Models: A philosophical primer. Dordrecht: Springer. 2016.
10. *de Regt H.W.* Understanding Scientific Understanding. Oxford: Oxford University Press. 2017.
11. *De Toffoli S.* ‘Chasing’ the Diagram—The Use of Visualizations in Algebraic Reasoning. // *The Review of Symbolic Logic*. 2017. № 10. P. 158–86.
12. *Martin B.* Identifying Logic Evidence. // *Synthese*. 2021. № 40. P. 9069–9095.
13. *Priest G.* Logical Disputes and the a Priori. // *Logique et Analyse*. 2016. № 59. P. 347–66.
14. *Wooldridge M.* An Introduction to MultiAgent Systems. London: John Wiley & Sons. 2009.
15. *Cook R.T.* Let a Thousand Flowers Bloom: A tour of logical pluralism. // *Philosophy Compass*. 2010. № 5. P. 492–504.

## References

1. *Woody A.I.* Re-orienting Discussions of Scientific Explanation: A functional perspective. // *Studies in History and Philosophy of Science*. – 2015. – № 52. – P. 79–87.
2. *Brigandt I.* Explanation in Biology: Reduction, pluralism, and explanatory aims. // *Science & Education*. – 2013. – № 22. – P. 69–91.
3. *Qiu R-Z.* Models of Explanation and Explanation in Medicine. // *International Studies in the Philosophy of Science*. – 1989. – № 3. – P. 199–212.
4. *Fisher G.* Explaining Explanation in Chemistry. // *Annals of the New York Academy of Sciences*. – 2003. – № 988. – P. 16–21.
5. *Carter J.* Philosophy of Mathematical Practice: Motivations, themes and prospects. *Philosophia Mathematica*. – 2019. – № 27. – P. 1–32.
6. *Hamami Y., Morris R.L.* Philosophy of Mathematical Practice: A primer for mathematics educators. // *ZDM Mathematics Education*. – 2020. – № 52. – P. 1113–1126.

7. *Soler L.* Introduction. // *Science After the Practice Turn in the Philosophy, History, and Social Studies of Science.* – 2014. – P.1–43.
8. *De Toffoli S.* (2020) Reconciling Rigor and Intuition. // *Erkenntnis.* – 2020. – № 86. – P. 1783–1802.
9. *Gelfert A.* How to Do Science with Models: A philosophical primer. Dordrecht: Springer. 2016.
10. *de Regt H.W.* Understanding Scientific Understanding. Oxford: Oxford University Press. 2017.
11. *De Toffoli S.* ‘Chasing’ the Diagram–The Use of Visualizations in Algebraic Reasoning. // *The Review of Symbolic Logic.* – 2017. – № 10. – P. 158–86.
12. *Martin B.* Identifying Logic Evidence. // *Synthese.* – 2021. – № 40. – P. 9069–9095.
13. *Priest G.* Logical Disputes and the a Priori. // *Logique et Analyse.* – 2016. – № 59. – P. 347–66.
14. *Wooldridge M.* An Introduction to MultiAgent Systems. London: John Wiley & Sons. 2009.
15. *Cook R.T.* Let a Thousand Flowers Bloom: A tour of logical pluralism. // *Philosophy Compass.* – 2010. – № 5. – P. 492–504.

### Информация об авторе

*Хлебалин Александр Валерьевич* – Институт философии и права Сибирского отделения Российской академии наук (ул. Николаева, 8, Новосибирск, 630090, Россия).  
sasha\_khl@mail.ru

### Information about the author

*Khlebalin Alexander Valerievich*– Institute of Philosophy and Law of the Siberian Branch of the Russian Academy of Sciences (8, Nikolaeva st., Novosibirsk, 630090, Russia).  
sasha\_khl@mail.ru

Дата поступления: 18.07.2025  
Принята в печать: 02.10.2025



*Философские проблемы логики  
и методологии науки*

УДК: 51: 101.8

DOI: 10.15372/ PS20250511

EDN: BCQCTN

**В.М. Резников**

**МЕШАЮЩИЕ И СПОСОБСТВУЮЩИЕ ФАКТОРЫ  
ФИЛОСОФСКОЙ АКТИВНОСТИ В НАУКЕ,  
И НЕКОТОРЫЕ ПЕРСПЕКТИВЫ  
УЧАСТИЯ ФИЛОСОФОВ В НАУКЕ**

В статье я защищаю тезис о недооценивании научным сообществом значимости философских идей и эффективного участия философов, обладающих знаниями математики, информатики и естествознания в научных исследованиях. В первой части статьи я показал, что мешающим фактором для развития науки и философии науки является отсутствие консенсуса философского сообщества о философском статусе методологии науки, и как следствие недостаточное внимание исследованиям по этой дисциплине. Я рассмотрел две известные группы аргументов против методологии науки. Первая группа основана на признании философским сообществом релятивистского характера физического знания после открытий теории относительности и квантовой физики. Вторая группа аргументов основана на отношении к науке классиков философии науки. Так, П. Фейерабенд отрицал универсальный характер научных методов, а Т. Кун зачастую отрицал объективное развитие науки. Я сформулировал следующие контраргументы. Во-первых, несмотря на определенный релятивизм науки, абстрагирование философов от научных достижений приводит к снижению качества

исследований в философии науки. Во-вторых, в своей автобиографии Фейерабенд признал иронический характер его знаменитого аргумента, обосновывавшего научный релятивизм.

Во второй части статьи я описал определенные перспективы участия философов, обладающих знаниями специальных наук в научной деятельности в основном в качестве критиков несовершенств науки, некорректного использования учеными философских идей. Однако имеются определенные перспективы для участия философского сообщества в проектах некритического характера, например, конструирование новых критериев объяснения и понимания для естественнонаучных и формальных дисциплин.

*Ключевые слова:* Философия науки; методология науки; физика; история квантовой физики; теория вероятностей; применение теории вероятностей; философия информатики.

**V.M. Reznikov**

## **INTEFERING AND CONTRIBUTING FACTORS OF PHILOSOPHICAL ACTIVITY IN SCIENCE AND SOME PROSPECTS FOR PARTICIPATION OF PHLOSOPHERS IN SCIENCE**

In the first part of the article, I showed that an interfering factor for the development of science and philosophy of science is the lack of consensus in the philosophical community about the philosophical status of the methodology of science, and, as a consequence, inadequate attention to research in this discipline.

I have considered two well-known sets of arguments against the methodology of science. The first group is based on the recognition by the philosophical community of the relativistic nature of physical knowledge after the discoveries of the theory of relativity and quantum physics. The second group of arguments is based on the attitude of the classics of the philosophy of science towards science. Thus, P. Feyerabend denied the universal nature of scientific methods, and T. Kuhn often denied the objective development of science. I have formulated the following counterarguments. First, despite a certain relativism in contemporary science, philosophers' abstraction from scientific achievements leads to a decline in the quality of research in the philosophy of science. Secondly, in his autobiography, Feyerabend acknowledged the ironic nature of his famous argument for scientific relativism.

In the second part of the article, I described certain prospects for the participation of philosophers with knowledge of specialized sciences in scientific activity, mainly as critics of the imperfections of science and the incorrect use of philosophical ideas by scientists. However, there are certain prospects for the participation of the philosophical community in projects of a non-critical nature, for example, the construction of new criteria of explanation and understanding for the natural sciences and formal disciplines.

*Keywords:* Philosophy of science; methodology of science; физика; history of quantum physics; probability theory; application of probability theory; philosophy of informatics.

Настоящее исследование посвящено актуальной и гипотетической деятельности философского сообщества в науке. Однако необходимо заметить, что философская активность не обязательно является прерогативой исключительно представителей философского сообщества. Так как центральной составляющей этой деятельности является критический подход по отношению к основаниям различных областей знания, включая и философию, поэтому философская активность проявляется не только в работах философов, но и в работах критически мыслящих исследователей, которые являются представителями самых различных областей знания. Здесь уместно привести одно высказывание, приписываемое знаменитому математику Д. Гильберту о физике. Физика слишком значимая наука, для того чтобы ее оставить физикам. Перефразируя Гильберта можно сказать, философия слишком значимая наука, чтобы ее оставлять философам. Так как, вопросы, исследуемые в данной работе, практически не представлены в современной литературе, поэтому предлагаемая статья является актуальной.

### **Мешающие факторы для участия философов в науке**

1) Научные дисциплины не имеют обязательного статуса для их изучения на философских факультетах. Известно, что серьезный вклад философов в научной сфере или в области философии науки предполагает солидную подготовку представителей философии в научных дисциплинах, однако в настоящее время математика, физика, информатика, методология науки и другие науки, необ-

ходимые для научной подготовки ученого, не являются обязательными для изучения дисциплинами на философских факультетах университетов. Очевидно, что единственная обязательная общеобразовательная дисциплина под названием: “концепции естествознания”, ни в коей мере не может заменить систематического изучения физики, химии или другой естественнонаучной дисциплины.

2) Методология науки находится на периферии интересов философов. Начиная с античной философии, исследования оснований философских, логических и естественнонаучных методов являлись составной частью философии. Однако серьезный интерес к методологии науки и выделение ее в самостоятельную область философии связано с работами Ф. Бэкона, Г. Галилея, Р. Декарта и И. Ньютона. Можно сказать, что с начала современной философии и вплоть до 20-30 гг. прошлого столетия методология науки являлась общепризнанной частью философии. Однако ситуация меняется после знакомства философов с выдающимися открытиями в области физики: частной теорией относительности, и особенно, с общей теорией относительности и квантовой физикой. Начиная с этого момента времени и до сих пор в философском сообществе нет консенсуса по вопросу о принадлежности методологии науки к философским наукам. Так, некоторые философы отрицают принадлежность методологии науки к философии. Известны две группы аргументов, отрицающих значение методологии науки для философии.

Первая группа аргументов опирается на релятивистский характер научного знания, в частности это касается математики, и особенно физики. Известно, что И. Кант считал вершиной познания эвклидову геометрию, однако Н. Н. Лобачевский построил независимую от геометрии Эвклида собственную геометрию, и геометрия Эвклида стала одной из возможных геометрий. Еще сложнее ситуация с физикой, так механика И. Ньютона начиная с ее создания имела статус солидно обоснованной, практически совершенной науки. Однако, после открытия А. Эйнштейном теории относительности, механика Ньютона оценивается физиками, как теория, которая является частным случаем теории Эйнштейна. Надо отметить, что создание квантовой физики оказало сильное влияние на изменение мировоззрения ученых, особенно физиков, так как законы физики микромира имеют неде-

терминистский, а вероятностный характер. Поэтому в связи с открытиями физики, некоторые философы пришли к мысли, что нет смысла изучать методы физики, так как при изменении теории старые методы оказываются на периферии интересов физики. Еще один аргумент против методологии науки основан на определенной аналогии деятельности философа с работой учителя в начальных классах. И учитель, и преподаватель философии помогают ученикам в овладении базовыми, устоявшимися идеями и интеллектуальными навыками. Другими словами, философы занимаются универсальными, вечными проблемами, а методология науки представляет интерес для философов, специализирующихся в области науки и для представителей науки.

Вторая группа аргументов опирается на отношение выдающихся философов, прежде всего П. Фейерабенда, К. Поппера и Т. Куна к методам науки. Начнем с Фейерабенда, так, на вопрос научных журналистов об универсальных методах науки, Фейерабенд отвечал фразой, ставшей знаменитой: «Все дозволено» [4, С. 197]. На первый взгляд, ответ Фейерабенда означал, что в науке нет универсальных методов познания, и в разных ситуациях могут оказаться подходящими разные методы. Однако необходимо ответить, что в своей интеллектуальной автобиографии Фейерабенд признает иронический характер его знаменитой фразы в ответ на докучливые вопросы журналистов [11]. Тем не менее, фактический отказ Фейерабенда от релятивизма научных методов не принимается во внимание в подавляющем множестве публикаций, и эта фраза используется как неосновательный аргумент против значимости методологии науки. От Фейерабенда к Попперу, так, К. Поппер считал, что единственным основательным методом науки является фальсификация знания. Противники методологии науки, используя его отношение к методам науки, выдвигали следующий аргумент против методологии науки. Какой смысл в исследованиях по методологии науки, если содержание этой дисциплины исчерпывается единственным методом фальсификации? Теория смены парадигм Куна тоже льёт воду на мельницу отрицателей методологии науки, так как в этой теории основное внимание уделяется не исследованию объективных методов, а анализу социальных факторов, ответственных за изменение направлений научных исследований. Еще один контраргу-

мент к методологии науки заключается в том, что философия до сих пор не справилась с решением проблемы индукции, Б. Рассел называл эту ситуацию скандалом в философии.

По моему мнению, рассуждения, представленные против значимости методологии науки, не являются особенно убедительными, поэтому я ограничусь краткими возражениями. Во-первых, аргумент, основанный на том, что философия занимается глобальными проблемами, является корректным, однако невозможно абстрагироваться от последних достижений науки, например, Интернета, искусственного интеллекта, так как они всё более глубоко и всестороннее пронизывают жизнедеятельность на всех уровнях от быта до профессиональной деятельности, охватывая и вечные философские вопросы. Во-вторых, действительно, в любой интеллектуальной деятельности, включающей науку, нет непогрешимых, универсальных законов, однако это не является рациональным основанием для отказа от методологии науки. В-третьих, Фейерабенд и Кун отказались от наиболее жестких аргументов против рациональности науки. В-четвертых, необходимо отметить различие в отношении к проблеме индукции в науке и философии. Так, в науке эта проблема не представляется особо острой и актуальной, более того, ученые считают, что эта проблема решена в некоторых научных дисциплинах, в частности в области распознавания образов [15]. Напротив, в философии пытались решить эту проблему в полном объеме и глубине, без использования специальных методов, поэтому практически не было шансов на успех.

После ответов на возражения против методологии науки, я приведу несколько собственных рассуждений о значимости методологии науки. Так, я связываю будущее философии науки не только с углублением исследований общих проблем философии науки, таких, как верификация и фальсификация знания, связь теории и практики научных исследований и других традиционных проблем философии, но и с учетом достижений и актуальных проблем науки. Необходимо напомнить, что все научные достижения основаны на применении научных методов, поэтому вполне очевидно, что методология науки является естественным мостом, связывающим науку и философию, отсюда следует, что отсутствие должного внимания к методологии науки является сдерживающим фактором для развития, как науки, так, и философии науки.

Я написал об отрицательных последствиях должного внимания философского сообщества к методологии науки. Теперь кратко опишу о позитивных последствиях для исследований в области методологии науки для философов. Во-первых, философ, который овладел методологией науки, и, в частности, научными методами, начинает понимать науку в целом. Во-вторых, при углубленном изучении методологии науки и научных дисциплин, исследователь узнает о запросах науки к философскому сообществу, и о научных проблемах, которые представляют интерес и для философии. В-третьих, специалисты в философии науки, владеющие научными методами, получают возможность работы в науке, например, выполняя, часто вакантную функцию критики несовершенств научных методов. Отмечу, что участие философов в научной деятельности способствует повышению социального статуса философии, значимости философской деятельности и общественной престижности профессии философа. В-четвертых, знание научных методов открывает возможность их использования для анализа философских проблем, а знание научных проблем, которые представляют интерес для философии, открывает возможности для их исследования на основе философского инструментария.

Теперь приведу некоторые соображения по поводу характера работы философов в науке. В наибольшей степени философская деятельность в науке имеет критический характер. Почему деятельность философов в научной сфере в целом имеет критическую направленность. Во-первых, обучение философии на философских факультетах является основополагающим фактором для овладения выпускниками этих факультетов критической методологии. Дело в том, что многие работы выдающихся философов, как правило, представляют собой критику предыдущих философских теорий, и создание новых теорий, которые являются развитием собственных идей, минимально связанных с работами предшественников. Если выпускники философских факультетов продолжают совершенствоваться в философской деятельности, то они со временем становятся отменными критиками. Во-вторых, в науке постоянно наблюдается дефицит в работах, представляющих собой корректную, содержательную критику несовершенств научных методов и научной деятельности в целом. Дело в том, что в науке не приня-

то критиковать несовершенные подходы и методы, не предлагая собственные альтернативные методы, или хотя бы конструктивные идеи об их построении. Как известно, в науке ценятся исключительно положительные результаты за исключением сугубо теоретических областей знания, поэтому в науке не часты публикации критического характера. В отличие от исследователей в области науки, у исследователей в сфере философии, как правило, нет комплекса положительных результатов, так как в философском сообществе ценятся не только новые идеи, но и представляют самостоятельный интерес глубокие работы критического характера.

### **Некоторые способствующие факторы для участия философов в науке**

1) Глубокое изучение ряда наук, в том числе, математики, физики, других естественнонаучных дисциплин, информатики, философии и методологии науки.

2) Особое значение имеют творческие контакты с философствующими учеными и изучение публикаций философствующих ученых.

Одна из целей настоящей работы состоит в том, чтобы продемонстрировать на примерах публикаций: философствующих физиков, математиков в области теории вероятностей, прикладной статистики и информатики истоки возникновения философских проектов в области науки. Начну с нобелевского лауреата в области физики С. Вайнберга, он неплохо разбирался в философских проблемах, имеющих отношение к науке. О его компетенции в области философии свидетельствует содержание и стиль его научно-популярной книги, ставшей знаменитой: «Мечты об окончательной теории», а также признательность Вайнберга в разнообразной поддержке при написании этой книги в предисловии к ней многим ученым, в том числе 12 философам [2]. Одна глава этой книги носит название: «Против философии», в ней критикуется значимость философии для науки. Сначала я кратко опишу претензии автора к философии, а потом отвечу на критику Вайнберга. Во-первых, он не обнаружил в философии идей, способствующих развитию физики, в лучшем случае философские

идеи помогают не повторить ошибки прошлого. Во-вторых, в качестве полезности философии, как редкого исключения из правил, Вайнберг приводит философию позитивизма, исключительно ориентированную на науку. Вайнберг вспоминает, что приверженность позитивизму помогла А. Эйнштейну сформулировать частную теорию относительности, однако Вайнберг отмечает, что следование этой философии не всегда оказывалась положительным фактором, как для достижения собственных результатов, так и для принятия новых результатов, полученных другими учеными. Так, известный физик Д. Томпсон является автором открытия электрона на основе ряда экспериментов, однако примерно в то же время физик Кауфман в серии испытаний, проведенных на качественно более высоком уровне, получил свидетельства о новой частице, получившей позже название электрона. Кауфман не стал анализировать, полученные им косвенные свидетельства новой частицы и поэтому не стал автором открытия, дело в том, что он был последовательным позитивистом, а в этой философии значимыми считаются открытия на основе прямых экспериментов, а не достижения путем домысливания косвенных свидетельств. Известный физик и философ Э. Мах, один из создателей логического позитивизма до конца жизни не принимал открытия атомов, так как они не были обнаружены на основе прямых экспериментов. В-третьих, философское восприятие оказывается консервативным, и минимально восприимчивым к новым достижениям. Как известно, все разделы классической физики редуцируемы к механике, причем сведение к ней свидетельствовало о понимании решаемой проблемы, и свой высокий статус механика сохраняла вплоть до открытия частной теории относительности. Вайнберг отмечал, что сохранение особого статуса механики столь длительное время стало возможным в результате проникновения идей позитивизма в культуру, в том числе в марксистскую философию и присущего философии консерватизма. В-четвертых, в философском арсенале нет наработок по формулированию объяснений, адекватных для физических наук. Вайнберг пишет: «Согласно Фейерабенду, понятие научного объяснения, разработанное рядом философов науки, столь узко, что невозможно говорить, что какая-то теория объясняется другой теорией. Эта точка зрения оставляет мое поколение физиков, занимающихся части-

цами, без работы» [4, с.98]. Последнее предложение в цитируемом нами абзаце из книги Вайнберга подчеркивает значимость проблемы объяснений для развития физики.

Теперь я приведу возражения в ответ на критику философии Вайнбергом. Во-первых, в работе современных методологов Р. Нола и Г. Санки приведен ответ известного философа В. Салмона на критику Вайнберга [11, с.4]. Так, по поводу первого критического замечания Вайнберга, Салмон отмечал, что философы отнюдь не стремятся заменить физиков, однако некоторые разработки философов, например, связанные с вопросами объяснения и понимания, могут оказаться полезными для физиков. Во-вторых, Вайнберг прав, что не существует универсальной философской доктрины, которая является наилучшей для применения во всех науках для решения всех проблем. Если бы такая универсальная концепция существовала, то рано или поздно ученые бы ее обнаружили, и применяли для решения всех научных проблем. Однако наилучших универсальных философских теорий не существует. Так, для А. Эйнштейна философия представлялась резервуаром плодотворных идей, однако физик имеет ограниченные интеллектуальные и временные ресурсы для исследования интересующей его проблемы, поэтому не исключено, что ученый использует не оптимальную философскую теорию, а ту, которой он владеет [12]. Однако в последствие серьезный физик углубляет и расширяет границы своего философского аппарата, и поэтому не исключено, что при решении другой физической проблемы, физик будет применять иную философскую теорию. А. Эйнштейн отмечал, что при формулировании частной теории относительности он опирался на философские идеи эмпириков: Э. Маха и в еще большей степени на Д. Юма, однако, в дальнейшем при создании общей теории относительности для него оказалась плодотворной философия конвенционализма [6]. В-третьих, утверждение о том, что философы являются консерваторами, не является универсально истинным, так как, по крайней мере, по отношению к философии оно оказывается ложным. Теперь возвращаемся к физике, приведу несколько фактов, опровергающих философский консерватизм по отношению к физике. Так, далеко не все философы после создания теории относительности и квантовой физики искали причины для того, чтобы не изучать современные разделы физики, и тем самым не отрицали значимость методологии науки и необхо-

димось изучения методов, используемых в физических науках. Например, известный философ Г. Рейхенбах вскоре после создания теории относительности и квантовой физики, написал содержательные монографии, посвященные этим открытиям [13]. Приведу еще один факт, убедительно опровергающий консерватизм физиков. Известно, что после открытия нескольких различных интерпретаций квантовой механики, многие физики оказались неготовыми воспринимать дальнейшие результаты, относящиеся к усовершенствованию полученных интерпретаций, и тем более, последующие открытия других интерпретаций. Так, по воспоминаниям известного физика Д. Белла, ведущие физические журналы не принимали к печати статьи, содержащие описания новых интерпретаций квантовой физики. Физикам неограниченную поддержку оказало философское сообщество, дело в том, что у философов нет предубеждений против множественности несовместимых объяснений одних и тех же феноменов. Так, физики долгое время публиковались в философских журналах, вплоть до признания физическим сообществом легитимности множественности интерпретаций квантовой физики [8]. В-четвертых, что касается проблем объяснения и понимания, то, действительно, в философии пока немного результатов, адекватных для наук, однако интенсивные исследования в этом направлении продолжаются.

Не только физики критикуют значимость философии для науки, так некоторые ученые, в частности, специалист в области теории информации и искусственного интеллекта Д. Барфут предъявляет претензии к философии, в частности, к попперовскому фаллибилизму, так как, по его мнению, он не определяет вектор развития науки [11]. Он полагает, что философия имеет значение, однако в современной философии нет значимых идей для развития науки, поэтому Барфут создал собственную теорию развития знания на основе достижений в тех областях знания, которыми он занимался. Так, для теории информации значимым является понятие колмогоровской сложности, напомним определение этого понятия. Пусть задана последовательность чисел, тогда колмогоровская сложность определяется длиной программы, которая печатает эту последовательность. Очевидно, что чем более явными и понятными закономерностями обладает эта последовательность, тем меньшей будет длина программы, и, наоборот, длина программы для хаотической, бессистемной последователь-

ности будет сопоставима с длиной этой последовательности. Вы скажу свои замечания к критерию развития науки на основе колмогоровской сложности. Конечно, теория, определяющая больше закономерностей, по сравнению с другими теориями, является наилучшей. Однако автор не предлагает, каких либо идей и методов по выявлению таких теорий. Кроме того, изложение таких теорий, отнюдь, не должно следовать критерию минимальной колмогоровской сложности, так как такой стиль изложения, по видимому, затрудняет овладение этой теорией. И, последнее, критика Барфутом фаллибилизма Поппера не является адекватной, так как фаллибилизм имеет самостоятельное значение и вполне совместим с критерием колмогоровской сложности.

Изучение научных дисциплин открывает потенциальные возможности для работы философа, так как научное знание не свободно от недостатков. Я обратился к теории вероятностей для описания ее некоторых несовершенств в контексте эпистемологии и корректной применимости этой науки. Для чего нужен критический анализ теории вероятностей в контексте ее приложений вне математики? Дело в том, что любая математическая наука является абстрактной дисциплиной, и заточена для исследования абстрактных сущностей, а не реальных объектов. Кроме того, в настоящее время нет ни одной известной теории в философии математики, предназначенной для определения, того, какими свойствами должны обладать формальные теории, адекватные для приложений вне математики? В тоже время известно несколько теорий в философии математики, предназначенных для исследования оснований математики, например, формализм, интуиционизм, платонизм и другие. В качестве базовой книги по теории вероятностей я выбрал известную книгу: “Основания теории вероятностей” А.Н. Колмогорова, так как знаменитый математик был философствующим математиком [3].

### **Критический анализ оснований теории вероятностей в контексте ее приложений**

1) Нужна ли вообще аксиоматика теории вероятностей для ее корректных и эффективных приложений? Здесь нет полного консенсуса, так С.Н. Бернштейн отмечал, что только обоснованная,

аксиоматизированная теория адекватна для приложений [1], а В.И. Арнольд отрицал значимость аксиоматического подхода для приложений математики. Я полагаю, что в целом аксиоматика не имеет значения для применения теории вероятностей, однако можно описать несколько ситуаций, для которых наличие аксиоматизации имеет некоторый смысл. Так, для применения известного знания аксиоматика не важна, но она значима для получения нового знания, и, если после этого оно будет применено и окажется адекватным для приложений, то это будет свидетельствовать о значимости аксиоматики для приложений.

2) Критика абстрактного характера аксиоматики А.Н. Колмогорова в контексте приложений. Самые разные исследователи отмечали абстрактный характер аксиоматики Колмогорова, среди них сам А.Н. Колмогоров [3], философ Хэмфри [8]. Фон Нейман - создатель первой формальной аксиоматики для квантовой физики утверждал, что для физики адекватна более содержательная аксиоматика.

3) Критика теории меры, математической науки, лежащей в основаниях теории вероятностей. Базовым понятием теории меры является понятие меры, оно представляет собой обобщение таких понятий, как: длина, площадь, объём и так далее. Понятие меры никак не учитывает специфику событий, изучаемых в стохастической математике. Кроме того, эта наука является более абстрактной, чем теория вероятностей. Некоторые прикладные математики отрицали ее значимость для теории вероятностей, например, Р. Мизес [9].

4) Критика выбора Колмогоровым понятия безусловной вероятности событий в качестве первичного понятия своей аксиоматики. В аксиоматике Колмогорова понятие вероятности  $P(A)$  для произвольного события  $A$  является первичным понятием, а понятие условной вероятности  $P(A/B)$  события  $A$ , при условии, что  $B$  произошло, оказывается производным понятием.  $P(A/B)$  определяется на основе вероятности  $P(A \cap B)$  совместного события  $A$  и  $B$  и вероятности  $P(B)$  события  $B$ :

$$P(A/B) = P(A \cap B) / P(B) \quad (1)$$

С эпистемологической точки зрения выбор в качестве первичных понятий безусловных или полных вероятностей событий не является корректным, так как, исследователи, как правило, обла-

дают знанием вероятностей событий при некоторых условиях, т.е. условными вероятностями. В настоящее время известно несколько вариантов построения аксиоматических теорий вероятностей, в которых первичным понятием является условная вероятность события, в качестве примера можно отметить теорию А. Реньи [14].

5) Об ограниченности понятия условной вероятности. Некоторые авторы критикуют ограниченность понятия условной вероятности, критика заключается в том, что вероятность условия не может равняться нулю. В ряде публикаций предложены решения по снятию этих ограничений. А.Н. Колмогоров отмечал особую значимость понятия независимости для создания и развития теории вероятностей, а также для ее отделения от теории меры. Это замечание Колмогорова является очень важным, имеет смысл его проанализировать. Первая часть утверждения Колмогорова является абсолютно справедливой, я уделю внимание последней его части о значимости независимости для отделения теории вероятностей от теории меры. Я полагаю, что А.Н. Колмогоров частично прав, так, с одной стороны, в теории меры не уделяется особого внимания применению понятия независимости, и, наоборот, в теории вероятностей это понятие используется, однако с другой стороны, выполнимость отношения независимости в теории вероятностей является следствием выполнимости этого отношения в теории меры.

6) Критика формального введения отношения независимости. Ряд известных математиков критиковали формальное введение отношения независимости в колмогоровской аксиоматической теории вероятностей. Так, Э. Борель, один из создателей теории меры, отмечал, что с одной стороны, теория вероятностей является не вполне математической наукой, так как она занимается определением вероятностей для исходов экспериментов по бросанию монет, кубиков. Однако, с другой стороны, понятие независимости введено формально и имеет абстрактный характер. По сути, Э. Борель не принимал определения независимости, так как не понимал, как оно соотносится с классическими объектами математики. Со временем Борель открыл, что отношение независимости выполняется для дробных частей чисел, и тем самым Борель обнаружил определенную связь классической математики и теории вероятностей. Другой известный математик Г. Штейнгауз тоже не принимал формального введения понятия независимости в теории вероятностей, вне его

связи с классической математикой. Он предложил своему ученику, математику В. Кацу, подумать для каких математических объектов выполняется это определение. Кац успешно справился с предложенной задачей, он обнаружил, что определение независимости выполняется для тригонометрических функций [9]. Насколько естественными являются результаты, полученные Борелем и Кацем? Во-первых, у Бореля оказываются независимыми числа, которые составляют дробную часть одного и того же числа. Этот результат оказывается не вполне интуитивным, так как не вполне естественно считать выполнимость независимости для чисел, имеющих генетическую связь. Во-вторых, согласно результатам Бореля и Каца, понятие независимости оказывается применимым для объектов, которые, по крайней мере, на первый взгляд, не имеют связи со стохастикой.

А теперь о преимуществе выбора философами книги А.Н. Колмогорова. Так, в своей книге Колмогоров отмечает, что к важнейшим задачам философии естествознания относятся выяснение и уточнение тех предпосылок, при которых можно изучаемые явления природы, рассматривать как независимые [3]. Замечание о независимости А.Н. Колмогорова оказывается абсолютно справедливым даже по отношению к настоящему времени, так как значимость и роль этого понятия не освоены в должной степени в естествознании, формальных науках и в философских дисциплинах. Кроме того, до сих пор не полностью исследованы связи этого понятия с другими понятиями и философскими категориями, а также сопричастность идеи независимости значимым событиям. Так, независимость естественным образом связана с причинностью, пусть  $A$  причина  $B$ , тогда  $A$  не зависит от  $B$ , это обстоятельство открывает возможность изучения независимости на основе причинных отношений. Хотя идея независимости не получила должного внимания, ее присутствие обнаруживается уже в античной метафизике. Так, если базовыми элементами античного мироздания являлись земля, вода, воздух, огонь, то каждый из этих элементов не зависит от других, аналогично все аксиомы геометрии Эвклида не зависят друг от друга. Еще идея независимости сопричастна ко многим знаменательным событиям в математике. Так, открытие независимости пятого постулата Эвклида от аксиом, привело к открытию неэвклидовой геометрии, и осознанию, что аксиомы не обязательно являются описа-

нием свойств реального мира, достаточно, что они формально корректны, и могут описывать свойства воображаемых миров. Отмечу, что в настоящее время идея независимости изучается, например, в рамках исследования понятия общей причины [7].

### **Заключение**

В работе я показал, что ограниченное внимание и интерес к методологии науки являются серьезными негативными факторами для развития науки, философии науки и активности философского сообщества в научной деятельности. Позитивными факторами для участия философов в науке являются контакты с философствующими учеными и изучение их работ. Так, на примерах нескольких теорий: теории всего С. Вайнберга, аксиоматизированной теории вероятностей А.Н. Колмогорова и теории вектора развития науки Д. Барфута, авторы которых являлись философствующими учеными, я продемонстрировал, что изучение их работ приводит к новым постановкам философских проблем и возможным проектам по их решению. Во-первых, критический анализ известных подходов к описанию независимости в естественнонаучных и формальных науках, а также создание философской теории независимости, адекватной этим дисциплинам. Во-вторых, изучение критериев объяснения и понимания, используемых учеными в естествознании и формальных науках, а также конструирование новых критериев, адекватных для этих наук. И, наконец, на основе истории квантовой физики я продемонстрировал, что приписывание С. Вайнбергом консерватизма, присущего философскому сообществу, по отношению к физике не является полностью обоснованным.

### **Литература**

1. *Бернштейн С.Н.* Современное состояние теории вероятностей и ее приложений // Бернштейн. Собрание сочинений в четырех томах: В Т. IV. – М.: Наука 1964. – С. 217-232.
2. *Вайнберг С.* Мечты об окончательной теории. – М.: Едиториал УРСС, 2004.

3. Колмогоров А.Н. Основные понятия теории вероятностей. – М.: ОНТИ НКТП СССР, 1936.
4. Фейерабэнд П. Против метода. – М.: Хранитель, 2007.
5. Burfoot D. Notes on a new philosophy of empirical science, 2011. URL: <https://arxiv.org/abs/1104.5466> (дата обращения: 29.11.2025).
6. Feierabend P. Killing Time. – Chicago and London.: The University Chicago Press, 1995.
7. Hofer-Szabo G., Redei M., Szabo L. The principle of the common cause. – Cambridge: Cambridge University Press, 2013.
8. Humphrey P. Probability theory and its models. // Probability and Statistics: Essays in Honor of David A. Freedman / Eds. by D. Nolan, T. Speed. Institute of mathematical statistics Collections. Vol. 2. – P. 1–11, 2008.
9. Кас М. Enigmas of chance. – California: California University Press, 1987. URL: <https://scilib-physics.narod.ru/Кас/index.html> (дата обращения 29.11. 2025).
10. Mises P. Mathematical theory of probability and statistics. – New York and London: Academic Press, 1964.
11. Nola R., Sankey H. Theories of scientific method. – Acumen, 2007.
12. Norton J. How Hume and Mach helped Einstein find special relativity // Discourse on a new method. Edited by Domsy M. and Dicrson, M. – Chicago: Open Court, 2010.
13. Reichenbach H. Philosophical foundations of quantum mechanics . – California: California University Press, 1944.
14. Renyi A. On a new axiomatic theory of probability // Acta mathematica Academiae Scientiarum Hungarica – 1955. –Vol. 6(3). – P. 285-335.
15. Vapnik V. Estimation of dependencies based on empirical data. –. New-York: Springer, 2006.

## References

1. Bernstein, S.N. (1964). Sovremennoe sostoyanie teorii veroyatnosti i ee prilozheniya. [The current state of probability theory and its applications]. In Bernstein. Sochineniya: V IV t. [Works in IV vol.], Vol. IV. Moscow, Nauka Publ.
2. Weinberg, S. (2004). Mechty ob okonchatelnoi teorii. [Dreams of a final theory]. Мечты об окончательной теории. Moscow, Editorial YRCC Publ. (In Russ).
3. Kolmogorov, A.N. (1936). Osnovnye ponyatoya teorii verotatnoستي. [Foundations of the theory of probability]. Moscow, ONTI NKTP SSSR.
4. Feierabend, P. (2007). [Protiv metoda]. [Against method]. Moscow, Custodian Publ. (In Russ).

5. *Burfoot, D.* (2011). Notes on a new philosophy of empirical science. Available at: <https://arxiv.org/abs/1104.5466> (date of access: 29.11. 2025).

6. *Feierabend, P.* (1995). Killing Time. Chicago and London, The University Chicago Press.

7. *Hofer-Szabo, G., Redei, M., Szabo, L.* (2013). The principle of the common cause. Cambridge, Cambridge University Press.

8. *Humphrey, P.* (2008). Probability and Statistics: Essays in Honor of David A. Freedman / Eds. D. Nolan, Speed T. Institute of mathematical statistics Collections, 2, 1–11.

9. *Kac, M.* Enigmas of chance. (1987). California, California University Press. Available at: <https://scilib-physics.narod.ru/Kac/index.html> (date of access: 29.11. 2025).

10. *Mises, P.* (1964). Mathematical theory of probability and statistics. New York and London, Academic Press.

11. *Nola, R., Sankey, H.* (2007). Theories of scientific method. Acumen.

12. *Norton, J.* (2010). How Hume and Mach helped Einstein find special relativity Discourse on a new method. / Eds. M. Domskey, M. Dicrson. Chicago, Open Court.

13. *Reichenbach, H.* (1944). Philosophical foundations of quantum mechanics. California, California University Press.

14. *Renyi, A.* (1955). On a new axiomatic theory of probability. Acta mathematica Academiae Scientiarum Hungarica, 6(3), 285-335.

15. *Vapnik, V.* (2006). Estimation of dependencies based on empirical data. New-York, Springer.

### **Информация об авторе**

*Резников Владимир Моисеевич* – Институт философии и права СО РАН (630090, г. Новосибирск, ул. Николаева 8).  
[mathphil1976@gmail.com](mailto:mathphil1976@gmail.com).

### **Information about the author**

*Reznikov Vladimir Moiseevich* – Institute of Philosophy and Law, SB RAS (8, Nikolaev st., Novosibirsk, 630090).  
[mathphil1976@gmail.com](mailto:mathphil1976@gmail.com).

Дата поступления 03.11..2025

Принята к печати 11.12.2025

УДК: 165.0:82

DOI: 10.15372/PS20250512

EDN: AJXHIQ

**Н.В. Головки****ПРИМЕР Б. УЭЗЕРСОНА,  
ПРАГМАТИЧЕСКИЕ СООБРАЖЕНИЯ  
И ЭПИСТЕМИЧЕСКОЕ ОТНОШЕНИЕ УКОРЕНЕНИЯ**

Цель работы – показать, что пример, который приводят Дж. Фантл и М. Макграт (2002), (прагматические соображения относительно риска приписать желаемый эпистемический статус данному убеждению в данной ситуации влияют на само понятие «эпистемический факт» и, как следствие, мы должны отказаться от эвиденциализма, как представления о том, что необходимые и достаточные условия обоснования определяются исключительно доступными эпистемическими данными), не достигает своей цели. Идея Б. Уэзерсона (2005) заключается в том, чтобы показать, что Дж. Фантл и М. Макграт не проводят различия между «степенью доверия по отношению к истинности пропозиции» (прагматические условия на обоснование) и «степенью доверия к убеждению» (прагматические условия на убеждение). Мы полагаем воспользоваться логикой поиска адекватного эпистемического отношения укоренения, т.е. «метаубеждений, которые объясняют то, почему эти основания являются именно теми основаниями, в силу которого убеждение принимается» (Г. Харман, К. Коркз и др.), связывающего соответствующее убеждение и его основания. Б. Уэзерсон указывает лишь на один возможный вариант фиксации отношения укоренения – через «аналогию с лотерей», – но в более общем случае логика поиска адекватного эпистемического отношения укоренения позволяет более полно указать на то, что мы можем обосновать «практический характер рассуждения» не прибегая к допущениям, которые уже предполагают, что мы обязаны интерпретировать обоснование в терминах «прагматических соображений».

*Ключевые слова:* эпистемология, эвиденциализм, обоснование, прагматические соображения, эпистемическое отношение укоренения.

**N.V. Golovko**

**WEATHERSON'S EXAMPLE,  
PRAGMATIC ENCOACHMENT,  
AND EPISTEMIC BASING RELATION**

The paper aims to show that the example given by J. Fantle and M. McGrath (2002) (pragmatic considerations regarding the risk of attributing the desired epistemic status to a given belief in a given situation affects the very concept of «epistemic fact» and, as a consequence, we must abandon evidentialism as the idea that necessary and sufficient conditions of justification are determined exclusively by available epistemic data) does not achieve its goal. The idea of B. Weatherson's example (2005) is to show that J. Fantle and M. McGrath do not distinguish between «the degree of confidence in the truth of a proposition» (pragmatic conditions on justification) and «the degree of confidence in a belief» (pragmatic conditions on belief). We want to use the logic of an adequate epistemic basing relation, i.e. «meta-beliefs that explain why these reasons are precisely the reasons by virtue of which the belief is held» (G. Harman, K. Korcz, etc.), connecting the corresponding belief and its reasons. B. Weatherson points out on only one possible option for assessing the basing relation – through the «analogy with lotteries», – but in a more general case, the logic of looking for an adequate epistemic basing relation allows us to more thoroughly indicate that we can justify the «practical nature of reasoning» without resorting to assumptions that already suggest that we are obliged to interpret justification in terms of «pragmatic encroachment».

*Keywords:* epistemology, evidentialism, justification, pragmatic encroachment, epistemic basing relation.

Статья «Данные, прагматика и обоснование» Дж. Фантла и М. Макграта, несомненно, является одной из знаковых работ контекстуализма, – в традиции закрепляется пример «Поезд», который теперь становится таким же классическим, как «Аэропорт» С. Коэна или «Банк» К. Дероса, выдвигается не просто аргумент против эвиденциализма, а целая альтернативная парадигма, требующая учета «прагматических соображений» (pragmatic encroachment) в ходе эпистемического обоснования [5].

Однако, в первом же «замечании на полях» (буквально в сноске номер 1), если присмотреться, встречаются сразу несколько нетривиальных тезисов, обсуждение которых может быть даже более интересным, чем итоговый вывод авторов. «Пример с лотереей никогда не представляется как один из вариантов примеров Гетье» [5, р. 87]. Если для вас основной результат Э. Гетье заключается в демонстрации ограниченности логической аргументации, в частности, принципа замыкания (*closure / transmissibility*), то та самая работа Дж. Хоторна «Знание и лотереи» (2004) неожиданно приобретают гораздо более фундаментальный смысл, раскрывая еще один (независимый!) паттерн аргументации в эпистемологии. Еще замечание: «Эвиденциализм представляет собой тезис, что если  $S$  не знает, то это не потому, что у  $S$  не хватает (lack) данных» [там же]. Утверждение, которое (кроме того, что заставляет еще раз обратиться к интуиции, стоящей за примерами Э. Гетье) буквально заставляет нас вернуться к определению эвиденциализма и еще раз проверить то, насколько хорошо мы его понимаем. Обе эти ремарки показывают, что «эвиденциализм» – это не просто известный тезис Э. Кони и Р. Фелдмана: «(EJ) Доксастическая установка  $D$  по отношению к пропозиции  $P$  эпистемически обоснована для  $S$  в  $t$ , если и только если обладание (*having*)  $D$  отвечает (*fits*) данным, которые есть у  $S$  в  $t$ » [4, р. 83]. Ключевое понятие, конечно же, – «данные», – которые есть не что иное как «убеждения», которые играют определенную роль в принятой модели эпистемического обоснования. Однако в общем случае, кроме прочего, всегда есть возможность по-разному понимать и то, что значит «обладать данными», и то, какие ментальные состояния следует считать данными, и то, как эти данные свидетельствуют в пользу  $P$ .

Ниже мы еще раз остановимся на логике рассуждений Дж. Фантла и М. Макграта. При определенных допущениях (в частности, если принять во внимание позицию Дж. Кванвига о том, что риск совершить ошибку является всего лишь еще одним фактором, подчеркивающим принципиально субъективный характер приписывания эпистемического статуса убеждению) у нас есть возможность представить пример «Поезд» как одну из форм скептического аргумента «от ошибки» [3]. Брай-

ан Уэзерсон предлагает рассмотреть «прагматические соображения» и соответствующую связь обоснования и данных в терминах лотереи, т.е. буквально воспользоваться первым из приведенных «замечаний на полях». Более того, принимая во внимание другое «замечание на полях», он предлагает развести «степенью доверия по отношению к истинности пропозиции» (прагматические условия на обоснование) и «степенью доверия к убеждению» (прагматические условия на убеждение). Как следствие, Б. Уэзерсон показывает, что у нас есть возможность обосновать «Если  $S$  обоснованно принимает что  $P$ , тогда  $S$  обоснован в использовании  $P$  как посылки в практическом рассуждении» *не прибегая* к допущениям, которые уже предполагают, что мы обязаны интерпретировать обоснование в терминах «прагматических соображений». На наш взгляд, рассуждения Б. Уэзерсона являются частным примером более общего случая, который показывает то же самое. Если перейти к логике обнаружения эпистемического отношений укоренения, которое, по сути, в рамках принимаемых ограничений, является «мета-убеждением, которое объясняет то, почему эти основания является именно теми основаниями, в силу которых данное убеждение принимается» (Г. Харман, К. Коркз и др.), то основания, связывающие «прагматический контекст» (например, акцентирующий внимание на концепции увеличения полезности действия) и собственно «эпистемический контекст» (касающийся истинности обсуждаемой пропозиции) обсуждения «степени доверия к убеждению» будут разными. А значит, никаких оснований говорить собственно о «прагматическом характере обоснования» не будет.

### Пример Б. Уэзерсона

Суть рассуждений Дж. Фантла и М. Макграта заключается в том, что они предлагают ввести в оборот еще одно необходимое условие эпистемического обоснования, а именно – прагматическое условие: «(PC)  $S$  обоснован в принятии (believing) что  $P$  только если для  $S$  рационально выбрать (refer) [оказаться в ситуа-

ции] как если бы  $P$ » [5, р. 77]<sup>1</sup>. В классической парадигме «обоснование» – это то, что переводит истинное мнение в знание и в этом смысле тезис о том, что «то, что переводит истинное убеждение в знание чувствительно (*sensitive*) по отношению к практическому окружению (*practical environment*), в котором находится субъект», по крайней мере, можно рассматривать как нетривиальное ее расширение. Однако, не все с этим согласны. Цель, которую ставит перед собой Б. Уэзерсон, заключается в том, чтобы показать, что эпистемическое обоснование не обязательно быть «прагматически чувствительным». Ошибка, которую по его мнению допускают Дж. Фантл и М. Макграт, заключается в том, что «(PC) не следует рассматривать как *прагматическое условие на обоснование*, оно должно пониматься как *прагматическое условие на убеждение*» (курсив наш – Н. Г.) [11, р. 417]. Логика рассуждений Б. Уэзерсона заключается в том, что нельзя просто так перейти от «(B)  $S$  принимает что  $P$  если и только если  $S$  выбирает (*refer*) [оказаться в ситуации] как если бы  $P$ » к «(JB)  $S$  обоснован в принятии что  $P$  только если  $S$  обоснован в выборе [оказаться в ситуации] как если бы  $P$ ». «Будет ошибкой полагать,

---

<sup>1</sup> Мы исходим из того, что «первое условие» (*belief condition*) традиционного трехчастного определения знания – « $S$  believes that  $P$ », как минимум, следует обозначать как «условие принятия»: « $S$  принимает убеждение что  $P$ ». (Нам известно, что у Э. Гетье, например, оно стоит на втором месте, но есть что-то странное рассуждать о характеристиках  $P$  – истинности и обоснованности до того, как это  $P$  зафиксирует для себя субъект). Речь идет о конкретном убеждении, которое по каким-то причинам сформировалось в сознании субъекта и, более того, выделяется субъектом (из всех прочих осознанных и неосознанных убеждений, которые мы предполагаем, что находятся в пределах когнитивного доступа субъекта) как релеватное (по каким-то причинам) для данной ситуации. Также подчеркнем, что форма «что  $P$ » (*that P*) в данном случае не случайна. Если мы принимаем пропозициональную теорию знания, то теоретически мы обязаны различать пропозицию « $P$ » и убеждение «что  $P$ ». Но здесь, конечно, многое зависит от того, насколько строго мы готовы (в первую очередь, в постоянной практике употребления выражений, содержащих соответствующие термины) принимать и постоянно напоминать себе о различных оговорках, в том числе, стоящую за семантической концепцией истины и  $T$ -предложением А. Тарского. Например, мы можем не писать «субъект  $S$  обоснован в принятии что  $P$ » и ограничиться «субъект  $S$  обоснован в принятии  $P$ », поскольку «принятие  $P$ » по определению означает формирование убеждения «что  $P$ », – но это не значит, что мы забыли о том, о чем действительно хотим сказать.

что мы вывели (JB) из (B) навесив нормативный оператор. (JB) не будет прагматическим условием на обоснование точно также как и (PC), которое мы получили заменив “обосновано” на “рационально”, поскольку мы просто вывели условное утверждение (conditional) из двухсторонней условной зависимости (biconditional)» [11, p. 418]<sup>1</sup>.

Предмет обсуждения Дж. Фантла и М. Макграта – эвиденциализм в самом широком смысле, например, как тезис о том, что «субъект  $S$  обоснован в принятии  $P$  если и только если у  $S$  есть данные в пользу  $P$ ». И в этом смысле, нет никаких проблем с тем, чтобы привести примеры, в которых будет нарушаться следующее представление: «если у нас есть два субъекта  $S_1$  и  $S_2$ , то необходимо, что если  $S_1$  и  $S_2$  обладают одними и теми же данными в пользу  $P$ , то они будут одинаково обоснованы в принятии  $P$  и более того,  $S_1$  будет обоснован в принятии  $P$  если и только если  $S_2$  также обоснован в принятии  $P$ ». Именно об этом пример «Поезд»: практическая значимость обстоятельств, в которых находится субъект, опасность сделать неверный вывод из имеющихся данных, и заставляют Дж. Фантла и М. Макграта говорить о том, что в одном контексте субъект обоснован, а в другом – нет, заставляют говорить о том, что прагматические соображения влияют на обоснование. С точки зрения Б. Уэзерсона проблема заключается в том, что Дж. Фантл и М. Макграта не говорят о том, о какой степени обоснованности  $P$  идет речь. По умолчанию считается, что мы говорим о знании  $P$ , но в общем случае это не так. «Получается, что вся претензия к эвиденциализму строится на том, что неверно утверждение вида: для любых двух субъектов  $S_1$  и  $S_2$  и для любой эпистемической степени  $\alpha$ , если  $S_1$  и  $S_2$  оба принимают в отношении  $P$  степень  $\alpha$ , то  $S_1$  будет обоснован в принятии  $P$  если и только если  $S_2$  также обоснован в принятии  $P$ . Но почему? Последнее утверждение сообщает нам нечто относительно прагматических соображений и их роли в психологии и в философии сознания, но никак не относительно прагматических соображений и их роли в эпистемологии» (курсив наш – Н. Г.) [11, p. 419]. В логике замечания « $S$  не знает не потому, что

---

<sup>1</sup> Можно заметить, что здесь Б. Уэзерсона также соглашается с невозможностью обойтись сугубо логической аргументацией в эпистемологии (см., например, [1]).

у него не хватает данных» Б. Уэзерсон предлагает разделить «прагматическую чувствительность понятие убеждения» и «обыденное, прагматически не-чувствительное понятие эпистемического обоснования».

Соответственно Б. Уэзерсон приводит следующий пример: «Гарри и Луиза стоят на платформе, на которой останавливаются два типа поездов – почтовые и экспрессы (не останавливаются на следующих пяти станциях). Проблема в том, что глядя на приближающийся поезд сложно сказать, какой именно в настоящий момент подходит к станции. По условию задачи им нужно попасть на станцию после следующей. При этом Луиза едет отдыхать и понимает, что в случае ошибки, для нее не будет проблем пересест на следующей остановке экспресса, – это небольшое путешествие не будет иметь никакого значения. Но у Гарри запланирована встреча, на которую нельзя опоздать. Стоя на платформе, они слышат разговор стоящих рядом ожидающих, и один из них говорит другому: “Это почтовый”. Времени проверить то, какой именно поезд подходит, нет» [11, р. 434]. Дж. Фантл и М. Маграт полагают, что несмотря на то, что эпистемически ситуация для Гарри и Луизы ничем не различается (у них один и тот же объем эпистемических данных), очевидно, сопутствующие «прагматические соображения» разные, а значит, Луиза обоснована в том, что приближающийся поезд почтовый, а Гарри – нет. «Но означает ли это, – спрашивает Б. Уэзерсон, – что само *понятие эпистемического обоснования должно быть чувствительным по отношению к прагматическому интересу субъекта?* Нет. Все происходящее можно описать как предложение сыграть в игру. Если поезд – экспресс, то они потратят на дорогу больше времени. Для Луизы это игра с низкими ставками, она ничем не рискует. Но для Гарри это игра с высокими ставками. Что каждый из них должен сделать? Какое убеждение каждый из них должен принять?» [11, р. 435].

Подчеркнем, Б. Уэзерсон сознательно интерпретирует ситуацию, описываемую Дж. Фантлом и М. Макгротом в терминах «примера с лотереей», а также формулирует не один вопрос: «кто обоснован?», а два: «что в этой ситуации предпочтительнее сделать?» и «о каком именно сформированном убеждении, которое

каждый из них принимает в данной ситуации, идет речь?»<sup>1</sup>. С точки зрения Б. Уэзерсона, описывая рациональное поведение Гарри и Луизы (и если мы воспользуемся критерием увеличения полезности действия), Луиза должна сесть на поезд, а Гарри – нет. Гораздо интереснее предлагаемый ответ на второй вопрос: «Фундаментальное допущения, которое я хотел бы сделать, заключается в том, что основной предмет эпистемической оценки – это степень доверия убеждению (*degree of belief*), либо то, что называется состоянием уверенности (*state of confidence*) по отношению к данному убеждению. С этой точки зрения Гарри и Луизы, они обоснованы принять *одну и ту же степень доверия убеждению* (это отвечает общей логике эвиденциализма, когда степень доверия должна быть пропорциональна имеющимся данным, а условный объем данных по отношению к формируемому состоянию уверенности у Гарри и у Луизы одинаковый – *Н. Г.*). Они оба формируют состояние убеждения, отвечающее определенной степени доверия, – которая не является степенью абсолютно максимальной (*certain*), – что поезд почтовый. И тогда почему мы говорим, что Луиза обоснована, а Гарри – нет? Потому что тут в рассмотрение включается контекст. Контекст вводит представление о *предпочтительном действии*, которое должны совершить Гарри и Луиза. Гарри не должен садиться на поезд, т.к. это действие (на основании имеющейся у него информации)

---

<sup>1</sup> В данном случае русский язык может сыграть злую шутку. Вопрос, который задает Б. Уэзерсон звучит как: «What should each of them believe?». И в данном случае, конечно, речь не может идти о том: «Во что каждый из них может верить?». Следуя Р. Чизолму, предмет эпистемологии – эпистемический статус «убеждений» (*belief*), а значит, никаких «вер» или «мнений». Эпистемический статус убеждения может быть как «низкий», например, «очевидно ложное» (*obviously false*), так и «высокий», например, «за рамками разумного сомнения» (*beyond reasonable doubts*), и в данном случае этимология, стоящая за словами «вера» или «мнение», на наш взгляд, будет только мешать. Отсутствие в русском языке адекватных средств фиксации предикатов, отвечающих понятию убеждение, – это эмпирический факт. Но это не означает, что мы должны, в угоду благозвучия, опускать то, о чем собственно говорится. В данном случае «What should each of them believe?» указывает на конкретные убеждения (а также на их соответствующие различные эпистемические статусы), которые возникнут (или по каким-то причинам выбираются ими как значимые) в головах Гарри и Луизы. И перевод «Какое именно убеждение каждый из них принимает?» с заменой «believe» на «принять убеждение» будет наиболее корректным.

не является для него рациональным (с точки зрения практической рациональности и концепции увеличения полезности действия – Н. Г.). В данном случае, практические соображения важны не потому, что они влияют на *степень доверия по отношению к истинности пропозиции* (что собственно и является предметом эпистемологии). Они важны потому что оказывают влияние на то, *относится ли эта разумная с точки зрения субъекта степень доверия к убеждению* (что является предметом исследования психологии и философии сознания). Здесь нет оснований считать прагматические соображения (concerns) частью эпистемологии» (курсив наш – Н. Г.) [11, р. 435]. Перед нами два вопроса. Эпистемологический вопрос об «обоснованности убеждения» (степень доверия по отношению к его истинности), ответ на который должен (в силу нормативности концепции обоснования) носить, в каком-то смысле, объективный характер. И вопрос о «степени доверия убеждению» в контексте его принятия в конкретной ситуации. Б. Уэзерсон интерпретирует этот второй (психологический в широком смысле) вопрос и соответствующий контекст как игру с низкими и высокими ставками. Это то, что не имеет никакого отношения к истинности пропозиции, т.е. к предмету эпистемологии.

Сложно сказать, удастся ли Б. Уэзерсону таким образом исключить прагматические соображения из эпистемологии. Здесь, в контексте обсуждения проблемы прагматических соображений, в том виде, как ее задают Дж. Фантл и М. Макграт, судя по всему, проблема действительно может быть в том, что они не придают значения тому, о каком эпистемическом статусе убеждения «на платформу прибывает почтовый поезд» идет речь. Абсолютность «знать» лишает нас возможности развести контексты, о которых говорит Б. Уэзерсон, – мешает развести «объективный эпистемический статус убеждения» (подчеркнем, «объективность» здесь крайне размытое понятие) и «прагматически чувствительное понятие убеждения». И как только мы начнем говорить применительно к данному убеждению о «любой эпистемической степени  $\alpha$ », необходимость переносить прагматические соображения на обоснование (именно то, что фиксирует эпистемический статус) отпадет. Тем более, что вопрос о том, можем ли мы действительно что-либо именно знать в обыденной жизни не является простым.

## Прагматически чувствительное понятие убеждения

Ранее мы уже обращали внимание на то, в каком смысле, теоретически, применительно к анализу рассуждений Дж. Фантла и М. Макграта, а также к сопутствующей парадигме «прагматических соображений», можно интерпретировать «прагматическую чувствительность понятие убеждения». С одной стороны, мы можем согласиться с рассуждением Дж. Кванвига о том, что «в случае с поездом [в работе Дж. Фантла и М. Макграта] подчеркивается акцент на риске совершить ошибку. Однако сам по себе акцент на риске совершить ошибку не вносит никакой анти-эвиденциалистский элемент в теорию знания. Привязка акцента на риске совершить ошибку к обладанию знанием всего лишь ведет к осознанию большей субъективности знания. Тут важно посмотреть на факторы, которые влияют на то, какой акцент мы выбираем. Если это акценты, касающиеся практических аспектов ситуации, которая влечет потерю (loss) знания, – это одно, но если акцент касается именно риска совершить ошибку, то очевидный фактор, в данном случае, – это субъективный характер принимаемого решения. И субъективность подобного рода не может быть частью критики эвиденциализма» [9, р. 84] (см. [3]). С другой стороны, мы можем согласиться с парадигмой, в которой работает А. Круглянский: «любая эпистемология – “обыденная”, “нормативная”, “научная” – отражает общий, одинаковый для всех психологический процесс получения (acquisition) знания, который подвержен влиянию одинаковых для всех психологических факторов, недостатков и уязвимостей. “Знание” – это пропозиция, по отношению к которой субъект обладает определенной степенью психологической уверенности (confidence). В этом смысле получение знания предполагает, что (1) пропозициональное содержание должно быть как-то произведено (generated) в сознании и что (2) данная степень уверенности должна быть как-то в сознании этой пропозиции придана (bestowed). Именно психологическая мотивационная составляющая *бессознательно* буквально управляет когнитивным процессом по производству, закреплению за убеждением соответствующего эпистемического статуса и его последующей переоценкой, а также обеспечивает основу для других сопутствующих когнитивных и аффективных, в том числе, поведенческих, реакций субъекта на подтверждение и/или опровержение имеющейся информа-

ции» (курсив наш – Н. Г.) [8, р. 9]. (см. [2]). На наш взгляд, существует и третий вариант того, как можно интерпретировать «прагматически чувствительное понятие убеждения» не смешивая (на чем настаивает Б. Уэзерсон) «степень доверия по отношению к истинности пропозиции» и «степень доверия к убеждению».

Рассмотрим пример «Цыганский адвокат», который приводит К. Лере: «Предположим, что адвокат, защищающий подсудимого, абсолютно доверяет предсказаниям на картах. Карты говорят, что клиент невиновен. Это повод для адвоката еще раз проанализировать улики, и ему неожиданно удается построить очень сложную, но обоснованную (valid), линию рассуждений, подтверждающую, что клиент невиновен. Однако эмоциональная вовлеченность адвоката, общественная шумиха вокруг обсуждаемого дела, а также то, что его линия рассуждений гораздо сложнее линии рассуждений прокурора, мешает остальным согласиться с ним. И если бы не непреложная (unshakable) убежденность адвоката в предсказаниях на картах, то она бы и сама согласилась с прокурором. Только уверенность в картах дает основание (reason) сказать, что адвокат знает, что клиент не совершал убийство» [10, р. 312]. Что произошло? Как отмечает К. Коркз, в данном случае «адвокат рационализирует свое убеждение о том, что его клиент невиновен, но он не имеет надлежащих оснований (not justified) полагать, что его клиент невиновен, – его убеждение, что клиент невиновен, основано на “плохой причине“, – он апеллирует к картам» [7, р. 528]. На наш взгляд, эту интерпретацию К. Коркза описанного примера К. Лере можно использовать в рассуждениях Б. Уэзерсона. Здесь К. Коркз говорит о «плохой причине» и использует это наблюдение для того, чтобы предположить, что «более адекватным будет отношение, которое свяжет основания и убеждение посредством некоторого *мета-убеждения, которое объясняет то, почему это основание является именно тем основанием, в силу которого убеждение принимается*» (курсив наш. – Н. Г.) [7, р. 532]. К слову, К. Лере также говорит о том, что «вопрос о том, дают ли основания знание, зависит от того, что данные основания *объясняют*» [10, р. 311]. Вернемся к примеру Б. Уэзерсона, здесь «мета-убеждение», о котором идет речь, – это то, что объясняет Луизе и Гарри то, почему для них рационально считать, что приближающийся поезд почтовый.

Прагматические соображения, не в последнюю очередь, появляются как следствие возможности смешать эпистемическую рациональность и практическую рациональность, которая в данном случае определяется практическим характером вывода, который делают Луиза и Гарри. В этом смысле, Б. Уэзерсон отмечает, что «наша задача состоит в том чтобы показать достоверность (JP): Если  $S$  обоснованно принимает что  $P$ , тогда  $S$  обоснован в использовании  $P$  как посылки в практическом рассуждении (reasoning), – не прибегая к допущениям, которые можно использовать как аргумент в пользу обоснования концепции прагматических допущений в отношении обоснования» (курсив наш – *Н. Г.*) [11, р. 436]. На наш взгляд, в данном случае апелляция к «мета-убеждению, которое объясняет то, почему это основание является именно тем основанием, в силу которого убеждение принимается» может объяснить эту интуицию Б. Уэзерсона, а также закрепить соответствующее понимание «прагматически чувствительного понятия убеждения».

Собственная логика рассуждений Б. Уэзерсона связана с анализом силлогизма:

P

Если P, тогда S должен (should) принять ставку

Таким образом, S должен принять ставку

и демонстрацией того, в каком смысле мы можем принять его или не принять, когда рассматриваем (PC) « $S$  обоснован в принятии что  $P$  только если для  $S$  рационально выбрать [оказаться в ситуации] как если бы  $P$ » Дж. Фантла и М. Макграта. В конце концов Б. Уэзерсон указывает на то, что «принимая во внимание роль импликации в ходе практического рассуждения, мы можем оказаться в ситуации, когда независимо от того, являются ли истинными обе посылки, субъект может быть обоснован в принятии (believing) заключения. Ведь проблема не в том, является ли принятие что  $P$  рациональным на основании данных, которые есть у субъекта в пользу  $P$ . Проблема в том, чтобы найти посылку, которая, принимая во внимание P, сделает заключение – решение принять ставку – рациональным. И очевидно Фантл и Макграт в качестве такой посылки (а значит и основания (reason), на основании которого принимается соответствующее убеждение – *Н. Г.*) выбирают максимизацию ожидаемой полезности» [11, р. 440]. Речь идет о том, сама природа рассматриваемого практического

силлогизма такова, что даже если обе посылки истинны (а для Дж. Фантла и М. Макграта  $P$  – истинно, т.к. по сути они работают только с одним эпистемическим статусом – знанием), то заключение из них может не следовать. Другими словами, здесь всегда есть опасность попасть в ситуацию «Цыганский адвокат», когда основания «на основании которых субъект обосновывает убеждение» (в собственно эпистемическом смысле, как отношение к имеющимся данным и к истинности соответствующей пропозиции) могут быть не теми основаниями «на основании которых субъект это убеждение принимает». Проблема может быть в природе практической рациональности. В данном случае «апелляция к максимизации ожидаемой полезности» – это и есть то самое «мета-убеждению, которое объясняет то, почему это основание является именно тем основанием, в силу которого убеждение принимается», которое принимают Дж. Фантл и М. Макграт. Для них обоснование достоверности ( $JP$ ), о котором говорит Б. Уэзерсон, изначально содержит посылку «прагматического допущения». В то же время, более честная постановка, согласно Б. Уэзерсону, заключается в том, чтобы посмотреть на ситуацию более широко, – с точки зрения «прагматических соображений и их роли в психологии и философии сознания»: «Мы исходим из того, что  $P$  обосновано если и только если степень уверенности в  $P$  достаточна для того, чтобы принять  $P$  в данном контексте, и субъект обоснован в приписывании данной степени уверенности для  $P$ . Мы можем сказать, что субъект принимает что  $P$  потому что (а) он предпочитает (в каком-то релевантном смысле) принять что  $P$ , чем не принять что  $P$ , и (б) имеющиеся условные утверждения, содержащие  $P$ , (conditionalising on  $P$ ) не могут изменить решения субъекта сделать ставку (субъект будет склонен сделать ставку независимо от условий, налагаемых на  $P$ ). Но *это еще не все* из того, что нужно для того, чтобы считать, что субъект принимает убеждение. Мы также должны спросить, будут ли условные утверждения, содержащие  $P$ , изменять содержание *любой другой посылки аргумента*» (курсив наш – Н. Г.) [11, р. 441]. В определенном смысле, здесь Б. Уэзерсон подчеркивает, что хороший аргумент должен быть не только «верным» (valid), но и «достоверным» (sound). Понятия «обоснование» и «рациональное принятие решения» – это разные понятия и попытка объединить их в одном

аргументе (как, по-видимому, и поступают Дж. Фантл и М. Макграт) буквально наталкивается на проблему достоверности (в смысле soundness) вывода. И в этом смысле, на чем настаивает Б. Уэзерсон, нет никакой необходимости рассматривать интуицию «прагматического соображения» как часть концепции обоснования.

\* \* \*

Приведенный пример К. Лере и его интерпретация К. Коркзом – это пример рассуждений, раскрывающих понятие «эпистемического отношение укоренения» (epistemic basing). Убеждение обосновано, только когда оно надлежащим образом опирается на адекватные основания. Само по себе наличие адекватных оснований – это лишь достаточное условие обоснования. Для того чтобы убеждение было эпистемически обосновано, необходимо показать, что предполагаемый эпистемический статус убеждения должным образом следует из соответствующих оснований, что эти основания, как отмечает Г. Харман: «являются именно теми основаниями, *в силу которых* (for which) убеждение принимается (held)» [6, р. 26]. Но именно об этом и говорит Б. Уэзерсон, разделяя «степень доверия по отношению к истинности пропозиции» и «степень доверия к убеждению». В рассуждениях Дж. Фантла и М. Макграта, равно как и в других рассуждениях тех, кто якобы стремится «расширить» представление об эпистемической рациональности (в данном случае, за счет привлечения логики практического рассуждения), может быть заложена ошибка не выполнения надлежащего эпистемического отношение укоренения, – которая приводит к ситуации, аналогичной описанной в примере «Цыганский адвокат». И в этом смысле, пример Б. Уэзерсона с акцентом на том, что прагматические соображения не обязательно должны касаться понятия обоснования, но затрагивают «прагматически чувствительное понятие убеждения» и вопросы его принятия (т.е. по сути вопросы «мета-убеждений, которые объясняют то, почему эти основания являются именно теми основаниями, в силу которого убеждение принимается») более чем показателен. В данном случае концепция эпистемического отношений укоренения – это инструмент, который объясняет то, почему рассуждения Б. Уэзерсона будут состоятельными. Потому что сопутствующие рассуждения будут объяснять (мы будем понимать

наличие соответствующего мета-убеждения как объяснение) субъекту, в каком смысле субъект принимает убеждение, – и этот контекст будет действительно «прагматически чувствительным». Сама апелляция к необходимости выписывания адекватного эпистемического отношения укоренения в каждом конкретном случае может считаться независимым аргументом в пользу интуиции «прагматических соображений», но только в смысле принятия убеждения, но не его обоснования. Б. Уэзерсон указывает лишь на один возможный вариант фиксации отношения укоренения – через «аналогию с лотерей», но в более общем случае логика поиска адекватного эпистемического отношения укоренения позволяет более полно указать на то, что мы можем обосновать «практический характер рассуждения» не прибегая к допущениям, которые уже предполагают, что мы обязаны интерпретировать обоснование в терминах «прагматических соображений».

### Литература

1. Головки Н.В. Чего нет в известной статье Эдмунда Гетье // Сибирский философский журнал. 2023. № 1. С. 105–126.
2. Головки Н.В. Контекстуализм, парадокс Круглянского и натурализованная эпистемология // Философия науки. 2024. № 3. С. 83–108.
3. Головки Н.В. Эвиденциализм, тезис Фантла–Макграта и аргумент от ошибки // *Respublica Literaria*. 2023. Т. 4, № 4. С. 5–13.
4. Conee E., Feldman R. *Evidentialism: Essays in Epistemology*. Clarendon Press, 2004.
5. Fantl J., McGrath M. Evidence, Pragmatics, and Justification // *Philosophical Review*. 2002. Vol. 111 (1). P. 67–94.
6. Harman G. *Thought*. Princeton, NJ : Princeton University Press, 1973.
7. Korcz K. The Causal-Doxastic Theory of the Basing Relation // *Canadian Journal of Philosophy*. 2000. Vol. 30, № 4. P. 525–550.
8. Kruglanski A. *Lay Epistemics and Human Knowledge*. Springer Science, 1989.
9. Kvanvig J. Against Pragmatic Encroachment // *Logos & Episteme*. 2011. Vol. 2. No. 1. P. 77–85.
10. Lehrer K. How Reasons Give us Knowledge, or the Case of the Gypsy Lawyer // *Journal of Philosophy*. 1971. Vol. 68, № 10. P. 311–313.
11. Weatherston B. Can We Do Without Pragmatic Encroachment? // *Philosophical Perspectives*. 2005. Vol. 19. No. 1. P. 417–443.

## Reference

1. *Golovko, N.V.* (2023). Chego net v izvestnoy statie Edmunda Getie [What is not mentioned in the famous article by Edmund Gettier]. *Siberian Journal of Philosophy*, 1, 105-126.
2. *Golovko, N.V.* (2024). Kontekstualizm, paradoks Kruglyanskogo i naturalizovannay epistemologia [Contextualism, Kruglanski's paradox and naturalized epistemology]. *Philosophy of Science*, 3, 83–108.
3. *Golovko, N.V.* (2023). Evidentsializm, tezis Fantla–Makgrata i argument ot oshibki [Evidentialism, Fantl–McGrath's argument and the argument from error]. *Respublica Literaria*, 4 (4), 5–13.
4. *Conee, E., Feldman, R.* (2004). *Evidentialism: Essays in Epistemology*. Clarendon Press.
5. *Fantl, J., McGrath, M.* (2002). Evidence, Pragmatics, and Justification. *Philosophical Review*, 111 (1), 67–94.
6. *Harman, G.* (1973) *Thought*. Princeton, NJ: Princeton University Press
7. *Korcz, K.* (2000) The Causal-Doxastic Theory of the Basing Relation. *Canadian Journal of Philosophy*, 30 (4), 525–550.
8. *Kruglanski, A.* (1989). *Lay Epistemics and Human Knowledge*. Springer Science.
9. *Kvanvig, J.* (2011). Against Pragmatic Encroachment. *Logos & Episteme*, 2 (1), 77-85.
10. *Lehrer, K.* (1971) How Reasons Give us Knowledge, or the Case of the Gypsy Lawyer. *Journal of Philosophy*, 68 (10), 311–313.
11. *Weatherson, B.* (2005). Can We Do Without Pragmatic Encroachment? *Philosophical Perspectives*, 19 (1), 417–443.

## Информация об авторе

*Головко Никита Владимирович.* Институт философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8).  
golovko@philosophy.nsc.ru

## Information about the author

*Golovko, Nikita Vladimirovich.* Institute of Philosophy and Law of the Siberian Branch of the Russian Academy of Sciences (8, Nikolaeva Str., Novosibirsk, 630090, Russia).  
golovko@philosophy.nsc.ru

Дата поступления 27.08.2025

УДК: 122

DOI: 10.15372/PS20250513

EDN: ALKQJH

**А.Ю. Сторожук**

**ПРОБЛЕМА ИЗМЕРЕНИЯ  
В КВАНТОВОЙ МЕХАНИКЕ  
С ТЕРМОДИНАМИЧЕСКОЙ ТОЧКИ ЗРЕНИЯ**

В статье исследуется проблема измерений в квантовой механике, тесно связанная с принципом неопределённости Гейзенберга и философскими дискуссиями об онтологической или эпистемологической природе квантовой неопределённости. Автор прослеживает эволюцию взглядов на данную проблему – от концепции скрытых параметров Эйнштейна до современных трактовок, рассматривающих измерение как нарушение замкнутости квантовой системы и процесс получения информации. Цель исследования – связать современные трактовки проблемы наблюдения с пониманием М. Планком основных термодинамических величин (в частности, энтропии) и сопоставить его с современным информационным подходом к проблеме квантовых измерений, в основе которого лежит понятие энтропии по Шеннону.

В работе излагается смысл принципа неопределённости Гейзенберга и его интерпретаций (онтологической – Бора и Гейзенберга, эпистемологической – Эйнштейна); показывается связь между термодинамической трактовкой процесса измерения (на основе идей М. Планка) и теорией информации (подход К. Шеннона); проводится сравнительный анализ понятий энтропии у Планка и Шеннона: их физической и информационной интерпретации, области применения и смысловых акцентов; рассматриваются конкретные примеры, иллюстрирующие изменение энтропии системы при удалении частицы (идеальный газ, кристалл при температуре, близкой к абсолютному нулю, система со связанными состояниями).

Основные выводы статьи состоят в том, что энтропия в термодинамике (Планк) и теории информации (Шеннон) имеет общую статистическую природу, но различается по контексту и направлению «неопределённости»: у Планка это свойство системы для наблюдателя, у Шеннона – свойство источника для получателя. Сопоставление термодинамического и информационного подходов открывает перспективы для нового понимания проблемы измерений в квантовой механике через призму теории информации. Значимость работы заключается в междисциплинарном синтезе идей термодинамики, квантовой механики и теории информации, позволяющем глубже осмыслить природу квантовой неопределённости и процесс измерения как информационный обмен.

*Ключевые слова:* проблема измерения, влияние наблюдателя, энтропия, информатика.

**Storozhuk A.Y.**

### **THE MEASUREMENT PROBLEM IN QUANTUM MECHANICS FROM A THERMODYNAMIC POINT OF VIEW**

This article explores the problem of measurement in quantum mechanics, closely related to the Heisenberg uncertainty principle and philosophical debates about the ontological or epistemological nature of quantum uncertainty. The author traces the evolution of views on this problem—from Einstein's concept of hidden variables to modern interpretations that view measurement as a violation of the closed nature of a quantum system and a process of obtaining information. The goal of the study is to connect modern interpretations of the observation problem with Max Planck's understanding of fundamental thermodynamic quantities (in particular, entropy) and compare it with the modern information approach to the problem of quantum measurements, which is based on Shannon's concept of entropy. The paper presents the meaning of the Heisenberg uncertainty principle and its interpretations (the ontological interpretation of Bohr and Heisenberg, and the epistemological interpretation of Einstein). The connection between the thermodynamic interpretation of the measurement process (based on the ideas of M. Planck) and information theory (K. Shannon's approach) is demonstrated. A comparative analysis of the concepts of entropy in Planck and Shannon is provided: their physical and informational interpretations, areas of application, and semantic emphases. Specific examples illustrating the change in

a system's entropy upon the removal of a particle (an ideal gas, a crystal at a temperature close to absolute zero, a system with bound states) are considered. The main conclusions of the article are that entropy in thermodynamics (Planck) and information theory (Shannon) have a common statistical nature, but differ in context and the direction of "uncertainty": for Planck, it is a property of the system for the observer, while for Shannon, it is a property of the source for the receiver. The comparison of the thermodynamic and information approaches opens up prospects for a new understanding of the problem of measurements in quantum mechanics through the prism of information theory. The significance of this work lies in its interdisciplinary synthesis of ideas from thermodynamics, quantum mechanics, and information theory, which allows for a deeper understanding of the nature of quantum uncertainty and the measurement process as an information exchange.

*Keywords:* measurement problem, observer influence, entropy, computer science.

## Введение

Проблема измерений в квантовой механике обсуждается в течение длительного времени и тесно связывается с непредсказуемым влиянием наблюдателя на систему. Измерение приводит к воздействию на систему недетерминистическим образом и невозможностью определить одновременно значения связанных параметров с произвольной наперед заданной точностью. Последняя ограничивается принципом неопределенности Гейзенберга: "имеется принципиальный предел для точности определения этих величин в один и тот же момент времени" [1, с. 224]. Этот предел связан со значением постоянной Планка, а произведение неопределенностей сопряженных величин, например, времени и энергии, или импульса и координаты не может быть меньше постоянной Планка.

Философские дискуссии касались источника данной неопределенности: присуща ли она природе *онтологически* или она *эпистемологическая*, то есть возникающая в процессе познания? В первом случае требуется признать в природе нарушение детерминистической причинности и ее принципиально вероятностный характер. Данная позиция вызвала бурный протест со стороны классической интерпретации квантовой механики. По этому по-

воду Эйнштейн сказал: "Бог не играет в кости" и, чтобы объяснить ситуацию неопределенности и вероятностный характер измерений, выдвинул концепцию скрытых параметров. Предложение Эйнштейна смещало рассмотрение данной проблемы в эпистемологическую плоскость, где речь шла о неполноте наших знаний. Однако, позже теория скрытых параметров была экспериментально опровергнута [2, 3, 4], что переместило рассмотрение проблемы измерения в онтологическую сторону.

Было предложено несколько подходов один из которых состоит в том, чтобы рассматривать процесс измерения с термодинамической точки зрения, как нарушение замкнутости квантовой системы, в ходе которого считывается информация. Интересно, что термодинамическая интерпретация процесса измерения близка к пониманию системы самим М. Планком - автором понятия "квант". Поэтому сначала мы обратимся к работам М. Планка, для поиска точек соприкосновения с современными информационными трактовками.

*Задачей* данной статьи является анализ понимания М. Планком основных термодинамических величин и сравнение с современным пониманием для обнаружения общих черт, которые могут стать в дальнейшем фундаментом для применения информационного подхода, с точки зрения которого будет рассматриваться проблема измерения в квантовой механики.

М. Планк занимался разработкой Термодинамики и решал проблему излучения абсолютно черного тела, известную как ультрафиолетовая катастрофа. Чтобы достичь соответствия с экспериментом, он вводит некую константу  $h$ , получившую в дальнейшем название постоянной Планка. Позже квант интерпретируется как порция поглощаемой и излучаемой энергии. Величина, непрерывная в термодинамике, становится дискретной при привязке ее Бором к строению атомов и требуется новый математический аппарат – операторы. Действие оператора описывает процедуру измерения, но операторы непрерывны. То обстоятельство, что операторы не коммутируют, находит свое выражение в принципе неопределенности Гейзенберга.

Принцип неопределенности говорит о невозможности получения полной информации о системе: чем точнее измеряется одна величина, тем менее определенной оказывается коммутирующая

с нею. Физики ведут философские дискуссии о принципе неопределенности, трактуя его либо только в эпистемологическом ключе, как делает Эйнштейн развивая теорию скрытых параметров (Бог не играет в кости), либо в онтологическом, как Бор и Гейзенберг. Последняя интерпретация противоречит классическому пониманию детерминистической причинности, так как допускает онтологическое существование в природе вероятности и случайности.

Неопределенность квантовых систем начинают связывать с процедурой измерения. Активно обсуждается проблема измерения в квантовой механике и влияние наблюдателя. В 2000-е гг. измерение в квантовой механике рассматривают как нарушение замкнутости системы в процессе получения информации о системе. Параллельно квантовой теории развивается информатика и математическая теория информации где фигурирует понятие энтропии по Шеннону.

Цель статьи состоит в том, чтобы сравнить понятие энтропии Планка и понятие энтропии по Шеннону, а затем, с точки зрения этого сравнения, рассмотреть проблему измерений.

### **Основные идеи М. Планка в области термодинамики**

Пытаясь применить классическую механику для описания тепловых явлений, М. Планк столкнулся с тем, что ее недостаточно для описания ряда процессов. Переосмысляя формулировку Клаузиусом первого начала термодинамики, Планк замечает недостаточность первого закона. Первый закон (сохранение энергии) не может определить, в каком направлении пойдет процесс (например, будет ли тепло перетекать от горячего тела к холодному или наоборот). Он лишь требует равенства энергий.

Но для ряда тепловых процессов характерна необратимость, которую вводит второй закон термодинамики. Он принципиально отличается от первого закона (закона сохранения энергии) тем, что он определяет направление, в котором происходят процессы в природе. Процесс считается необратимым, если невозможно вернуть и систему, и окружающие тела в исходное состояние, не оставив никаких других изменений. Примерами необратимых

процессов являются выделение тепла из-за трения, расширение газа, теплопроводность (передача тепла от горячего тела к холодному). Планк подчеркивает, что безнадежно пытаться вывести второй закон из принципов механики (например, из стремления потенциальной энергии к минимуму), так как в термодинамике не обнаружено состояния с абсолютным минимумом энергии, которое бы однозначно определяло направление всех процессов. Второй закон термодинамики – это не следствие определений или манипуляций с понятием энергии, а самостоятельный закон природы, основанный на экспериментальных фактах и формулируемый в виде проверяемых утверждений о необратимости определенных процессов. «Процесс, который никоим образом не может быть полностью обратим, называется необратимым, все остальные процессы – обратимыми. Для того, чтобы процесс был необратимым, недостаточно того, чтобы его нельзя было непосредственно обратить. Это относится к многим механическим процессам, которые не являются необратимыми (ср. § 113). Полное требование состоит в том, чтобы было невозможно, даже с помощью всех факторов природы, восстановить везде точное начальное состояние, когда процесс когда-то произошел. [5, 83]. Таким образом, второй закон термодинамики – это независимый и фундаментальный принцип, устанавливающий "стрелу времени" для природных процессов через критерий необратимости, который не выводим из одного лишь закона сохранения энергии.

Первый закон термодинамики (закон сохранения энергии) был сформулирован как принцип, доказывающий невозможность вечного двигателя, совершающего полезную работу. Первоначально Планк действует по аналогии, привязывая формулировку второго закона к невозможности вечного двигателя второго рода, основанного на циклической передаче тепла. План указывает на необратимость процесса передачи тепла от более горячего тела к более холодному. «Поток тепла всегда направлен от более горячего к более холодному. Опять же, частным случаем этого типа процесса является прямая передача тепла посредством теплопроводности между тепловыми резервуарами, без какого-либо фактического участия системы, которая должна пройти через цикл операций. Видно, что это необратимое изменение, поскольку оно приводит к увеличению суммы энтропий двух тепловых резервуаров» [5, с. 108].

Затем Планк отходит от чисто термодинамических формулировок к формулировке комбинаторной, когда энтропия начинает пониматься как мера неупорядоченности системы. «Энтропия, подобно температуре, давлению и плотности, не может быть определена как абсолютная, непрерывная величина, а как некое среднее значение большого числа отдельных величин» [5, с. 104]. Классическое статистическое определение энтропии по Больцману-Планку:  $S = k * \ln(W)$ , где  $W$  – число микросостояний системы, а  $k$  – постоянная Больцмана (равна  $1,38 \cdot 10^{-23}$  Дж/К).

Таким образом, второй закон термодинамики, фундаментально независим от эмпирических явлений. Подобно первому закону, который утвердился не просто как обобщение экспериментальных данных о вечном двигателе, а как самостоятельный и непреложный принцип, второй закон – с его принципом возрастания энтропии – также претендует на статус объективного естественного закона, существующего независимо от человеческого разума. Последующее применение принципа энтропии к анализу цикла Карно для систем любого рода является ярким примером такого плодотворного пути, ведущего к углублению понимания природы.

### **Проблема измерений в квантовой механике**

Одной из актуальных проблем квантовой механики является проблема измерений. В последние годы появилась попытка трактовать процесс измерения как нарушение замкнутости квантовой системы [6]. Данная трактовка отходит от традиционного понимания процесса измерения в копенгагенской интерпретации, когда наблюдатель рассматривался как часть системы, а квантовое описание распространялось на всю систему. Согласно новой трактовке, система теряет энергию в процессе измерения параметров частицы, то есть получение информации о системе приводит к нарушению ее замкнутости.

Данная интерпретация процесса измерения связывает квантовую систему, с одной стороны, с термодинамикой, а с другой стороны, с теорией информации.

Как было указано выше, термодинамическое понимание системы было особенно близко основателю квантовой теории М. Планку который занимался проблемами термодинамики. Собственно, само понятие «квант» и постоянная Планка было введено при решении проблемы излучения абсолютно черного тела. Планку принадлежит интерпретация второго начала и формулировка третьего начала термодинамики, а так же статистическая интерпретация понятия энтропии.

Первые идеи о «физичности» информации появляются в мысленном эксперименте, известном как «Демон Максвелла». Демон совершает работу по разделению газа на холодные и горячие молекулы, на основе информации об их скоростях. В истории науки связь между термодинамикой и информацией подробно анализировалась на заре формирования кибернетики.

Изучая процессы передачи информации, Н. Винер отмечал, что передаваемое сообщение часто искажают посторонние помехи, шумовой фон. Поэтому передача информации возможна лишь как передача альтернатив, расшифровка которых возможна лишь с некоторыми погрешностями. Тем самым системы связи следует рассматривать статистически. Поэтому Винеру пришлось «разрабатывать статистическую теорию количества информации. Понятие количества информации совершенно естественно связывается с классическим понятием статистической механики – понятием энтропии. Как количество информации в системе есть мера организованности системы, точно так же энтропия системы есть мера дезорганизованности системы; одно равно другому, взятому с обратным знаком» [7, с. 55].

То есть, информация, понимаемая как выбор одной из многих альтернатив, «отождествляется Винером с отрицательной энтропией и становится, подобно количеству вещества или энергии, одной из фундаментальных характеристик явлений природы. Таков второй краеугольный камень кибернетического здания. Отсюда толкование кибернетики как теории организации, как теории борьбы с мировым хаосом, с роковым возрастанием энтропии. Действующий объект поглощает информацию из внешней среды и использует ее для выбора правильного поведения. Информация никогда не создается, она только передается и принимается, но при этом может утрачиваться,

исчезать. Она искажается помехами, “шумом”, на пути к объекту и внутри его и теряется для него. Борьба с энтропией – борьба с шумом, искажающим информацию» [7, с. 17].

Поскольку квантовое измерение представляет собой получение информации, то есть понижение энтропии измерительного прибора, постольку квантовая система претерпевает обратное воздействие в ходе процесса измерения: рост энтропии и потерю энергии. В связи с повышением энтропии измерение квантовой системы меняет ее состояние необратимым образом. Человек воздействует на свободную эволюцию квантовой системы, **«гася энтропию извлеченной из окружающей среды отрицательной энтропией – информацией»**. Познание – часть жизни, более того – самая ее суть. “Действенно жить – это значит жить, располагая правильной информацией”» [7, с. 41].

Винер писал, что если некоторая функция может принимать «одинаковые значения при разных значениях аргумента, или в случае функции нескольких переменных позволяем некоторым из них свободно пробегать их естественную область изменения, мы теряем информацию. Никакая операция над сообщением не может в среднем увеличить информацию. Здесь мы имеем точное применение второго закона термодинамики к технике связи. Обратно, уточнение в среднем неопределенной ситуации приводит, как мы видели, большей частью к увеличению информации и никогда – к ее потере» [7, с.125].

Заметим лишь, что понятие энтропия у Планка и в теории информации у Шеннона [8] имеет несколько различий. Так область применения понятия энтропии у Планка - физические системы: газы, кристаллы, излучение. У Шеннона применение энтропии происходит к системам связи и зашумленным каналам, искажающим данные пересылаемых сигналов.

Планк описывает, прежде всего, меру необратимости процесса и рассеяния энергии. Позже Планк дает статистическую интерпретацию энтропии, связывая последнюю с числом возможных состояний системы: степень беспорядка, неопределенности в расположении и движении частиц (молекул, атомов). Шеннон характеризует степень неопределенности, неожиданности или информационной насыщенности сообщения, а также меру количества информации. Например, сигнал, состоящий из

одних единиц имеет низкую информативность и высокую упорядоченность, то есть характеризуется низкой энтропией. А вот сигнал, состоящий из случайной последовательности символов, имеет высокую информативность и высокую энтропию. В данном случае энтропия характеризует степень неопределенности в выборе следующего символа. Чем менее предсказуемо следующее сообщение (или символ в нем), тем выше его информационная энтропия.

Планк вывел формулу Больцмана  $S = k \ln(W)$ , где  $S$  – энтропия,  $k$  – постоянная Больцмана,  $W$  – число микросостояний. Здесь энтропия – это логарифм числа микросостояний ( $W$ ), которые соответствуют одному макросостоянию системы. Эта формула связывает энтропия со степенью неупорядоченности данной системы. Чем больше способов (микросостояний) реализовать данное макросостояние, тем выше энтропия. Микросостояние – это точное описание состояния каждой отдельной частицы (ее положение, скорость). Например, точные координаты и скорости всех молекул воздуха в комнате. Предоставленная самой себе система стремится прийти к самому вероятному состоянию – состоянию хаоса и беспорядка. Наибольшую энтропию имеет газ, наименьшую – правильный кристалл.

Формула Шеннона для энтропии:

$$H = - \sum p_i * \log_2(p_i)$$

где  $H$  – энтропия,  $p_i$  – вероятность появления  $i$ -го символа.

Формула Шеннона является обобщением и абстракцией статистического подхода Планка-Больцмана. Если все микросостояния равновероятны, формулы становятся математически эквивалентными (с точностью до константы и основания  $\log$ ). Энтропия – это средняя "степень неожиданности" символа в сообщении. Она измеряет неопределенность того, какое сообщение будет передано. Высокая энтропия по Шеннону означает, что каждый следующий символ сообщения непредсказуем. Сообщение несет много информации. Низкая энтропия означает, что сообщение предсказуемо и содержит мало новой информации. Таким образом, энтропия – это мера потенциальной "информационной емкости" источника, мера неожиданности посылаемых сообщений.

В итоге мы видим, что оба подхода оперируют вероятностями. И Планк, и Шеннон используют логарифмическую меру для описания неопределенности, связанной с распределением вероятностей. Формула Шеннона – это гениальное обобщение статистической физики на абстрактный мир информации. Эта глубокая связь показывает, что понятия порядка, хаоса и информации фундаментально связаны в нашем мире.

Дополнительным мостом между термодинамикой и теорией информации может послужить принцип Ландауэра, согласно которому при любой логически необратимой операции по манипулированию информацией, такой как стирание небольшого количества памяти, энтропия увеличивается, и соответствующее количество энергии рассеивается в виде тепла. Поскольку любая вычислительная операция сопровождается записью данных в память и последующим стиранием их оттуда, необходимо затрачивать энергию на проведение данной операции. Рассеивание энергии и рост энтропии указывают на термодинамическую необратимость данной операции с информацией [9].

### Заключение

Проблема измерений в квантовой механике – одна из ключевых и наиболее дискуссионных тем современной физики. Многолетние дискуссии по поводу фундаментальных физических принципов (прежде всего – принципа неопределённости Гейзенберга) и философских интерпретаций природы квантовой неопределённости (онтологической и эпистемологической) не достаточно для удовлетворительного решения. Анализ показал, что для её полного осмысления необходимо применять междисциплинарные подходы, связывающие квантовую механику с термодинамикой и теорией информации.

В итоге была выявлена глубинная связь между термодинамическим подходом М. Планка и информационным подходом К. Шеннона, оба из которых оперируют понятием энтропии как меры неопределённости, но в разных контекстах (характеристика необратимости процессов физической системы, или непредсказуемость источника сообщений). Математические формулировки

( $S=k*\ln(W)$  у Планка и  $H=-\sum p_i*\log_2(p_i)$  у Шеннона) имеют общую статистическую природу, различаясь лишь интерпретацией вероятностей и областью применения. Их математическое родство намекает на глубинную связь между бытием и знанием: процесс измерения в квантовой механике оказывается не просто «считыванием» данных, а созданием информации через взаимодействие наблюдателя и системы.

Сопоставление энтропии Планка и Шеннона открывает неожиданную перспективу в области метафизические следствий. Во первых, реальность предстает как процесс, где акт наблюдения – не пассивное созерцание, а активное участие в формировании действительности. Поднимается вопрос границы объективности. Идея «объективной реальности, независимой от наблюдателя» оказывается иллюзией. Мы не можем отделить познающего от познаваемого – они взаимодействуют и в ходе их взаимодействия возникает не только информация о реальности, но и ее новые состояния.

Обоснована применимость информационного подхода к проблеме измерений: сопоставление термодинамической и информационной энтропии позволяет трактовать процесс измерения как обмен информацией между системой и наблюдателем, что открывает новые пути для понимания квантовой неопределённости. По крайней мере этот подход позволяет преодолеть подход Н.Бора, рассматривавшего наблюдателя в качестве части квантовой системы, что приводило к появлению субъективистских трактовок и введению неопределенного состояния наблюдателя.

Проведённый анализ подтверждает, что проблема измерений в квантовой механике не сводится к чисто физическому вопросу, а затрагивает фундаментальные аспекты познания: природу случайности и детерминизма; роль наблюдателя в формировании и влиянии на реальность; границы применимости классических понятий в микромире.

Междисциплинарный синтез идей термодинамики, квантовой механики и теории информации демонстрирует, что квантовая неопределённость имеет не только операциональный, но и информационный смысл. При этом процесс измерения можно рассматривать как снятие неопределённости (в шенноновском смысле) через взаимодействие системы с прибором.

Дальнейшие исследования в этом направлении могут привести к:

- созданию более целостной интерпретации квантовой механики;
- разработке новых экспериментальных методик, учитывающих информационный аспект измерений;
- углублению понимания связи между физическими процессами и обработкой информации на фундаментальном уровне.

Таким образом, проблема измерений остаётся живой исследовательской зоной, где физика, философия и теория информации взаимно обогащают друг друга, открывая новые горизонты познания природы реальности.

### Литература

1. Вихман Э. Квантовая физика. Берклевский курс физики. т. IV. – М.: Наука, 1977.
2. *Einstein, A.; Podolsky, B.; Rosen, N.* (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? // *Physical Review*. Vol. 47, №10. P. 777–780.
3. *Bell, J. S.* On the Problem of Hidden Variables in Quantum Mechanics. // *Reviews of Modern Physics*. 1966. Vol. 38, № 3. P. 447–452.
4. *Clauser, J. F.* Experimental Investigation of a Polarization Correlation Anomaly. *Phys. Rev. Lett.* 36, 1223 – Published 24 May, 1976 DOI: <https://doi.org/10.1103/PhysRevLett.36.1223>.
5. *Planck, M.* Treatise on thermodynamics, 1858-1947; Topics Thermodynamics. - London: Longmans, Green Collection cdl; americana Contributor University of California Libraries. 1903.
6. *Желтиков А.М.* Критика квантового разума: измерение, сознание, отложенный выбор и утраченная когерентность // УФН. 2018. Т. 188. № 10. С. 1119-1128.
7. *Винер Н.* Кибернетика, или управление и связь в животном и машине. – 2-е издание. – М.: Наука; Главная редакция изданий для зарубежных стран, 1983.
8. *Shannon, C. E.* A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948. Vol. 27, № 3. P. 4190487/
9. *Landauer, R.* (1961). Irreversibility and Heat Generation in the Computing Process. // *IBM Journal of Research and Development*. №5. P. 183–191.

### References

1. *Wichmann, E. H.* (1971) Quantum physics (Berkeley physics course) Vol. 4. Hardcover.
2. *Einstein, A.; Podolsky, B.; Rosen, N.* (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? // *Physical Review*. Vol. 47, №10. P. 777–780.
3. *Bell, J. S.* On the Problem of Hidden Variables in Quantum Mechanics. // *Reviews of Modern Physics*. 1966. Vol. 38, № 3. P. 447–452.
4. *Clauser, J. F.* Experimental Investigation of a Polarization Correlation Anomaly. *Phys. Rev. Lett.* 36, 1223 – Published 24 May, 1976. DOI: <https://doi.org/10.1103/PhysRevLett.36.1223>.
5. *Planck, M.* Treatise on thermodynamics, 1858-1947; Topics Thermodynamics. - London: Longmans, Green Collection cdl; americana Contributor University of California Libraries. 1903.
6. *Zheltikov, A. M.* Criticism of Quantum Mind: Measurement, Consciousness, Delayed Choice, and Lost Coherence. // *Physics-Uspekhi*, 2018, vol. 188, no. 10, P. 1119-1128. (In Russ)
7. *Wiener, N.* (1948). *Cybernetics: Or control and communication in the animal and the machine*. Cambridge, Massachusetts: MIT Press.
8. *Shannon, C. E.* A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948. Vol. 27, № 3. P. 4190487/
9. *Landauer, R.* (1961). Irreversibility and Heat Generation in the Computing Process. // *IBM Journal of Research and Development*. № 5. P. 183–191.

### Информация об авторе

*Сторожук Анна Юрьевна* – доктор философских наук. ИФПР СО РАН, Россия 630090, г. Новосибирск, ул. Николаева, 8.  
stor71@mail.ru

### Information about Author

*Storozhuk Anna Yurievna* – Doctor of Philosophy. IFPR SO RAN, Russia 630090, Novosibirsk, Nikolaeva str., 8.  
stor71@mail.ru

Дата поступления 12.11.2025  
Принята к печати 11.12.2025

УДК: 168.5+141.32

DOI: 10.15372/PS20250514

EDN: ALEQZF

**О.А. Власова**

## **ДИАЛОГ НАУК В ЭНАКТИВИСТСКОЙ ПСИХИАТРИИ: НЕЙРОФЕНОМЕНОЛОГИЯ Т. ФУКСА И ВОПРОС О МЕЖДИСЦИПЛИНАРНОСТИ<sup>1</sup>**

В статье анализируется проблема междисциплинарности на примере нейрофеноменологии Т. Фукса. Воссоздается история диалога философии и психиатрии как история антиномизма научных подходов. Вскрываются основания энактивизма как основания преодоления антиномизма в философии психиатрии, выделяются две волны энактивизма в психиатрии. Уточняются энактивистские установки современной нейрофеноменологии. Показывается значение нейрофеноменологии в дискуссиях о «трудной проблеме сознания». Последовательно анализируется проект нейрофеноменологии Томаса Фукса, его критические установки, в качестве которых называются: критика нейробиологизма, редукционизма и репрезентационизма. Воссоздается структура нейрофеноменологии и ставится центральная для нее проблема воплощенности. Обсуждаются основные патологические трансформации сознания как поля интенциональности: изменения телесности, темпоральности и социальных отношений. Показывается, что особенно продуктивным является учение Фукса о реляционной психиатрии и полидисциплинарности. Проясняются особенности взаимодействия нейробиологии и философии, психотерапии в их анализе и последующей коррекции психических состояний. Подчеркивается продуктивность модели Фукса для интеграции специалистов помогающих профессий и возможности ее развития в контексте отечественной традиции нейропсихологии и психотерапии.

---

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда № 24-28-01172 «Антиномизм и методологические основания философии психиатрии».

*Ключевые слова:* междисциплинарность, нейрофеноменология, трудная проблема сознания, энактивизм, нейробиология, интенциональность, опыт, воплощенность.

**O.A. Vlasova**

## **DIALOGUE OF SCIENCES IN ENACTIVIST PSYCHIATRY: T. FUCHS' NEUROPHENOMENOLOGY AND THE QUESTION OF INTERDISCIPLINARITY**

The paper analyzes the problem of interdisciplinarity on the example of T. Fuchs' neurophenomenology. The author reconstructs the history of the dialogue the philosophy and psychiatry as the history of antinomianism of scientific approaches, reveals the foundations of enactivism as the basis for the cessation of antinomianism in the philosophy of psychiatry, discovers two waves of enactivism in psychiatry. Enactivist attitudes of modern neurophenomenology are clarified. The significance of neurophenomenology for discussions on the "hard problem of consciousness" is shown. The paper includes the consistent analysis of the project of Thomas Fuchs' neurophenomenology and its serious attitudes: criticism of neurobiologism, reductionism and representationism. The structure of neurophenomenology is reconstructed and the central problem of embodiment is posed. The paper discusses the main pathological transformations of consciousness as a field of intentionality: changes in corporeality, temporality and social relations. It is shown that Fuchs' ideas on relational psychiatry and polydisciplinarity is especially productive. The author studies the interaction of neurobiology and philosophy, psychotherapy in their analysis and the correction of mental illnesses. The productivity of Fuchs' model for international specialists and the possibilities of its development in the Russian tradition of neuropsychology and psychotherapy are emphasized.

*Keywords:* interdisciplinarity, neurophenomenology, hard problem of consciousness, enactivism, neurobiology, intentionality, experience, embodiment.

Поле психиатрии уже около тридцати лет являет собой одно из передовых пространств диалога наук, в которое вовлечены психологи и нейрофизиологи, специалисты в социальной работе и праве, антропологи и критики культуры, биологи и философы. Так было и ранее, в течение XX века, однако начиная с 1990-х гг.,

благодаря открытиям нейронаук, развитию парадигмы энактивизма и философии психиатрии диалог усложнился и обогатился. И теперь он ведется не на периферии ее поля, а в сердцевине теории и практики, вовлекая крупнейшие институты, финансовые программы и ведущих специалистов.

Одну из значимых ролей в этом диалоге играет философия. Она, как и прежде, снабжает психиатрию методами, понятиями и первичными гипотезами, посредством которых эта наука концептуализирует собственные проблемы, и непосредственно формирует поле диалога. В.В. Васильев определяет такую функцию как «роль служанки наук», очень метко обозначая ее следующим образом: «...Философы должны налаживать диалог между разными областями науки, имеющими выход на сознание, – между нейробиологами и когнитивистами, эволюционистами и психологами, между лингвистами и квантовыми физиками – сочетания могут быть самыми необычными. Кроме того, они могут снабжать ученых разного рода методологическими принципами, задавать хорошие вопросы (ответ на которые, возможно, не их дело) и прояснять понятия, использующиеся при изучении сознания» [3, с. 32].

Несмотря на то, что для самой философии такая функция посредничества не самоцель, и в последние годы (главным образом, в англо-американском мире) она претендует не на «прислуживание», а на «поиск ответов» в этом поле, преуменьшать ее важность в контексте психиатрии не следует. Тот поворот к опыту, которому в последние тридцать лет, содействует философия, способствует изменению теории, практики лечения, и, что важнее всего, оценке характера психических недугов и изменению отношения к ним. Энактивистская парадигма закладывает основания для принципиально другой конфигурации психиатрической помощи и взаимодействия специалистов помогающих профессий: психологов, психотерапевтов, эрготерапевтов, социальных работников.

### **Диалог в психиатрии, энактивизм и «трудная проблема»**

Если посмотреть в прошлое, отчетливо видно, что исток диалога – антиномизм научных подходов, покоящийся на психофизическом дуализме тела/сознания. Его акцентирует в своих ранних

работах уже Карл Ясперс, который настаивает на том, что психиатрия имеет двойственный статус и одновременно представляет ее как науку о болезни (медицинскую) и о человеке (антропологическую). Томас Фукс следует этой традиции в трактовке идентичности психиатрии. «С момента своего зарождения в 1800 году психиатрия разрывалась между полюсами естественных и гуманитарных наук, с одной стороны, устремляя свое внимание на субъективный опыт, а с другой, на нейронный субстрат» [14, р. 252], – подчеркивает он.

С 1910-х гг. до 1990-х гг. две научные перспективы не могут сойтись в единство. Антиномизм сохраняется в ситуации развития когнитивной науки и на волне философии психиатрии в 1990-х гг. За это время психиатрия переживает две так называемые «биологические волны», связанные с открытиями в исследованиях мозга и акцентированием биологической парадигмы: в 1900–1920-х гг. и в 1990-е гг. За обеими волнами идет «реакция»: вслед за первой развивается экзистенциально-феноменологическая психиатрия, вслед за второй – нейрофеноменология.

Начиная с 1930-х гг., в рамках развития экзистенциально-феноменологической психиатрии, идет стихийное движение к теории воплощенного переживания. В центральной для направления категории опыта это движение предвосхищает многие установки энактивизма которые, вслед за Е.Н. Князевой [7, с. 7] (и продолжая сделанные ею обобщения применительно к конкретному материалу), можно обозначить как: акцентирование перспективы первого лица и описательной методологии; попытку снятия оппозиции биологического и личностного, субъекта и объекта в категории опыта и признание его множественной разноразмерной детерминации; обращение к непосредственному переживанию и телесности, метаперспективе коммуникации [4]. Однако экзистенциально-феноменологическая психиатрия становится слишком преждевременным предшественником энактивизма.

Последний приносит философии психиатрии шанс снятия антиномизма. В знаковой книге «Отелесненный ум» Франциско Варела, Эван Томпсон и Элеонора Рош прямо заявляют: «Наша цель – создать пространство возможностей, в котором в полной мере можно будет оценить связь когнитивной науки и человеческого опыта, а также стимулировать способствующие трансформации

свойства человеческого опыта в научной литературе» [2, с. 89]. В работах 1990-х гг. (Й. Парнас, Дж. Стангеллини, Л. Сасс, Д. Захави, К. Мундт и др.) идеи экзистенциально-феноменологической психиатрии дополняются идеями социальной феноменологии (Р. Лэйнг и др.), которые ранее расценивались, скорее, в рамках антипсихиатрии и критики институции. Это позволяет выйти в поле интерсубъективности, и эти два аспекта в единстве – феноменология субъективного и межсубъективного опыта – сразу же включаются в практику, примером чего являются разработанные под руководством Й. Парнаса опросники для исследования переживания себя (EASE) и мира (EAWM) [20; 21]. Однако в это время происходит, скорее, антиномичное уравнивание нейробиологической психиатрии.

Расширение сферы опыта поворачивает психиатрию к поиску новых оснований, и в качестве таковых она обращается не только к работам Ф. Варелы, но и к экологическому подходу Грегори Бэйтсона [1], а также Дж. Гибсона [5]. Именно Бэйтсон когда-то, еще в начале 1960-х гг., вдохновляет Лэйнга на развитие межличностной феноменологии, а Гибсон разрабатывает терминологию «энактивности» и «отелеснивания». В.В. Лекторский характеризует специфику нового поворота: «Гибсон, так же как и Варела с коллегами, считал, что перцепция должна быть понята не как результат простой обработки мозгом информации, получаемой из внешней среды, а как процесс активного извлечения этой информации, в котором участвуют не только органы чувств и мозг, а все тела познающего субъекта» [9, с. 14]. В этом поле в 2000-х гг. оформляется энактивистская психиатрия Томаса Фукса, одна из книг которого носит название «Экология мозга» [14].

Энактивистская психиатрия логически следует за самим энактивизмом в решении «трудной проблемы сознания», которую В.В. Васильев кратко описывает так: «...как ментальные состояния, в их качественном аспекте, соотносятся с соответствующими им нейронными процессами и почему они вообще существуют в привязке к этим процессам?» [3, с. 33]. Для психиатрии вопрос звучит аналогично: как интенсивно переживаемый опыт пациентов соотносится с открываемыми нейронными механизмами психических заболеваний и каков характер этой связи. Однако психиатрия «материализм» классической философии сознания [12]

и добавляет очень важный аспект: можем ли мы, воздействуя на опыт, трансформировать нейробиологические механизмы и, таким образом, встроить «психическое лечение» в целостный комплекс нейрофеноменологической психиатрии? Это не просто логическое развитие вопроса на практический план психиатрии. Такая формулировка позволяет изменить сам подход к лечению психических заболеваний: «зайти» не через поведение, а через опыт, ведь в малой психиатрии мы снова видим доминирование бихевиоризма.

### **Томас Фукс: философская концептуализация в психиатрии**

В нейрофеноменологии Томаса Фукса, профессора и клинициста Гейдельбергского университета, проступает первое последовательное осознание происходящих в психиатрии изменений. В основе его теории лежит критика: 1) нейробиологической установки, 2) редуccionизма, 2) репрезентационизма.

По убеждению Фукса, нейробиология осуществляет по отношению к мозгу программу последовательной натурализации: она показывает, что наш опыт и нашу субъективность можно описать в нейробиологических терминах, демонстрирует на фМРТ визуализации психических процессов и переживаний, язык нашего жизненного мира в объективированных научных терминах. Сам он называет это «ниспровержением жизненного мира» [14, р. xiv]. Основным принципом натурализации при этом является редуccionизм: все сложные явления опыта, общения, поведения, раскладываются на составляющие или сводятся к элементам, которые можно локализовать и привязать к простейшим нейробиологическим механизмам. На основании неких ментальных репрезентаций, как считается, мы реагируем на окружающий мир, людей и выстраиваем свое поведение.

Фукс критикует эту позицию, и, по его глубочайшему убеждению, мы должны понять, что репрезентации, схемы, не опосредуют нашу реальность и не являются ею, мы взаимодействуем с окружающим миром, включая других людей, непосредственно. В своих многочисленных работах он подчеркивает «...мозг есть,

в первую очередь, орган живого существа, а не разума» [14, р. 5]. В этой перспективе мозг трактуется как орган-посредник в сенсомоторном взаимодействии с миром и другими людьми, орган, который открывает мир. Функция посредничества возможна благодаря фундаментальному качеству человека – воплощенности, отелесненному характеру ума.

В ряде своих работ Фукс развивает феноменологическую теорию психопатологии, связывая ее с новейшими экспериментальными нейропсихологическими исследованиями и одновременно концептуализируя эти нарушения на философском (феноменологическом) языке. По убеждению Фукса, необходимо обратиться к тому чувству общности, которое предшествует разделению проприоцепции, восприятия и эмоций, и которое в своей «Феноменологии восприятия» Мерло-Понти называет интенциональной дугой [18]. Именно трансформации этого чувства приводят к трансформациям 1) воплощенности, 2) темпоральности и 3) метаперспективы.

Воплощенность – исходная сфера, и, следуя М. Мерло-Понти и Р. Лэйngu, Фукс подчеркивает, что развоплощение приводит к разрушению ощущения себя, отождествления с телом и органичной включенности в межтелесное взаимодействие [15]. В группе к этому измерению добавляется еще измерение метаперспективы, т.е. способность ставить себя на место другого и представлять себе мысли и чувства других людей. Фукс уточняет содержание метаперспективы применительно к психопатологии: в норме она связана с постоянным перемещением из воплощенной перспективы собственного тела в децентрированную перспективу другого, а также со способностью приостанавливать переживание непосредственности и достоверности опыта. В психических расстройствах обращение перспектив затруднено, сознание разворачивается в конкретности и не может обратиться к воображаемому плану. Мерло-Понти в «Феноменологии восприятия» приводит ряд подобных клинических случаев при неврологической патологии [11, с. 137–197], и Фукс подтверждает эти данные ссылкой на исследования, в которых фиксируется изменения фМРТ-активности при переходе от перспективы первого лица к перспективе второго или третьего (изменения мозговой активности с правой височно-теменной коры на левую височно-полярную при постоянной активации передней поясной извилины) [22].

Оба указанных нарушения связывает, по Фуксу, трансформация интенциональности, которая становится центральной в нарушении темпоральности. Последние есть следствие обращения интенции: направленность событий поворачивается в противоположную сторону, прямо на больного. Человек больше не реализует собственный план, активными силами становятся другие люди и вещи, мир в целом. В качестве нейрофизиологических коррелятов Фукс называет изменения в префронтальной коре, которые влекут снижение концентрации внимания, изменение объема рабочей памяти, что в совокупности приводит к распаду единства сознания [19].

Однако Фукс стремится уйти от традиционного представления о психопатологии как искаженного взгляда на реальность и разрабатывает гипотезу интересубъективной конституции общей реальности, выделяя два уровня последней. На первом уровне воплощенное существование дает гибкость и множественность перспектив, субъективный релятивизм, позволяющий переключаться между событиями жизни и объектами. На втором уровне взаимодействие с другими приводит к необходимости отказа от перспективы первого лица и замене ее на перспективу третьего. Лабильность первого и третьего лица уходит в психопатологии, где опыт эгоцентричности и субъективация восприятия реальности и других замыкают человека в себе, в неспособности со-конституирования реальности с другими. Существование утрачивает «свою интенциональную и децентрированную структуру» [13, р. 128]. В психопатологии невозможным становится взаимное преобразование «я», «ты», «они» перспектив, спонтанная и свободная субъективностей, если использовать терминологию, которую, вслед за Ж.-П. Сатром, предложил Р. Лэйнг [10].

Проанализированная многоуровневая теория опыта и его патологии показывает два обозначенных Фуksom плана: молекулярной дисфункции и молярного описания, которые сходятся в его описаниях патологического опыта. Между ними воссоздается жизнь опыта. Промежуточный статус феноменологии напоминает тот, что был у экзистенциально-феноменологической психиатрии. Однако там основным недостатком как раз было исключение естественнонаучного пласта. Нейрофеноменология снима-

ет это противоречие и, благодаря обращению к парадигме сознания, позволяет выстроить полноценную меж- и даже поли- дисциплинарную теорию душевной жизни в норме и патологии.

### Проблема междисциплинарности

Развивая идеи Варелы о воплощенной когнитивной науке и значимости исследования переживания вне разделения на мозг и тело, Фукс указывает на необходимость для психиатрии представления о целостности опыта. «...Адекватная наука об опыте, – подчеркивает он, – имеет фундаментальное значение как для психиатрии, так и для когнитивной нейробиологии. Без исследования феноменологии субъективности мы не можем прояснить соответствующие надличностные процессы. Когнитивная нейробиология будет оставаться слепой к своему предмету до тех пор, пока методологически не осмыслит то, что пытается объяснить. Если мы не преодолеем нынешнюю объективистскую, редуccionистскую эпистемологию в психиатрии, мы не увидим прогресса эмпирических исследований» [17, р. 321].

Акцентируя концепцию «мозга-посредника» и сводя нейробиологию и философию (в лице феноменологии) в едином поле, Фукс выделяет три уровня развертывания воплощенного опыта человека: 1) микроуровень (уровень мозга) – нейронные и молекулярные процессы, 2) мезоуровень (уровень организма) – взаимодействие мозга, организма и окружающей среды, 3) макроуровень (психосоциальный) – социальное взаимодействие. Связи между этими уровнями идут вверх и вниз, что позволяет говорить о круговой, петлеобразной детерминации опыта человека. Это характерная для энантикизма трактовка, Д.В. Иванов так описывает ее специфику у Варелы, Томпсона и Рош: «Эти авторы подчеркивают, что восприятие и действия представляют собой петлю. Содержание восприятия существенными образом зависит от характера деятельности живого организма, в свою очередь, те действия, которые совершает существо, зависят от того, как оно воспринимает мир. И, конечно же, то, как организм воспринимает окружающую

среду и взаимодействует с ней, определяется его конкретной телесной организацией» [6, с. 96].

В динамической системе Фукс выделяет горизонтальные и вертикальные круговые (петлеобразные) связи: вертикальные (двунаправленные) связывают уровни разного порядка; горизонтальные представляют циклы отрицательной обратной связи и сопряжены с невозможностью изменения перспективы и взаимодействия внутри уровня. «Оба вида циркулярных причинных процессов связаны с посредничеством мозга, но не могут быть локализованы внутри него. (...) То, что все мышление реализуется как нейронная активность, еще не означает, что оно идентично мозговым процессам» [14, р. 256, 258], – подчеркивает он. Мозг опосредует процессы означивания, смыслополагания, интенциональные акты в контексте социальной ситуации в смене социальных перспектив и в контексте функционирования тела в окружающей среде. Динамическую систему отношений может, по Фуксу, исследовать только приблизительно обрисовываемая им «релятивная медицина» как «наука и практика, изучающая биологические, психологические и социальные отношения и их нарушения» [14, р. 255].

Фукс следующим образом описывает взаимодействие специалистов в этой области: «Феноменология трактует субъективность не как объект, который необходимо описать, а как поле, в котором являет себя мир. Поэтому ее целью является постижение не содержания или объекта, но *формы* и *структуры* опыта сознания. Вероятно, измененные формы опыта патогенетически подводят нас ближе к биологическому субстрату, в то время как содержание лично окрашено и носит более индивидуальный характер. (...) Таким образом, феноменологический подход конституирует промежуточный уровень, который связывает уровень молекулярных дисфункций, изучаемых экспериментальной нейропсихологией (к примеру, снижение непосредственной памяти, функций контроля, внимания), и молярный уровень описательной психопатологии и ее нозологических синдромов. Микродисфункции влияют на целостную картину трансформированного опыта» [16, р. 548].

Самым важным следствием междисциплинарной контекстуализации психических заболеваний является пересмотр су-

ти терапии. Фукс следующим образом описывает ее: «Если мы рассматриваем психические расстройства в двунаправленной трактовке человека как одновременно переживающей целостности и физического организма, то все терапевтические воздействия должны быть представлены в *обоих аспектах*. (...) Это значит, что всякое терапевтическое воздействие носит одновременно физиологический, *а также* психологический характер» [14, р. 268]. И фармакологическая терапия, и психотерапия работают в пространстве воплощенного функционирования человека. Психотерапия действует через трансформации «сверху вниз», по нисходящему механизму (связанному с префронтальной корой), от субъективного опыта к вегетативному уровню. Медикаментозная терапия действует через трансформацию «снизу вверх», по восходящему механизму (связанному с подкорковыми слоями: вентральной областью и лимбической системой), от фармакологической терапии через изменение подкоркового метаболизма к трансформации организации познания и взаимодействия с миром. Медикаментозная терапия может только купировать симптомы и воздействовать на ранее образованные нейронные связи, она напрямую не влияет на сложившиеся патологические модели взаимодействия человека. Ни один из путей не является самодостаточным для лечения, напротив, один предполагает другой, поскольку именно их сочетание может целостно трансформировать петлеобразные связи и изменить все три уровня воплощенного функционирования.

Полиперспективистский подход, который делает возможным опора на энгаживизм, начинает развиваться только в последние годы. Он подпитывается необходимостью преодоления бихевиоризма как устаревшего основания реабилитационных программ. По мысли сторонников-психиатров, он может выступить объединяющим стержнем для институциональной интеграции разнородных специальностей: психотерапии и нейропсихологии, социальной работы и эрготерапии. Сейчас такой базой интеграции выступает только АВА-подход.

Для нашей российской сцены это такая же возможность налаживания полидисциплинарных связей, которые не до конца выстроены в рамках бихевиористской парадигмы, и которые ча-

сто исключают психотерапевтов, нейропсихологов, дефектологов. Энактивизм чрезвычайно близок отечественной психологии с ее культурно-исторической психологией (Л.С. Выготский), теорией динамических систем (А.Р. Лурия), психологией личности (А.Г. Асмолов, Ф.Е. Василюк) смысло-деятельностной теорией (А.Н. и Д.А. Леонтьевы) [8]. Он может укрепить развивающийся (в рамках когнитивных наук и философии сознания) диалог нейрофизиологии, психологии и философии.

### Литература

1. *Бейтсон Г.* Шаги в направлении экологии разума. Пер. Д. Федотова, М. Папуша. М.: УРСС, 2005. 232 с.
2. *Варела, Ф., Томпсон, Э., Рош, Э.* Отелесненный ум. Когнитивная наука и человеческий опыт. Пер. К. Тулуповой. М.: Фонд «Сохраним Тибет», 2023. 456 с.
3. *Васильев В.В.* Трудная проблема сознания. М.: Прогресс-Традиция, 2009. 269 с.
4. *Власова О.А.* Старые проблемы и новые перспективы нейрофеноменологии в психиатрии: хроника радикального поворота // Вестник Российского университета дружбы народов. Серия: Философия. 2024. №4. С. 964–978.
5. *Гибсон Дж.* Экологический подход к зрительному восприятию. Пер. Т.М. Сокольской. М.: Прогресс, 1988. 464 с.
6. *Иванов Д.В.* Энактивизм и проблема сознания // Эпистемология и философия науки. 2016. Т. 49. № 3. С. 88–104.
7. *Князева Е.Н.* Энактивизм: новая форма конструктивизма в биологии. М.; СПб.: Центр гуманитарных инициатив, Университетская книга, 2014. 352 с.
8. *Лекторский В.А.* Познание, действие, реальность // Вопросы философии. 2017. № 9. С. 5–23.
9. *Лекторский В.В.* Предисловие к первому русскому изданию / Варела, Ф., Томпсон, Э., Рош, Э. Отелесненный ум. Когнитивная наука и человеческий опыт, пер. К. Тулуповой. М.: Фонд «Сохраним Тибет», 2023. С. 9–16.
10. *Лэйнг Р.Д.* Я и Другие. Пер. Е. Загородной. М.: Класс, 2002. 192 с.
11. *Мерло-Понти М.* Феноменология восприятия. Пер. с франц. под ред. И.С. Вдовиной, С.Л. Фокина. СПб.: Ювента, Наука, 1999.

12. Юлина Н.С. Головоломки проблемы сознания: концепция Дэниела Деннета. М.: Канон+, 2004. 544 с.

13. Fuchs T. Delusion, Reality and Intersubjectivity: A Phenomenological and Enactive Analysis // *Phenomenology and Mind*. 2020. Vol. 18. P. 120–143.

14. Fuchs T. Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind. Oxford: Oxford University Press, 2017. 370 p.

15. Fuchs T. Pathologies of Intersubjectivity in Autism and Schizophrenia // *Journal of Consciousness Studies*. 2015. № 1–2. P. 191–214.

16. Fuchs T. Phenomenology and Psychopathology / Schmicking D., Gallagher S. (eds.) *Handbook of Phenomenology and Cognitive Science*. Dordrecht, Ney York, Heidelberg, London: Springer, 2010. P. 547–573.

17. Fuchs T. The Challenge of Neuroscience: Psychiatry and Phenomenology Today // *Psychopathology*. 2002. Vol. 35, № 6. P. 319–326.

18. Fuchs T. The Tacit Dimension. Commentary to Wolfgang Blankenburgs “First Steps Toward a Psychopathology of Common Sense” // *Philosophy, Psychiatry & Psychology*. 2001. № 4. P. 323–326.

19. Fuchs T. The Temporal Structure of Intentionality and Its Disturbance in Schizophrenia // *Psychopathology*. 2007. № 40. P. 229–235.

20. Parnas J., Møller P., Kircher T., Thalbitzer J., Jansson L., Handest P., Zahavi D. EASE: Examination of Anomalous Self-Experience // *Psychopathology*. 2005. Vol. 38. P. 236–258.

21. Sass L., Pienkos E., Skodlar B., Stanghellini G., Fuchs Th., Parnas J., Jones N. EAWE: Examination of Anomalous World Experience // *Psychopathology*. 2017. Vol. 50. P. 10–54.

22. Vogeley, K, Bussfeld, P, Newen, A, et al. Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective // *Neuroimage*. 2001. Vol. 14. P. 170–181.

## References

1. Bateson G. Steps to An Ecology of Mind. Moscow.: URSS, 2005. (Russian Trans.)

2. Fuchs T. Delusion, Reality and Intersubjectivity: A Phenomenological and Enactive Analysis // *Phenomenology and Mind*. 2020. Vol. 18. P. 120–143.

3. Fuchs T. Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind. Oxford: Oxford University Press, 2017. 370 p.

4. Fuchs T. Pathologies of Intersubjectivity in Autism and Schizophrenia // *Journal of Consciousness Studies*. 2015. № 1–2. P. 191–214.

5. *Fuchs T.* Phenomenology and Psychopathology / Schmicking D., Gallagher S. (eds.) Handbook of Phenomenology and Cognitive Science. Dordrecht, Ney York, Heidelberg, London: Springer, 2010. P. 547–573.

6. *Fuchs T.* The Challenge of Neuroscience: Psychiatry and Phenomenology Today // Psychopathology. 2002. Vol. 35, № 6. P. 319–326.

7. *Fuchs T.* The Tacit Dimension. Commentary to Wolfgang Blankenburgs “First Steps Toward a Psychopathology of Common Sense” // Philosophy, Psychiatry & Psychology. 2001. № 4. P. 323–326.

8. *Fuchs T.* The Temporal Structure of Intentionality and Its Disturbance in Schizophrenia // Psychopathology. 2007. № 40. P. 229–235.

9. *Gibson J.J.* The Ecological Approach to Visual Perception. Moscow.: Progress Publ., 1988. (Russian Trans.)

10. *Ivanov D.V.* Enaktivizm i problema soznaniya [Enactivism and the Problem of Consciousness] // Epistemologiya i filosofiya nauki [Epistemology & Philosophy of Science] . 2016. № 3. P. 88–104. (In Russ.)

11. *Knyazeva E.N.* Enaktivizm: novaya forma konstruktivizma v biologii [Enactivism: A New Form of Constructivism in Biology]. Moscow, St Petersburg: University Book Publ., 2014. (In Russ.)

12. *Laing R.D.* Self and Others. Moscow: Class Publ., 2002. (Russian Trans.)

13. *Lectorskiy V.V.* Poznanie, deystvie, real'nost' [Cognition, Action, Reality] // Voprosy filosofii. [Questions of Philosophy] 2017. № 9. P. 5–23.

14. *Lectorskiy V.V.* Predislovie k pervomu russkomu izdaniyu [Preface to First Russian Edition] / Varela F., Thompson E., Rosch E. The Embodied Mind: Cognitive Science and Human Experience. Moscow: Save Tibet Foundation Publ., 2023. P. 9–16.

15. *Merleau-Ponty M.* Phenomenology of Perception. St. Petersburg: Yuventa, Nauka Publ., 1999. (Russian Trans.)

16. *Parnas J., Möller P., Kircher T., Thalbitzer J., Jansson L., Handest P., Zahavi D.* EASE: Examination of Anomalous Self-Experience // Psychopathology. 2005. Vol. 38. P. 236–258.

17. *Sass L., Pienkos E., Skodlar B., Stanghellini G., Fuchs Th., Parnas J., Jones N.* EAW: Examination of Anomalous World Experience // Psychopathology. 2017. Vol. 50. P. 10–54.

18. *Varela F., Thompson E., Rosch E.* The Embodied Mind: Cognitive Science and Human Experience. Moscow: Save Tibet Foundation Publ., 2023. (Russian Trans.)

19. *Vasilyev V.V.* Trudnaya problema soznaniya [The Hard Problem of Consciousness]. Moscow: Progress Tradition, 2009. (In Russ.)

20. *Vlasova O.A.* Starye problemy i novye perspektivy neyrofenomenologii v psikiatrii: khronika radikal'nogo povorota [Old Problems and New Perspectives for Neurophenomenology in Psychiatry: The Chronicle

of the Radical Turn] // Vestnik Rossiyskogo universiteta druzhby narodov. Seriya: Filosofiya [RUDN Journal of Philosophy]. 2024. № 4. P. 964–978. (In Russ.)

21. *Vogeley, K, Bussfeld, P, Newen, A, et al.* Mind reading: Neural Mechanisms of Theory of Mind and Self-Perspective // Neuroimage. 2001. Vol. 14. P. 170–181.

22. *Yulina N.S.* Golovolomki problemy soznaniya: kontseptsiya Deniela Denneta [Puzzles of the Problem of Consciousness: Daniel Dennett's Concept]. Moscow.: Kanon+ Publ., 2004. (In Russ.)

### **Информация об авторе**

Власова Ольга Александровна - Санкт-Петербургский государственный университет, кафедра истории философии, Институт философии, профессор, ведущий научный сотрудник, доктор философских наук, доцент, Университетская набережная, д. 7–9, Санкт-Петербург, 199034, Российская Федерация.

[o.a.vlasova@gmail.com](mailto:o.a.vlasova@gmail.com)

### **Information about the author**

Vlasova Olga A. - Saint-Petersburg State University, Department of History of Philosophy, Institute of Philosophy, Professor, leading researcher, DSc in Philosophy 7–9, Universitetskaya emb, St. Petersburg, 199034, Russian Federation.

[o.a.vlasova@gmail.com](mailto:o.a.vlasova@gmail.com)

Дата поступления 01.03.2025

Принята к печати 11.12.2025

УДК: 1/14

DOI: 10.15372/PS20250515

EDN: IWOUOG

**А.К. Гуц**

**МОДАЛЬНЫЙ РЕАЛИЗМ,  
СПОСОБЫ ПОРОЖДЕНИЯ ВОЗМОЖНЫХ МИРОВ,  
КРИТЕРИЙ ФИЗИЧНОСТИ ВОЗМОЖНЫХ МИРОВ**

В докладе дана история появления различных концепций возможных миров, которые вошли в науку разными путями – как открытие в квантовой механике, так и в ходе решения семантических проблем в математической логике, или через желание математиков внедрить в физику категорный аппарат, альтернативный аппарату, основанному на теории множеств. Критерием реальности возможных миров в теоретическом плане, по мнению автора, может стать, применительно к ним, развитие такой же физики, как и в случае известной нам теоретической физики, прародителями которой были Ньютон и Лейбниц.

*Ключевые слова:* возможные миры, модальный реализм.

**A.K. Guts**

**MODAL REALISM,  
WAYS OF GENERATING POSSIBLE WORLDS,  
CRITERION OF PHYSICALITY OF POSSIBLE WORLDS**

In this report a history of the emergence of various concepts of possible worlds that have entered science in different ways – as discovery in quantum mechanics, as well as in the course of solving semantic problems in

mathematical logic, or through the desire of mathematicians to introduce into physics a categorical apparatus that is alternative to the apparatus based on set theory is provided. According to the author, the criterion of the reality of possible worlds in theoretical terms may be the development of the same physics applied to them as in the case of theoretical physics known to us, the progenitors of which were Newton and Leibniz.

*Keywords:* possible worlds, modal realism.

Возможные миры, которые были открыты и восприняты как нечто физически возможное в физике в то время, когда физики возвысились над лириками и были кумирами молодежи, то есть в 1950-годы, были практически сразу же забыты. Но в 1970-е годы они вновь появились в науке благодаря Брайсу ДеВитту (DeWitt, 1973), но уже под названием *параллельных вселенных*. Их первоначальное название – *relative states*, и знали об их существовании весьма немногие физики, но это были Джон Арчибальд Уилер и Нильс Бор. Нильс Бор был напуган дерзкой идеей аспиранта Джо Эверетта III (Everett, 1957). Общение последнего с Бором, а Эверетт летал к нему в Данию, закончилось тем, что Эверетт ушел из науки и подался в атомное ведомство. Диссертацию по теории *relative states* он всё-таки защитил, но благодаря настойчивости Уилера (Everett, 1973). Слава создателя возможных параллельных миров в квантовой механике пришла к нему в 1972, когда ДеВитт собрал конференцию по *многомировой* интерпретации квантовой механики, или как ее стали позже называть *эвереттовской интерпретации* квантовой механики.

Но миры Эверетта никак не описаны физически в рамках математического аппарата квантовой механики – они всего лишь слагаемые бесконечного ряда, и лишь одно из них есть наш реальный мир, согласно *боровской*, или копенгагенской интерпретации квантовой механики. Точнее, миры Эверетта такие же, как наш мир, но что-то в них *не так*, и это «не так» не случилось, поскольку *маловероятно*. Однако, Эверетт выдвинул концепцию, что в определенный момент, когда все эти миры *реализуются*, то есть Реальность разветвляется, одновременно разветвляется и наше сознание, и каждый возникший мир наблюдается какой-то частью нашего сознания. И таких маловероятных «не так» беско-

нечно много. Однако, где они существуют, где их искать, квантовая механика ничего не говорит. Впрочем, о разветвлении, или, как говорят, о раздвоении личности мы что-то слышали от психиатров, которые считают это заболеванием.

Тем не менее, попытку наполнить возможные миры Эверетта физическим содержанием предпринял один из создателей квантовых вычислений Давид Дойч, говоря о *теневых* частицах, заполняющих *возможные теневые миры*. Книга Дойча «Fabric of Reality» (1997) широко обсуждалась, однако эти теневые миры никак физически не описаны, они фактически являются мирами Эверетта, поскольку Дойч использовал ту же квантовую механику, что и Эверетт.

Здесь важно отметить, что миры Эверетта рождаются в рамках классической двузначной логики, поскольку оговоренный выше ряд – это всего лишь классическое решение классического дифференциального уравнения Шрёдингера. Разговоры же о существовании особой квантовой логики, лучше сказать квантовых логик, поскольку их придумано множество, мало соотносится в какой-то мере с классической логикой исследований специалистов по квантовой механике. Миры Эверетта вне модальных логик.

Теневым частицам Дойча было найдена чисто физическая интерпретация как фермионам с нулевым тензором энергии-импульса. Сделала это молодой математик Елена Палешева (Палешева, 2001), защитившая на эту тему кандидатскую диссертацию по теоретической физике. Эти частицы, а, точнее, соответствующее им решение уравнений Дирака были обнаружены в 1970-е годы и известны под названием призраки или духи (ghosts). Было показано, что призраки могут взаимодействовать с частицами нашего мира (Палешева, 2002). И более того, возможны переходы из теневого мира в наш мир и обратно (Гуц, 2020). Таким образом, тут мы видим расхождение с утверждениями отдельных известных философов, что возможные миры не взаимодействуют – каждый сам по себе. Но в данном случае, установление взаимодействия возможных миров друг с другом – это факты, полученные в рамках квантовой механики, то есть, общепринятыми в физике научными методами. Появляется возможность просчитывать самые различные детали проникновения одного мира в другой, находить условия таких проникновений,

определять как последствия таких общений, так их преимущества, делая, быть может, неожиданные открытия.

Вспомним, что есть чисто логический способ порождения возможных миров. И найден он был не физиками, а логиками. Этот способ действует вне классической логики, он обращается к познанию Мира с помощью модальной логики. Так рождаются возможные миры Льюиса, возрождаются миры Лейбница, рождаются при интерпретации формул модальной логики, известной как семантика Соула Крипке. Хотя было ясно, что интерпретации модальных логик многозначны, блестящее обращение Крипке к монадам Лейбница породило многообразие логически возможных миров с булевой логикой.

Вот только физика в работах Кларенса Льюиса и Крипке, по сути дела, отсутствует. Эти миры возможны, но в них нет физики, нет дифференциальных уравнений, нет материи. Неясно, где их искать, и как они выглядят. Впрочем, Льюис, и особенно Крипке, не придавали своим возможным мирам характера реального физического существования. Хотя другой Льюис, Дэвид Льюис, – авторитетный философ в области аналитической философии, будучи материалистом, был убежден в полной реальности возможных миров.

Попытку построить физику с возможными мирами и с уравнениями в них предпринял в 1970-е годы американский логик Уильям Ловер, обратившись к математической теории категорий. Но даже математики, интересующиеся физикой, его идеи не восприняли. Тем не менее, на свет появилась теория топосов, которая, однако, в категорном исполнении, то есть в рамках математической теории категорий, по-прежнему отпугивала искателей реальных возможных миров.

Тут интересно отметить, что хотя теория топосов как раз и задает многообразие возможных миров, поскольку ей была предназначена роль семантических моделей для формальной *синтетической дифференциальной геометрии* (СДГ), в рамках которой и пишутся столь необходимые дифференциальные уравнения, в которой содержится аналог дифференциального и интегрального исчисления, и в которой все функции дифференцируемы, что всегда было тем, на что не обращали внимания физики, но что так остро воспринимали математики (Коск, 1981).

Синтетическая дифференциальная геометрия, а лучше сказать, синтетическое дифференциальное и интегральное исчисление – это тот же математический аппарат, который подарили физикам Ньютон и Лейбниц, и благодаря чему были написаны уравнения Максвелла и Эйнштейна и все уравнения математической физики, вот только логика там интуиционистская, или, если хотите, модальная. И, следовательно, при семантических интерпретациях теорий, созданных с помощью синтетической дифференциальной геометрии, всегда будут появляться многообразия возможных миров. Но в этих мирах, уже всё есть – и уравнения, и физика, и, если пожелаете, раздвоение сознания. Однако, в СДГ не действует закон исключенного третьего, логика в СДГ интуиционистская, и это открывает путь к порождению многообразия возможных *классических* миров, то есть миров с классической логикой, как и в случае семантики Крипке.

Как видим, это тот же путь, что путь порождения возможных миров Крипке. Вначале, на языке синтетической дифференциальной геометрии пишется формальная теория  $T$  пространства-времени с уравнением Эйнштейна, – аналог общей теории относительности (Гуц, 2012), – затем в качестве ее интерпретации  $i$  берется один из гладких топосов, скажем, топос  $\mathbf{Sets}^{Lop}$ , или символически

$$i: \mathbf{Sets}^{Lop} \models T,$$

где  $L$  – категория так называемых локусов, состоящей из конечно порожденных гладких колец  $lA = C^\infty(\mathbf{R}^n)/I$  (Moerdijk, 1991),  $\mathbf{Sets}$  – категория множеств Кантора.

При такой интерпретации объектам топоса  $\mathbf{Sets}^{Lop}$ , например кольцу  $R$  – множеству, заменяющему в СДГ поле вещественных чисел, множеству функций в нем  $R^R$  и другим объектам ставятся в соответствии функторы  $F$  рассматриваемого топоса  $\mathbf{Sets}^{Lop}$  (категории), как-то,  $i(R) = F_R \in \mathbf{Sets}^{Lop}$ ,  $i(R^R) = F^F \in \mathbf{Sets}^{Lop}$  и пр.

При интерпретации  $i$  элементам  $x$  кольца  $R$  ставятся в соответствие «элементы»  $i(x)$  функтора  $F_R$  из  $\mathbf{Sets}^{Lop}$ . Но сделать это не так просто потому, что функтор  $F_R$  определен на категории локусов  $L$ ,  $F_R: L \rightarrow \mathbf{Sets}$ , то есть переменной (аргументом) является произвольный локус  $lA$ , а значением будет множество

$F_R(lA) \in \mathbf{Sets}$ . Иначе говоря, выбор того, что поставить в соответствие «числу»  $x \in R$  зависит от выбора локуса  $lA$ .

Выход из затруднения заключается в определении *обобщённых элементов*  $f \in_{lA} F$  для любого функтора  $F$  в *стадии* (at stage)  $lA$ , который по определению есть элемент  $f \in F(lA)$ .

Теперь сопоставляем элементу  $x \in R$  обобщённый элемент  $i(x) \in_{lA} lA$ . Но, как видим, таких элементов  $i(x)$  столько, сколько локусов. При переходе к интерпретации (модели) в топосе  $\mathbf{Sets}^{L_{op}}$  происходит «размножение», или «разветвление» элемента  $x$ . Он начинает существовать в бесконечном числе вариантов  $\{i(x) : i(x) \in_{lA} F_R, lA \in L\}$ .

Другими словами, появляется бесконечное число возможных миров с классической булевой логикой, и в каждом из них имеет место классическое дифференциальное и интегральное исчисления и, естественно, классическая теоретическая физика.

Чем различается математика в этих мирах? И в чем различие физических законов в них? Дело в том, что кольцо  $R$  отличается от поля вещественных чисел  $\mathbf{R}$  добавлением к последнему множества бесконечно малых элементов  $D = \{d \neq 0 : d^2 = 0\}$ . Поэтому известный всем факт, что любая фундаментальная физическая константа  $k$  измеряется лишь с некоторой точностью, в теории  $T$  сводится к постулату, что  $k = k_0 + d$ ,  $d \in D$ , и это сразу дает разветвление на множество различных возможных миров. В принципе, в каждом из них, физические константы точно не измеряются, но их отклонения от величины  $k_0$  у всех различны.

Отсюда следует, что доказательством истинности (реальности) существования множества возможных миров, по-разному отличающихся от нашего, является то, что, как всем известно, физические константы мы, в нашем мире, точно измерить не можем, и лишь только уточняем их. Заметим, что если слово stage перевести как «сцена», то возможные миры появляются в *Реальности* как *разные* спектакли с одним и тем же сценарием  $T$  на *разных сценах* и в значительной степени за счет участия актеров с *отличающимся по силе талантом*, который мы называем фундаментальной константой.

Подводя итог, можем сказать, что скепсис Крипке по поводу реальности возможных миров был оправдан, поскольку в его возможных мирах не было «материи» (физики), а была только булева

логика, а оптимизм Дэвида Льюиса был преждевременным, в его руках были только слагаемые из бесконечного ряда, дающего решения уравнению Дирака, но и то, если он был знаком с эвереттовской интерпретацией квантовой механики.

### Литература

1. *Everett, Hugh III.* Relative State. Formulation of Quantum Mechanics. // *Reviews of Modern Physics.* 1957. Vol. 29, №3. P. 454-462.
2. *Everett, Hugh III.* Theory of the universal wave function. // *The Many-Worlds Interpretation of Quantum Mechanics / Eds. B. DeWitt, N. Graham.* Princeton, New Jersey: Princeton University Press. 1973. P. 3-140.
3. *DeWitt, B. S.* The many-universe interpretation of quantum mechanics. // *The Many-Worlds Interpretation of Quantum Mechanics / Eds. B. DeWitt, N. Graham.* Princeton/New Jersey: Princeton University Press. 1973. P. 167-218.
4. *Палешева, Е. В.* Спинорные духи, теневые электроны и мультиверс Дойча. // *Математические структуры и моделирование.* 2001. Вып.8. С. 66-75.
5. *Палешева, Е.В.* Вклад спинорных духов в интерференцию квантовых частиц // *Математические структуры и моделирование.* 2002. Вып. 9. С. 142-157.
6. *Гуц, А. К.* (2020). Частицы-призраки, сцепленность исторических эпох и машина времени. *Математические структуры и моделирование.* №3 (55). С. 12-21.
7. *Kock, A.* Synthetic Differential Geometry. Cambridge Univ. Press. 1981.
8. *Moerdijk, I., Reyes, G. E.* *Models for Smooth Infinitesimal Analysis.* Springer-Verlag, 1991.
9. *Гуц, А. К.* (2012). *Физика реальности.* Омск: Изд-во КАН. 2012.

### References

1. *Everett, Hugh III.* (1957). "Relative State» Formulation of Quantum Mechanics. // *Reviews of Modern Physics.* Vol. 29, no.3. pp. 454-462.
2. *Everett, Hugh III.* (1973). Theory of the universal wave function. In *The Many-Worlds Interpretation of Quantum Mechanics / Eds. B. DeWitt, N.Graham.* Princeton/New Jersey: Princeton University Press, pp. 3-140. (In Eng.)

3. DeWitt, B. S. (1973). The many-universe interpretation of quantum mechanics. // *The Many-Worlds Interpretation of Quantum Mechanics*. / Eds. B. DeWitt, N.Graham. Princeton/New Jersey: Princeton University Press. Pp. 167-218.
4. Palesheva, E. V. (2001). Spinor ghosts, shadow electrons, and the Deutsch multiverse. *Mathematical Structures and Modeling*. 2001. № 8. Pp. 66-75. (In Russ.)
5. Palesheva, E.V. (2002). Contribution of spinor ghosts to the interference of quantum particles. / *Mathematical structures and modeling*. Issue 9. Pp. 142-157. (In Russ.)
6. Guts, A. K. (2020). Ghost particles, entanglement of historical epochs and a time machine. *Mathematical structures and modeling*. No. 3 (55). P. 12-21. (In Russ.)
7. Kock, A. (1981). *Synthetic Differential Geometry*. Cambridge Univ. Press.
8. Moerdijk, I., Reyes, G. E. (1991). *Models for Smooth Infinitesimal Analysis*. Springer-Verlag.
9. Guts, A. K. (2012). *Physics of Reality*. Omsk: Publishing House of the KAN. (In Russ.)

### Информация об авторе

Гуц Александр Константинович – доктор физ.-мат. наук, профессор, Профессор кафедры информационных технологий и математики Сочинский государственный университет.  
aguts@mail.ru

### Information about author

Guts Alexander Konstantinovich – Dr. Science in Physics and Mathematics, Professor Full Professor Department of Information Technology and Mathematics Sochi State University.  
aguts@mail.ru

Дата поступления 12.03.2025  
Принята к печати 11.12.2025

УДК: 165.1 + 004.8

DOI: 10.15372/PS20250516

EDN: JDJFTO

**О.И. Елхова**

**«ЧЕЛОВЕК В КОНТУРЕ»  
И ПРЕДЕЛЫ РАЦИОНАЛЬНОСТИ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

В статье рассматриваются философские основания и ограничения рациональности искусственного интеллекта в контексте принятия решений. Автор анализирует различие между эпистемической и практической рациональностью, акцентируя внимание на последней как основе работы рациональных агентов. Центральное место занимает концепция ограниченной рациональности, согласно которой принятие решений осуществляется в условиях неполной информации, когнитивных ограничений и ограниченных вычислительных ресурсов. Обосновано, что идеальная рациональность недостижима, а наиболее адекватной моделью для ИИ является ограниченная рациональность. Рассмотрены четыре типа рациональности: *идеальная*, *вычислительная*, *ограниченная* и *ограниченная оптимальность*, с выводом о практической применимости именно *ограниченной рациональности* в разработке интеллектуальных систем. Сделан вывод о практической применимости ограниченного подхода при разработке интеллектуальных систем. Автор отмечает, что необходимо учитывать указанные ограничения при создании адаптивных алгоритмов, а также значимость участия человека в процессе принятия решений для повышения надёжности результатов. Модель «человека в контуре» трактуется не просто как техническая форма взаимодействия, а как выражение ситуационной рациональности, предполагающей учёт контекста, моральных последствий и уникальности каждой конкретной ситуации. Присутствие человека придаёт алгоритмическому мышлению ценностный и интерпретативный характер, восстанавливая связь между рациональностью и практической мудростью (*phronesis*). В условиях, когда алгоритмы

ограничены ресурсами и подвержены ошибкам, именно человек способен выявлять контекстуальные нюансы и осуществлять осмысленную корректировку решений. Таким образом, включение человека в контур является важнейшим фактором, обеспечивающим повышение надежности решений, принимаемых искусственным интеллектом.

*Ключевые слова:* человек в контуре, ограниченная рациональность, искусственный интеллект, практическая рациональность, рациональный агент, принятие решений

**O.I. Elkhova**

## **«HUMAN IN THE LOOP» AND THE LIMITS OF ARTIFICIAL INTELLIGENCE'S RATIONALITY**

The article examines the philosophical foundations and limits of artificial intelligence rationality within the context of decision-making. The author analyzes the distinction between epistemic and practical rationality, emphasizing the latter as the fundamental basis for rational agent functioning. Central to the discussion is the concept of bounded rationality, according to which decision-making occurs under conditions of incomplete information, cognitive constraints, and limited computational resources. The article argues that ideal rationality is unattainable, making bounded rationality the most appropriate model for artificial intelligence. Four types of rationality are explored: ideal, computational, bounded rationality, and bounded optimality. The study concludes that bounded rationality is practically most applicable in developing intelligent systems. Furthermore, it highlights the necessity of considering these limitations when designing adaptive algorithms and underscores the significance of human involvement in the decision-making process to enhance reliability. The «*human-in-the-loop*» model is interpreted not merely as a technical mode of interaction, but as an embodiment of situational rationality, taking into account context, moral implications, and the uniqueness of each specific situation. Human participation adds a value-based and interpretative dimension to algorithmic reasoning, restoring the connection between rationality and phronesis. Given that algorithms are limited by resources and prone to errors, it is precisely the human who is capable of identifying contextual nuances and providing meaningful corrections. Thus, human inclusion in the loop emerges as a critical factor in enhancing the reliability of artificial intelligence-driven decisions.

*Keywords:* human-in-the-loop, bounded rationality, artificial intelligence, practical rationality, rational agent, decision-making

## Введение

Концепция рациональности в философии восходит к Аристотелю, который выделял два ее вида: эпистемическую и практическую. Первая связана с обоснованностью убеждений, вторая – с действиями и принятием решений. Такое разграничение отражает различие между процессом рассуждения и фактическим поведением. В эпистемическом смысле рациональность предполагает наличие аргументированного и надежно сформированного убеждения. Однако даже противоречивые суждения могут сохранять рациональный характер, если они обоснованы с точки зрения используемых методов познания.

Аристотель подчеркивает, что процесс рассуждения и само поведение представляют собой разные феномены. В настоящее время большее внимание уделяется рациональному поведению, поскольку оно связано с практическими результатами, для его описания используется понятие «рационального агента» – субъекта, принимающего решения, направленные на достижение оптимального результата, в том числе в условиях неопределенности. Практическая рациональность существенно отличается от эпистемической рациональности, которая сосредоточена на соответствии рассуждений логическим принципам. В отличие от эпистемической, она охватывает не только логические аспекты, но и мотивацию, желания, учитывает реальные ограничения, направлена на принятие решений и выбор наилучшего варианта действий в конкретных условиях. Именно эта функция закладывается в основу работы искусственного интеллекта, который должен анализировать данные, прогнозировать возможные последствия и определять наиболее эффективные решения. Такой процесс во многом напоминает человеческое практическое мышление, ориентированное на достижение целей, а не только на обоснование убеждений. Понятие «рационального агента» стало ключевым в дискурсе об искусственном интеллекте [1, с. 11–16]. В контексте **искусственного интеллекта** под рациональным агентом полагается система, способная анализировать данные, прогнозировать последствия действий и выбирать оптимальное решение с учетом заданных параметров в условиях неопределенности. В настоящей работе целью исследования является выявление философских ос-

нований и пределов рациональности искусственного интеллекта, поскольку в современных условиях все более актуальным становится детальный анализ его ограничений при обработке информации и принятии решений. Активное внедрение интеллектуальных систем в критически важные сферы жизнедеятельности человека требует пересмотра границ их рациональности и факторов, влияющих на точность и обоснованность принимаемых решений.

### Теоретический анализ

Несмотря на то, что искусственный интеллект изначально создавался для моделирования человеческого интеллекта, стоит признать, что решение человека не всегда может быть рациональным и приводить к оптимальным результатам, зачастую оказывается предвзятым. В этой связи стоит вспомнить концепцию *ограниченной рациональности* Г. Саймона, объясняющую процесс принятия решений через призму когнитивных ограничений индивидов и неполноты доступной информации. Данные взгляды существенно отличаются от традиционных представлений, когда человек рассматривается как полностью рациональный агент, способный всегда выбирать наилучший вариант, основываясь на объективном анализе. Классическая теория рациональности, словно не замечая своих скрытых предпосылок, исходит из убеждения, что индивиды обладают исчерпывающим знанием всех возможных альтернатив и их последствий. В свою очередь, Г. Саймон подчеркивает, что в реальных условиях люди редко стремятся к абсолютному оптимуму. Вместо этого они руководствуются принципом удовлетворительности, предпочитая решения, которые кажутся достаточно хорошими, но не обязательно являются наилучшими. В этом контексте рациональность играет ключевую роль, объясняя механизмы принятия решений. Когнитивные ограничения человека Г. Саймон образно сравнивал с одним лезвием ножниц, а структуру окружения – с другим лезвием: разум «режет» эффективно только благодаря взаимодействию ограниченных способностей с подсказками среды [8, с. 198–199]. Теория *ограниченной рациональности* ставит под сомнение пред-

ставление о том, что человек способен принимать решения, опираясь исключительно на объективные данные, без воздействия эмоций и когнитивных искажений, подчеркивая, что рациональность определяется не только итоговым результатом, но и самим процессом выбора. В действительности принятие решений включает анализ доступной информации, учет временных рамок и оценку значимости возможных альтернатив. Когнитивные барьеры, нехватка знаний и влияние внешней среды также препятствуют достижению идеального результата. Актуальность данных положений подтверждается исследованиями, посвящёнными вызовам цифровой эпохи, которые показывают, что развитие современных информационно-коммуникационных технологий, в том числе искусственного интеллекта, порождает принципиально новые вызовы, ставящие под вопрос устоявшиеся представления о рациональности и требующие их фундаментального переосмысления [3, с. 27–30].

Проблема рациональности искусственного интеллекта неразрывно связана с особенностями обработки информации. Такие аспекты проявляются не только в когнитивных ограничениях алгоритмов, но и в специфике взаимодействия человека с цифровой средой, включая виртуальную реальность. В данной среде структурные метрики феноменологического опыта оказывают влияние на субъективные механизмы принятия решений. Анализ феноменологии виртуального опыта показывает, что процесс восприятия цифрового пространства основывается на принципах, схожих с концепцией ограниченной рациональности. В ходе взаимодействия с виртуальной средой человек синтезирует сенсорные данные из обоих миров: физического и цифрового, что способствует формированию адаптивных когнитивных стратегий. Концепция полевой интерференции реального и виртуального в исследовании феноменологического опыта демонстрирует, что рациональность в цифровом пространстве нельзя свести к простому алгоритму оптимизации [2, с. 1003–1005]. Напротив, рациональность формируется под влиянием множества взаимосвязанных факторов: когнитивных, сенсорных и социальных, требующих комплексного подхода к изучению.

В ситуациях, когда количество возможных альтернатив слишком велико для исчерпывающего анализа, индивиды часто прибе-

гают к упрощенным эвристикам, что неизбежно приводит к отклонению от оптимального выбора. Кроме того, субъективные факторы, такие как личные ценности, цели и предпочтения, оказывают значительное влияние на процесс принятия решений, порой искажая объективную оценку возможных вариантов. Ошибки в суждениях могут возникать по следующим причинам: человек не обладает безграничными когнитивными ресурсами и может не знать о более выгодных альтернативах либо, напротив, ошибочно считать невыгодный вариант предпочтительным, т.е. в условиях неопределенности люди удовлетворяются с их точки зрения наиболее подходящим вариантом, а не максимизируют результат. В результате принятые решения не всегда соответствуют объективному максимуму.

Концепция Г. Саймона оказала глубокое влияние на ИИ, было выявлено, что и машины сталкиваются с теми же принципиальными барьерами, что и люди. Теория ограниченной рациональности остается актуальной и в современных исследованиях, особенно при разработке систем с ИИ. Можно выделить два подхода к моделированию рациональности в ИИ. Первый, «снизу-вверх», использует данные и входные сигналы от человека, что может вносить предвзятость. Искажения происходят из-за того, что данные, на которых обучаются модели, несут в себе ложные идеи и ограниченные представления, отражающие человеческие предрассудки. Второй, «сверху-вниз», основан на логике и формальной рациональности, но человеческие предубеждения могут быть заложены в самих правилах модели. Например, в системах машинного перевода могут быть использованы определенные языковые паттерны, которые искажают результаты. Современные исследования активно развивают гибридные подходы, которые пытаются объединить преимущества обеих моделей, сочетая данные и логику, чтобы минимизировать предвзятость и повысить рациональность решений.

В дискуссиях о возможностях искусственного интеллекта выделяют четыре типа рациональности: *идеальную, вычислительную*, а также концепции *ограниченной рациональности и ограниченной оптимальности* [7, с 36–38, с. 58–59]. *Идеальная рациональность* подразумевает принятие решения на основе полного знания всех возможных альтернатив и их последствий.

Только в реальных условиях такая рациональность остается недостижимой из-за неполноты информации и сложности решаемых задач. Кроме того, некоторые задачи требуют учета огромного количества переменных, что делает их расчет невозможным даже для самых мощных современных систем. В связи с этим более применимой концепцией становится *вычислительная рациональность*, которая учитывает доступные ресурсы и позволяет находить решения, приемлемые в рамках вычислительных ограничений. Однако и здесь существуют барьеры: чем сложнее задача, тем больше ресурсов требуется для ее решения, что делает невозможным анализ всех альтернатив в разумные сроки. Например, в задачах, связанных с распознаванием образов или прогнозированием, система вынуждена использовать эвристики и приближенные методы, а не полный анализ всех возможных решений. Многие задачи, с которыми сталкивается искусственный интеллект, относятся к классу вычислительно сложных, что означает их экспоненциальный рост по количеству возможных решений или даже неразрешимость за полиномиальное время. Ограниченность вычислительных ресурсов в сочетании со сложностью реальных задач принятия решений делает достижение оптимальных решений практически невозможным. Одним из проявлений данной проблемы является комбинаторный взрыв: стремительное увеличение числа возможных вариантов, которое делает полный перебор решений неосуществимым даже для самых мощных вычислительных систем. В системах искусственного интеллекта это особенно заметно в задачах оптимизации и планирования, где экспоненциальное разрастание возможных комбинаций требует применения эвристических методов и приближенных алгоритмов. В результате ИИ приходится ограничивать область рассуждений некоторым *фреймом* (рамкой), что неизбежно упрощает картину мира, но предотвращает паралич от избыточных данных.

Концепция *ограниченной оптимальности* (*bounded optimality*), сформулированная Стюартом Расселом, предполагает, что оптимальным является не агент, действующий в абстрактных условиях, а тот, который достигает наилучшего возможного результата в пределах заданных ресурсных ограничений. Согласно его определению, «ограниченная оптимальность расширяет традиционное понимание рациональности, вводя в него эксплицит-

ный учёт ограниченных вычислительных ресурсов реальных агентов» [6, с. 13]. Данные представления легли в основу ряда исследований по рациональному мета-рассуждению, когда интеллектуальные агенты выделяют часть ресурсов не только на само решение задачи, но и на оценку эффективности дальнейшего поиска (например, когда целесообразно прекратить поиск решения и перейти к его реализации). Таким образом, от Саймона берет начало линия исследований, рассматривающая интеллектуального агента как рационально действующего в пределах своих возможностей, а не в абсолютном выражении. Сегодня практически все успешные системы ИИ в том или ином виде реализуют принципы ограниченной рациональности, срезая неперспективные варианты из-за ресурсных ограничений.

Однако ограниченность ИИ определяется не только техническими факторами, но и фундаментальными математическими принципами. Еще А. Тьюринг доказал существование неразрешимых задач, одной из которых является проблема остановки – задача, которую никакая универсальная вычислительная машина не способна решить [11; 12]. Этот факт имеет принципиальное значение: у любой формальной системы интеллекта существуют границы, обусловленные не только доступными ресурсами, но и самими законами вычислимости. Следовательно, ограничения искусственного интеллекта – не просто следствие недостаточных вычислительных мощностей, а фундаментальное свойство любой алгоритмической системы. Так же принципиальным ограничением являются фундаментальные барьеры, выявленные теоремами К. Гёделя о неполноте. Если в любой достаточно мощной формальной системе существуют утверждения, которые невозможно ни доказать, ни опровергнуть, то сходные проблемы могут возникать и в системах с искусственным интеллектом [10, с 265–266, с. 276]. Все это означает, что определенные классы задач останутся принципиально нерешаемыми для машинных алгоритмов, независимо от их сложности и вычислительных мощностей.

Можно сказать, что *ограниченная оптимальность* представляет собой компромисс между точностью и затратами: она не требует исчерпывающего анализа всех возможных решений, но гарантирует нахождение приемлемого результата в разумные сроки. Данный подход особенно актуален для реальных вычислительных систем,

которые должны функционировать в условиях ограниченного времени, памяти и вычислительных мощностей. Но даже этот уровень рациональности не всегда достижим, поскольку алгоритмы могут сталкиваться с ресурсными ограничениями, а вычислительная сложность некоторых задач делает поиск удовлетворительного решения невозможным. Современные интеллектуальные системы проектируются с учетом ограниченной рациональности: признается, что идеально рациональный алгоритм часто неосуществим, поэтому алгоритмы должны быть эффективными в пределах доступных ресурсов. Например, в системах планирования и поиска решений используются эвристики, отсечение нерелевантных вариантов и «anytime-алгоритмы», которые могут прерывать вычисление при достижении приемлемого решения.

Поскольку идеальная рациональность остается недостижимой, вычислительная рациональность ограничена ресурсами, а даже ограниченная оптимальность не всегда возможна, то наиболее реалистичной и практически применимой становится *ограниченная рациональность*, которая позволяет адаптировать искусственный интеллект к реальным условиям и существующим ограничениям. В этой модели агент принимает решения не на основе поиска глобально оптимального решения, а в рамках доступных возможностей и ограничений. Другими словами, система выбирает удовлетворяющее решение (*satisficing*) вместо поиска абсолютного оптимума. Такой подход широко применяется в алгоритмах искусственного интеллекта, особенно в многокритериальном принятии решений, где важно найти баланс между качеством решения и затратами на его вычисление.

Концепция *ограниченной рациональности* также играет ключевую роль в междисциплинарных исследованиях, особенно в изучении взаимодействия человека и искусственного интеллекта. Внедрение ИИ ставит перед исследователями ряд этических проблем, связанных с распределением ответственности за ошибки искусственного интеллекта. При таких обстоятельствах требуется определить, на ком лежит ответственность: на разработчиках, операторах или самой системе. Значимой задачей становится достижение оптимального соотношения между автономностью алгоритмов и необходимостью человеческого контроля, что предоставит возможность минимизировать вмешательство и в то

же время предотвратить возможные риски. Так, современные исследователи, в частности Л. Флориди, подчеркивают значимость концепции гибридного интеллекта, который предполагает сочетание машинных алгоритмов с человеческой интерпретацией и корректировкой [4, с. 163–165].

Одним из центральных аспектов данной проблематики является роль «человека в контуре» (*human-in-the-loop*), т.е. участие человека в принятии критически важных решений для минимизации рисков. Поскольку искусственный интеллект не обладает способностью к сознательному восприятию информации, то его следует рассматривать не как самостоятельного агента, а как инструмент, чья эффективность зависит от взаимодействия с человеком. Выделяются три основные модели такого взаимодействия. Первая модель обозначается как «человек над контуром» (*human-on-the-loop*) и предполагает, что человек контролирует решения ИИ и может вмешаться при необходимости. Вторая, «человек в контуре» (*human-in-the-loop*), предусматривает активное участие человека в процессе принятия решений, а также контроль и корректировку результатов работы алгоритмов. Третья модель, «человек вне контура» (*human-out-of-the-loop*), описывает ситуацию, когда система функционирует полностью автономно, без вмешательства человека, что повышает риски, особенно в критически значимых сферах.

По мере развития искусственного интеллекта важно не только совершенствовать алгоритмы, но и придавать большое значение философски взвешенному проектированию. Нужно помнить, что исключение человека из цикла принятия решений не снимает с него ответственности, а лишь скрывает её и затрудняет контроль. В ситуациях, где последствия носят выраженный моральный характер (к примеру, распределение медицинских ресурсов, определение виновности в судопроизводстве или поведение автономного автомобиля при угрозе ДТП), ответственность должна оставаться за человеком. Такой подход основывается на принципиальных отличиях между вычислительными процессами и человеческим опытом, включающим эмпатию, совесть и совокупное жизненное знание. В качестве иллюстрации можно привести следующие примеры. ИИ активно применяется в медицинской диагностике, однако его выводы требуют проверки врачом. Так, ал-

горитмы Google Health в 2021 году продемонстрировали 94% точности при выявлении рака молочной железы, но не распознавали редкие формы опухолей, которые специалисты определяли лучше [5]. Данный факт подтверждает необходимость гибридного подхода, при котором врач оценивает результаты ИИ перед постановкой диагноза. Похожие проблемы наблюдаются в сфере беспилотного транспорта. В 2018 году автомобиль Uber не распознал пешехода с велосипедом в темное время суток, что привело к смертельному ДТП [9]. В настоящее время автомобильные компании, включая Tesla, внедряют модель *human-on-the-loop*, требующую обязательного контроля со стороны водителя даже при активированном автопилоте. Современные исследования свидетельствуют о необходимости рассматривать искусственный интеллект не как полностью автономную систему, а как инструмент, работающий в тесном взаимодействии с человеком.

Благодаря коллективным усилиям исследователей и практиков, стремящихся повысить эффективность совместной работы человека и искусственного интеллекта, был разработан подход *Human-AI Teams*, целью которого является оптимизация взаимодействия человека и технологий, минимизация рисков автоматизированных решений и повышение качества результатов в различных сферах деятельности. Сочетание машинной аналитики и человеческого опыта открывает новые перспективы в медицине, транспорте, экономике и других критически важных сферах. В ближайшие годы ожидается совершенствование технологий адаптивного ИИ, способного не только выполнять поставленные задачи, но и постепенно обучаться взаимодействию с человеком, учитывая его предпочтения и стиль работы. Поэтому концепция ограниченной рациональности и модель «человека в контуре» становятся неотъемлемой частью развития искусственного интеллекта, обеспечивая баланс между эффективностью и этической ответственностью. При этом такое участие невозможно без обращения к концепции ситуационной рациональности, согласно которой принятие решений не может быть полностью алгоритмизировано, ведь каждая ситуация содержит в себе элементы неповторимости и открытости. В отличие от автономных машин, человек не просто исполняет функцию контроля или вмешательства, а привносит в процесс решения ценностный и интерпретативный компонент. Это означает,

что рациональность в модели *human-in-the-loop* не сводится к сугубо инструментальному мышлению: она опирается на способность субъекта учитывать контекст, неопределённость, моральные последствия и уникальность каждой ситуации. Человек в контуре не только оценивает работу алгоритма, но и обнаруживает контекстуальные нюансы, недоступные машине, работающей исключительно на основе статистических зависимостей. Его присутствие восстанавливает связь между рациональностью и практической мудростью (*phronesis*), то есть пониманием конкретной жизненной ситуации, в которую вписано каждое решение.

Таким образом, модель *human-in-the-loop* позволяет не только повысить безопасность и адаптивность систем, но и сохранить человеческое измерение решений в условиях цифровой рациональности. Участие человека в контуре представляет собой философски обоснованное требование, подчёркивающее важность интеграции рационального выбора с экзистенциальной, этической и социокультурной обусловленностью человеческих поступков. Это, в свою очередь, свидетельствует о необходимости согласования логики алгоритмических процедур с «логикой жизни» – сопряжения формальной рациональности и того, что философия называет «рациональностью второго порядка», предполагающей способность к саморефлексии и обоснованию не только результатов, но и исходных предпосылок выбора.

### **Заключение и выводы**

В заключение еще раз подчеркнем, что рациональность искусственного интеллекта не может определяться идеальной моделью. Ограниченность вычислительных ресурсов и высокая сложность решаемых задач предопределяют необходимость обращения к концепции ограниченной рациональности. В этой связи применение концепции ограниченной рациональности становится необходимым условием для эффективного функционирования интеллектуальных систем. Их развитие требует сбалансированного сочетания формальной логики, доступных вычислительных мощностей и способности адаптироваться к динамично изменяющимся условиям окружающей среды. Перспективными направ-

лениями дальнейших исследований являются углубленный философский и методологический анализ гибридных моделей рациональности, объединяющих эвристические и формально-логические подходы. Важное значение также приобретает изучение этических последствий и социальных аспектов интеграции ИИ, что открывает широкие возможности для междисциплинарных исследований и практического применения полученных результатов. На основе анализа различных моделей рациональности установлено, что концепция «человека в контуре» (*human-in-the-loop*) играет ключевую роль в обеспечении контроля и корректировки решений, принимаемых искусственным интеллектом. Таким образом, ограниченная рациональность не только задает технические границы работы интеллектуальных систем, но и формирует концептуальный фундамент для разработки алгоритмов, учитывающих не только вычислительные, но и социальные и этические ограничения.

### Литература

1. *Бабич М.Ю.* Понятие рационального агента и многоагентные системы // Проблемы информатики в образовании, управлении, экономике и технике: сб. статей XVII Междунар. науч.-техн. конф. Пенза: ПДЗ, 2017. С. 11–16.

2. *Елхова О.И.* Метрики феноменологического виртуального опыта // Вестник Российского университета дружбы народов. Серия: Философия. 2024. Т. 28. № 4. С. 997–1013.

3. *Елхова О.И., Кудряшев А.Ф.* Современные вызовы информационно-коммуникационных технологий // Вестник Самарского государственного технического университета. Серия: Философия. 2024. Т. 6. № 3. С. 27–34.

4. *Allo P.* (ed.). *Putting Information First: Luciano Floridi and the Philosophy of Information.* Wiley-Blackwell, 2011. 208 p.

5. *Forster V.* Artificial Intelligence Improves Breast Cancer Diagnosis [Электронный ресурс] / V. Forster. Режим доступа: <https://www.forbes.com/sites/victoriaforster/2025/01/13/artificial-intelligence-improves-breast-cancer-diagnosis/> (дата обращения: 15.03.2025).

6. *Russell S.* Rationality and Intelligence: A Brief Update // In: Müller V.C. (ed.). *Fundamental Issues of Artificial Intelligence.* Cham: Springer, 2016. P. 7–28.

7. *Russell S., Norvig P.* Artificial Intelligence: A Modern Approach. Pearson Series in Artificial Intelligence, 2022. 1115 p.

8. *Simon H.* Models of Man: Social and Rational. New York: John Wiley and Sons, Inc., 1957. 287 p. URL: <https://archive.org/details/modelsofman0000herb> (дата обращения: 15.01.2025).

9. The Guardian. Uber self-driving car kills woman in Arizona [Электронный ресурс] // The Guardian. Режим доступа: <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe> (дата обращения: 15.03.2025).

10. *Tourlakis G.* Gödel's First Incompleteness Theorem via the Halting Problem // *Computability*. Springer, 2022. P. 265–280.

11. *Turing A.* Computing Machinery and Intelligence // *Mind*. 1950. Vol. 49. P. 433–460.

12. *Turing A.* On Computable Numbers, with an Application to the Entscheidungsproblem // *Proceedings of the London Mathematical Society*. 1936. Vol. 42. P. 230–265.

## References

1. *Babich M.Yu.* Ponyatie ratsional'nogo agenta i mnogoagentnye sistemy [The Concept of a Rational Agent and Multi-Agent Systems] // *Problemy informatiki v obrazovanii, upravlenii, ekonomike i tekhnike: sb. statei XVII Mezhdunar. nauch.-tekhn. konf. [Problems of Informatics in Education, Management, Economy, and Engineering: Proceedings of the 17th International Scientific and Technical Conference]*. Penza: PDZ, 2017. P. 11–16. (In Russ.)

2. *Elkova O.I.* Metriki fenomenologicheskogo virtual'nogo opyta [Metrics of Phenomenological Virtual Experience] // *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Filosofiya. [RUDN Journal of Philosophy]*. 2024. Vol. 28. No. 4. P. 997–1013. URL: <https://doi.org/10.22363/2313-2302-2024-28-4-997-1013>. (In Russ.)

3. *Elkova O.I., Kudryashev A.F.* Sovremennye vyzovy informatsionno-kommunikatsionnykh tekhnologii [Modern Challenges of Information and Communication Technologies] // *Vestnik Samarskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Filosofiya. [Bulletin of Samara State Technical University. Series: Philosophy]*. 2024. Vol. 6. No. 3. P. 27–34. (In Russ.)

4. *Allo P.* (ed.). *Putting Information First: Luciano Floridi and the Philosophy of Information*. Wiley-Blackwell, 2011. 208 p.

5. *Forster V.* Artificial Intelligence Improves Breast Cancer Diagnosis [Electronic resource] / V. Forster. Available at: <https://www.forbes.com/sites/victoriaforster/2025/01/13/artificial-intelligence-improves-breast-cancer-diagnosis/> (accessed: 15.03.2025).

6. *Russell S.* Rationality and Intelligence: A Brief Update // In: Müller V.C. (ed.). *Fundamental Issues of Artificial Intelligence*. Cham: Springer, 2016. P. 7–28.

7. *Russell S., Norvig P.* *Artificial Intelligence: A Modern Approach*. Pearson Series in Artificial Intelligence, 2022. 1115 p.

8. *Simon H.* *Models of Man: Social and Rational*. New York: John Wiley and Sons, Inc., 1957. 287 p. URL: <https://archive.org/details/modelsofman0000herb> (accessed: 15.01.2025).

9. The Guardian. Uber self-driving car kills woman in Arizona [Electronic resource] // The Guardian. Available at: <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe> (accessed: 15.03.2025).

10. *Tourlakis G.* Gödel's First Incompleteness Theorem via the Halting Problem // *Computability*. Springer, 2022. P. 265–280.

11. *Turing A.* Computing Machinery and Intelligence // *Mind*. 1950. Vol. 49. P. 433–460.

12. *Turing A.* On Computable Numbers, with an Application to the Entscheidungsproblem // *Proceedings of the London Mathematical Society*. 1936. Vol. 42. P. 230–265.

### **Информация об авторе**

*Елхова Оксана Игоревна* – Уфимский университет науки и технологий, доктор философских наук, доцент, профессор кафедры философии и культурологии (Республика Башкортостан, 450076, Уфа, ул. Заки Валиди, 32).

[oxana-elkhova@yandex.ru](mailto:oxana-elkhova@yandex.ru)

### **Information about the author**

*Elkhova Oxana Igorevna* – Ufa University of Science and Technology, Doctor of Philosophy, docent, Professor at the Department of Philosophy and Cultural Studies (Republic of Bashkortostan, 450076, Ufa, Zaki Validi Street, 32).

[oxana-elkhova@yandex.ru](mailto:oxana-elkhova@yandex.ru)

Дата поступления 10.12.2024

Принята к печати 11.12.2025

УДК: 41.1 (141.12)  
DOI: 10.15372/PS20250517  
EDN: JMXLAF

**Е.В. Ускова**

## **О МЕТАФИЗИЧЕСКИХ ОСНОВАНИЯХ НАТУРАЛИЗМА В ОТНОШЕНИИ СОЗНАНИЯ**

В статье рассмотрены метафизические основания натурализма по отношению к сознанию. Натурализм - это самая естественная и распространённая, но не лишённая противоречий философская позиция. Её слабыми местами является утверждение приоритета научного знания, в то время как критерии научности часто оказываются неопределёнными; провозглашение редукции как основного принципа, следование которому может лишить предмет исследования его специфики; отрицание значимости онтологических оснований при решении гносеологических вопросов. Натурализм в отношении сознания в свою очередь сталкивается с необходимостью утверждения онтологического статуса сознания: оно является отдельной онтологической сущностью и тогда по отношению к нему должны действовать все известные науке законы, либо не является отдельной онтологической сущностью и должно рассматриваться как свойство мозга или подобным образом. Нередуктивный натурализм в отношении сознания предстаёт как внутренне противоречивая философская позиция: отказываясь от редукции сознания к мозгу в попытке отстоять уникальность качественных свойств ментальных состояний, мы тем самым нарушаем один из основных постулатов натурализма. Получается, что онтологический статус сознания не может быть оставлен без внимания при рассмотрении сознания как особого феномена, от определения этого статуса напрямую зависит и любая наша теория сознания.

*Ключевые слова:* философия сознания, метафизика сознания, онтологический статус сознания, натурализм, физикализм, сциентизм, редукционизм.

E.V. Uskova

## ON THE METAPHYSICAL FOUNDATIONS OF NATURALISM IN RELATION TO CONSCIOUSNESS

The article examines the metaphysical foundations of naturalism in relation to consciousness. Naturalism is the most natural and widespread, but not devoid of contradictions, philosophical position. Its weak points are the assertion of the priority of scientific knowledge, while the criteria of scientificity often turn out to be vague; the proclamation of reduction as the main principle, adherence to which can deprive the subject of study of its specificity; denial of the significance of ontological foundations in solving epistemological issues. Naturalism in relation to consciousness, in turn, is faced with the need to assert the ontological status of consciousness: it is a separate ontological entity and then all laws known to science must apply to it, or it is not a separate ontological entity and must be considered as a property of the brain or something similar. Non-reductive naturalism regarding consciousness appears as an internally contradictory philosophical position: by refusing to reduce consciousness to the brain in an attempt to defend the uniqueness of the qualitative qualities of mental states, we thereby violate one of the main postulates of naturalism. It turns out that the ontological status of consciousness cannot be ignored when considering consciousness as a special phenomenon; any of our theories of consciousness directly depend on the definition of this status.

*Key words:* philosophy of consciousness, metaphysics of consciousness, ontological status of consciousness, naturalism, physicalism, scientism, reductionism.

В данной статье мы рассмотрим натурализм как философскую позицию по отношению к сознанию, а также попробуем указать на те метафизические установки, которые лежат в её основании.

Натурализм представляет собой наиболее распространённую философскую позицию, которую занимает большинство современных философов сознания, работающих в рамках аналитической философии. Несмотря на то, что натуралистическое рассмотрение сознания представляется как наиболее очевидное, проблема обоснования и согласования этой позиции с научными данными о сознании, с одной стороны, и феноменальными аспектами субъективного сознательного опыта, с другой стороны, остаётся.

Взгляд на сознание как на естественный феномен не требует никакого дополнительного подтверждения, это ясно любому учёному, изучающему работу сознания в его конкретных аспектах, это то, что Д. Чалмерс обозначает как «лёгкие» проблемы сознания. А вот согласование выводов и результатов научных исследований с рассмотрением сознания от первого лица, а также сам факт качественной окрашенности наших ментальных состояний, представляется гораздо более «трудной» проблемой сознания. Эти трудности связаны с онтологическим статусом сознания в том числе, потому что требуют от нас определённости по отношению к сознанию. Сознание признаётся как нечто физическое и имеющее самостоятельный онтологический статус по отношению к мозгу или не обладающее таким статусом и не существующее без своего материального носителя, либо имеющее иную, нефизическую природу и нуждающееся в объяснении. Вопрос об онтологическом статусе сознания является метафизическим по своей сути и его актуальность в последнее время осознаётся всё больше и больше, о чём свидетельствует увеличивающееся количество публикаций по метафизике сознания и не только. Представляется, что от нашего понимания природы сознания зависит не только философия сознания, но и другие направления философии, а также конкретные области естественно-научного исследования и работы в области искусственного интеллекта.

Натурализм выступает как заглавие для определённой философской позиции по отношению к сознанию, но суть этой позиции видится отдельным философам по-разному. В частности, интересный взгляд на историю натурализма как явления культуры, а не только философии, мы находим в статье У. Селларса «Появление натурализма» [8]. Каждая из натуралистических концепций по отношению к сознанию имеет как свои внутренние логические несогласованности, так и в целом натурализм представляет собой достаточно противоречивое явление.

Одни авторы полагают, что объяснительные возможности натурализма напрямую связаны с тем, что он собой представляет. Если мы понимаем натурализм как некий тезис, то его практически невозможно защищать в силу логической неопределённости составляющих его терминов и проблематичных следствий, которые из него вытекают. Более выигрышной позицией в этом смысле бу-

дет рассмотрение натурализма как «исследовательской программы», что позволяет сделать положения такой позиции менее чёткими и в целом не претендующими на чрезмерную научную строгость. «Исследовательская программа – это набор методологических диспозиций, диспозиций, которые принимают определённого вида аргументы или убеждения в качестве источников очевидного знания. С моей точки зрения, натурализм – это разделяемая многими исследовательская программа, подмножество максимального набора методологических диспозиций, которая рассматривает методы науки и только их в качестве основных источников очевидного знания» [7, с. 107]. Однако, как отмечает автор, данная позиция также не является логически непротиворечивой. Таким образом, вопрос об успешности натурализма, поставленный автором в заглавии своей статьи решается скорее отрицательно, т.е. натурализм не может рассматриваться как логически непротиворечивая позиция сама по себе, а значит она не может быть применена нами для объяснения чего-либо (например, сознания).

В целом многие авторы также сходятся на том, что возможно выделить три направления в натурализме: онтологический – пытающейся ответить на вопрос о сущности и природе сознания как такового, эпистемологический – рассматривающий сознание с точки зрения возможности его познания, и методологический натурализм – предлагающий конкретные методы и пути его осмысленного изучения. На данный момент нас будет интересовать именно онтологический натурализм.

При рассмотрении сознания натуралистическую позицию по отношению к нему занимают те философы и учёные, которые придерживаются монизма в его материалистическом варианте. Предполагается, что сознание однозначно имеет причину своего существования в головном мозге, но по-разному видится суть этого причинного отношения. Представители теории тождества (Ю. Плейс, Дж. Сمارт) полагают, что все процессы, происходящие в сознании соответственно тождественны физическим процессам в мозге, различия лишь в их описании (См. описание теории тождества и других теорий сознания в аналитической философии: Крейн Т.А. «Краткая история философии сознания в XX веке» – [3]). Философы, придерживающиеся концепции дуализма свойств, также исходят из того, что есть единственный матери-

альный субстрат – мозг, являющийся причиной сознания, при этом физиология мозга и сознание будут выступать в качестве разных свойств одного и того же органа. Биологический натурализм Дж. Сёрла также представляет собой материалистическую позицию по отношению к сознанию, автор полагает сознание естественным биологическим феноменом, появившимся в ходе эволюционного развития человеческого вида, однако он отдельно подчёркивает наличие феноменальных свойств сознания и невозможность их элиминации из нашего описания сознания. Позиция физикализма, рассматривающая сознание как функцию мозга или иного (возможно и не только биологического) субстрата, также отвечает духу материализма и является натуралистической.

Натурализм – это философская позиция, которая мыслится как во всём следующая за науками и разделяющая их истины. Конечно же, это объясняется прежде всего историческими причинами, а именно «логическим позитивизмом», который был обозначен его основателями (Р. Карнапом и У. Куайном) как проект, направленный против метафизики в философии, отстаивающий идею логической непротиворечивости и ясности. (См. P.M.S. Hacker: *Passing by the Naturalistic Turn: On Quine's Cul-De-Sac// How Successful is Naturalism?* Georg Gasser (Ed.), 2007 о том, как формировались идеи логического позитивизма в Венском кружке [5]).

Например, У. Куайн так и пишет: «...Реальность должна быть идентифицирована и описана самой наукой, а не предшествующей ей философией» [6, р. 21]. Основная задача, которая тогда остаётся за философией – это анализ и прояснение. Этот философский проект имеет уже более чем вековую историю. Сложности, с которыми сталкивается данное представление об идеях и задачах философского исследования, могут быть раскрыты при анализе предпосылок и следствий этой философской позиции.

О том, как связаны между собой темы метафизики и сознания, мы находим упоминание в программной статье Д. Иванова «Философия сознания». В ней он рассматривает изменения в предмете и методах, которые прошла философия на пути исследования сознания. В частности, он отмечает, что первоначально исследовательский проект по изучению сознания в рамках «логического позитивизма» исходил из отрицания метафизики, её изгнания. Однако те философы, которые стояли у истоков этого проекта, тем не

менее имели некоторые общие представления об устройстве мира и отталкивались от них (а значит метафизика неявным образом присутствовала в их размышлениях). Постепенно это было осознано и самими представителями аналитической философии и некоторые из них стали отказываться от антиметафизических позиций. «Возможно, именно подобного рода размышления привели аналитических философов во второй половине XX века к отказу от антиметафизических установок, которых придерживались некоторые представители ранней аналитической философии, например позитивисты» [1, с. 226]. Данного рода размышления ещё раз подчёркивают важность исследования метафизических (онтологических) оснований философских позиций.

Например, по мнению уже Г. Гассера и М. Стефана, которое они изложили в своей статье «Тяжёлое бремя доказательства в пользу биологического натурализма» [4], натурализм выступает в качестве некоторого набора программных стратегий понимания мира. Его центральная идея – это анализ мира исключительно научным способом. Только наука даёт истинное знание о реальности. Только наука обладает максимально точным знанием о нашем мире, его устройстве и каузальных взаимодействиях в нём. Вопросами о наиболее общих предельных основаниях нашего мира занимается конечно же онтология. Именно в рамках онтологии мы можем получить ответы на вопросы о том, то «реально», а что «нереально» в мире. С точки зрения натурализма, именно методы естественных наук осуществляют гегемонию по отношению ко всем способам поиска истины, именно науки могут дать нам наиболее точную информацию о том, что существует. Поэтому получается, что именно онтология зависит от того, что признаётся реальными самими науками.

При поисках ответа на вопрос о природе сознания представители конкретных научных дисциплин (нейрофизиологии, когнитивной психологии и других) обращаются к философии в попытках добыть некоторое общее представление о сознании. Но они не получают его, поскольку философия сама обращается к конкретным наукам за основными истинами и идеями по отношению к сознанию. Получается замкнутый исследовательский круг.

Чтобы миновать эти неизбежные онтологические противоречия, многие современные философы-натуралисты, как отмечают Г. Гассер и М. Стефан, «отказываются от исследований в области

онтологии, считая их необязательными, т.к. натурализм должен быть нейтральной онтологической позицией» [4, с. 160]. Конечно, учёных как таковых, может совершенно не интересовать онтологическая проблематика, но философы не могут оставить её без внимания и не попытаться ответить на вопрос об онтологических основаниях натурализма. «Онтология – это не маргинальная, а центральная тема для натурализма», полагают они [4, с. 160].

Стратегия избегания самого разговора об онтологии свойственна многим современным сторонникам натурализма. Один из таких вариантов натурализма предлагает Й.Л. Брандл, называя его «скромным натурализмом». Он не претендует ни на какие метафизические выводы и обобщения, являясь метафизически нейтральным, претендует только на объяснение сознания как естественного феномена, появившегося у живых существ по ходу их развития. Данный натурализм базируется на следующем объяснительном принципе: «Всё знание об эмпирическом мире получается из естественных источников знания, то есть из источников, чьё функционирование может быть рационально объяснено» [2, с. 259].

Укажем ещё на одну проблему натурализма, которая по мнению Г. Гассера и М. Стефана, заключается в сциентизме как философской позиции, из которой он исходит. Точно также как натурализм, сциентизм, лежащий в его основе, оказывается неоднозначным термином по своему объёму и содержанию. Зачастую не удаётся обозначить точный список самих наук, выделить принципы выбора одних наук по отношению к другим, учесть изменчивость самих наук и их методов и т.д. Избежать этой неоднозначности возможно, только если мы попробуем выделить ряд зарекомендовавших себя наук, таких как физика, химия и биология. Это «натурализм основных наук», который помогает избежать как крайнего «физикализма» (сведения всего многообразия явлений к взаимодействиям на уровне микромира), так и варианта «примирительного натурализма», расширяющего границы понятия до социальных и гуманитарных наук.

При непосредственном рассмотрении натуралистической позиции по отношению к сознанию мы приходим, как полагает Г. Гассер и М. Стефан к невозможности отрицания двух тезисов:

1) Редукционизма по отношению к сознанию, т.е. идеи сведения высшего (сознания) к низшему (мозгу);

2) Каузальной замкнутости физического, т.е. представления о том, что всё происходящее в мире имеет свои физические причины и может быть объяснено исключительно с помощью законов физики.

Если мы утверждаем 1-е положение, но отрицаем 2-е или наоборот (утверждаем 2-е, но отрицаем 1-е), то это ведёт к логическому противоречию [4, с. 174]. То есть, по мнению авторов, нередуктивный натурализм внутренне противоречив. Сама по себе попытка обоснования феноменальной природы сознания и желание учесть эту его характеристику несовместима с позицией редукционизма. Одновременно «спасти феномен» сознания, указав на его уникальную субъективную природу и качественные качества, с одной стороны, и объяснять его в терминах естественных наук, с другой, представляется непоследовательным. В этом смысле более последовательным выступает эпифеноменализм Д. Чалмерса, постулирующего сознание отдельным нефизическим свойством всего материального, которое должно быть не исключено из физического мира, а наоборот, включено в него. Но в этом случае нам придётся пересмотреть всё наше представление о мире, сформулированное в рамках науки. В попытках избежать логических противоречий в теории сознания мы приходим к идее пересмотра фундаментальных оснований всей научной картины мира, и онтологии как таковой. Изменения на уровне гносеологии влекут за собой изменения на уровне онтологии, что достаточно показательно.

Таким образом, мы полагаем, что рассмотрение метафизических оснований натурализма как философской позиции приводит нас к следующим выводам. Натурализм в отношении сознания не отменяет ответа на вопрос об онтологическом статусе сознания. Рассмотрение сознания как естественного биологического феномена помогает нам увидеть историю его появления и развития, снимает покров мистики с самого феномена сознания. Но естественно-научный подход оставляет проблему сознания неразрешённой. Нередуктивный натурализм в отношении сознания является интересной, но непоследовательной позицией по отношению к сознанию. Как нам избежать онтологических противоречий в отношении сознания (не утверждая его как отдельную онтологическую сущность) и одновременно учесть качественную уни-

кальность сознания (субъективный сознательный опыт), при этом оставаясь на позиции натурализма по отношению к сознанию, будет являться предметом дальнейших исследований.

### Литература

1. *Иванов, Д.В.* Философия сознания. // Философская антропология. 2018. Т. 4. № 2. С. 218–250.
2. *Brandl J.L.* The Unmysteriousness of Consciousness: A Case Study in Naturalistic Philosophy. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4, P. 243-266.
3. *Crane, T.* A Short History of the Philosophy of Consciousness in the Twentieth Century. // Philosophy of Mind in the Twentieth and Twenty-First Centuries. The History of the Philosophy of Mind. 2018. Vol. 6. London: Routledge, P. 78–103.
4. *Gasser G., Stefan M.* The Heavy Burden of Proof for Ontological Naturalism. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4. Lancaster, New Brunswick:ontos verlag, P. 159-182.
5. *Hacker P.M.S.* Passing by the Naturalistic Turn: On Quine's Cul-De-Sac. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4. Lancaster, New Brunswick:ontos verlag, P. 143-158.
6. *Quine, W.V.O.* Things and their Place in Theories. // Theories and Things. Cambridge, MA: Harvard University Press. 1981.
7. *Rea M. C.* How Successful Is Naturalism? // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. Lancaster, New Brunswick:ontos verlag, P. 105-116.
8. *Sellars R.W.* The Emergence of Naturalism. // The International Journal of Ethics. Vol. XXXIV (4), July 1924. P. 309-338.

### References

9. *Ivanov, D.V.* (2018). Philosophy of Consciousness. // Philosophical Anthropology. Vol. 4. No. 2. P. 218–250.
10. *Brandl J.L.* The Unmysteriousness of Consciousness: A Case Study in Naturalistic Philosophy. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4, P. 243-266.

11. *Crane, T.* A Short History of the Philosophy of Consciousness in the Twentieth Century. // Philosophy of Mind in the Twentieth and Twenty-First Centuries. The History of the Philosophy of Mind. 2018. Vol. 6. London: Routledge, P. 78–103.

12. *Gasser G., Stefan M.* The Heavy Burden of Proof for Ontological Naturalism. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4. Lancaster, New Brunswick: ontos verlag, P. 159-182.

13. *Hacker P.M.S.* Passing by the Naturalistic Turn: On Quine's Cul-De-Sac. // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. 4. Lancaster, New Brunswick: ontos verlag, P. 143-158.

14. *Quine, W.V.O.* Things and their Place in Theories. // Theories and Things. Cambridge, MA: Harvard University Press. 1981.

15. *Rea M. C.* How Successful Is Naturalism? // How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society. New Series. 2007. Vol. Lancaster, New Brunswick: ontos verlag, P. 105-116.

16. *Sellars R.W.* The Emergence of Naturalism. // The International Journal of Ethics. Vol. XXXIV (4), July 1924. P. 309-338.

### **Информация об авторе**

*Ускова Екатерина Викторовна* (Uskova Ekaterina Victorovna) – к.ф.н., доцент кафедры Онтологии и теории познания, Уральский Федеральный университет им. Первого президента России Б.Н. Ельцина, 620002, г. Екатеринбург, ул. Мира, 19.

[e.v.uskova@urfu.ru](mailto:e.v.uskova@urfu.ru)

### **Information about author**

*Uskova Ekaterina Victorovna* – Associate Professor of the Department of Ontology and Theory of Knowledge, Ural Federal University named after the First President of Russia B.N. Yeltsin.

[e.v.uskova@urfu.ru](mailto:e.v.uskova@urfu.ru)

Дата поступления 18.03.2025

Принята к печати 11.12.2025

УДК: 001.89

DOI: 10.15372/PS20250518

EDN: JNVDHO

**А.Ю. Миронкина**

### **ФОРМУЛА НАХОЖДЕНИЯ ЧИСЛА ЦИТИРОВАНИЯ СТАТЕЙ УЧЁНЫХ ПРИ ИЗМЕНЕНИИ ИНДЕКСА ХИРША**

В статье предлагаются методические рекомендации для наиболее точного отслеживания наукометрического показателя индекса публикационной активности и востребованности публикаций учёного, известного в науке как индекс Хирша (h-индекс). Автор знакомит с разработанной формулой нахождения роста прибавленных за определённый период времени цитат учёного, влияющих на повышение индекса Хирша, рассказывает о параметрах его зависимости, даёт советы по практическому неавтоматизированному и автоматизированному расчёту индекса, описывает область его применения. Ежегодный, или же полугодовой, мониторинг индекса роста активных цитирований, по мнению автора, призван стимулировать учёных к дальнейшей публикационной деятельности.

*Ключевые слова:* РИНЦ, наукометрия, индекс цитирования, индекс Хирша.

**A.Yu. Mironkina**

### **THE FORMULA FOR FINDING THE NUMBER OF CITATIONS OF ARTICLES BY SCIENTISTS WHEN THE HIRSCH INDEX CHANGES**

The article offers methodological recommendations for the most accurate tracking of the scientometric index of publication activity and demand for publications of a scientist known in science as the Hirsch index

(h-index). The author introduces the developed formula for finding the growth of the scientist's citations added over a certain period of time that affect the increase in the Hirsch index, talks about the parameters of its dependence, gives advice on practical non-automated and automated calculation of the index, describes the scope of its application. The annual, or semi-annual, monitoring of the growth index of active citations, according to the author, is intended to stimulate scientists to further publication activities.

*Keywords:* RSCI, scientometry, citation index, Hirsch index.

## Введение

Не первый год уже во многих странах мира, в том числе и в Российской Федерации, в сообществе учёных большое внимание уделяется определению темпов исследовательской деятельности в науке. Для этого подсчитываются публикации и цитирования (ссылки на публикации), разработаны ряд индексов учёта активности учёных.

Наиболее распространённый среди всех индексов является индекс Хирша (h-индекс). Данный индекс предложен в 2005 г. аргентино-американским физиком Хорхе Хиршем. Индекс Хирша легко рассчитывается, он основан на числе публикаций и количестве цитирований данных публикаций в других научных работах. Индекс Хирша равен  $h$ , если максимальное число  $h$  опубликованных автором научных работ, процитировано исследователями в научной литературе не менее  $h$  раз [Hirsch, 2005]. Так, если индекс Хирша учёного равен 8, значит, у автора есть максимальное число 8 научных работ, каждая из которых цитируется минимум 8 раз в других научных изданиях, т.е. у автора 64 активных цитирований.

Ведущее место в системе учёта деятельности учёных в нашей стране занимает Российский индекс научного цитирования (РИНЦ). Данный индекс размещён в электронной библиотеке РИНЦ на сайте [elibrary.ru](http://elibrary.ru). В нём содержится широкая база данных о научных учреждениях, научных изданиях, современных учёных, научных публикациях и т.п. Общее число публикаций в изданиях РИНЦ, число цитирования публикаций в изданиях РИНЦ и индекс Хирша – основные показатели научной продуктивности учёных и научных учреждений. Библиотека РИНЦ даёт возможность детально ознакомиться с анализом публикаций и цитирований учёного.

Помимо индекса Хирша в научной литературе разработаны ещё несколько индексов, позволяющих в определённых случаях мониторить деятельность учёных.

*g*-индекс. Все статьи учёного отсортировываются в порядке убывания по количеству цитирований. Наибольшее число самых цитируемых статей *g*, имеющих суммарно не менее  $g^2$  цитирований является *g*-индексом [Egghe, 2006].

*AR*-индекс. В расчётах индекса Хирша включается возраст публикации [Jin, 2007].

*e*-индекс. Учитывает публикации не учтённые при расчётах индекса Хирша [Zhang, 2009].

*hg*-индекс. Данный индекс комбинируется на основе *h*-индекса и *g*-индекса [Alonso et al., 2010].

*Sh*-индекс. «В *Sh*-индексе целая часть является обычным индексом Хирша, а дробная показывает, насколько ученый приблизился к следующему значению индекса Хирша. [Штовба, Штовба, 2011].

*h+* индекс. Данный индекс также представляет дробную модификацию индекса Хирша «Индекс Хирша+ зависит от количества напечатанных цитируемых статей в журналах РИНЦ и индекса цитирования, обозначающий полное количество ссылок на научные работы автора (или соавтора) в других научных источниках» [Миронкина, 2016. С. 127].

*Ра*-индекс (индекс ПАИВПУ). Индекс публикационной активности и востребованности публикаций учёного. Предлагаемый индекс зависит от четырёх параметров: напечатанного числа публикаций в изданиях индексируемых РИНЦ, коэффициента участия автора в написании данной публикации, числа цитирований каждой из напечатанных публикаций в изданиях индексируемых РИНЦ и количества лет, прошедших с года первой напечатанной публикации. [Миронкина, 2016. С. 176].

### **Расчёт цитирований, повышающий индекс Хирша**

С каждым годом в учебных исследовательских заведениях руководители всё больше внимания уделяют индексу Хирша своих работников, ведь, этот показатель определяет состояние научной деятельности учёного, научной деятельности всего коллектива заведения. Для этого разрабатываются стимулирующие методы заин-

тересовать соискателей к написанию научных трудов, причём, как можно лучшего качества. Стимулирование, как правило, затрагивает число написавших статей, с учётом издания их опубликовавшее и число цитирований (ссылок), позволивших учёному увеличить свой индекс Хирша. С первым вопросом дело обстоит проще: администрация устанавливает размер доплаты на каждую статью и по итогам года (или семестра) производит подсчёт по данным базы электронной библиотеки РИНЦ. Несколько сложнее обстоит дело с учётом цитирований, повышающих индекс Хирша.

Рассмотрим зависимость числа цитирований от повышения индекса Хирша на один шаг.

Таблица 1

**Зависимость числа цитирований  
от номера перехода индекса Хирша к следующему номеру**

Увеличение индекса Хирша	Число цитирований
с 0 до 1	1
с 1 до 2	3
с 2 до 3	5
с 3 до 4	7
с 4 до 5	9
с 5 до 6	11
с 6 до 7	13
с 7 до 8	15
с 8 до 9	17
с 9 до 10	19
с 10 до 11	21
с 11 до 12	23
...	...
с 20 до 21	41
с 21 до 22	43
...	...
с 30 до 31	61
с 31 до 32	63
...	...
с 40 до 41	81
с 41 до 42	83
...	...

Источник: таблица составлена автором.

Анализируя данные таблицы 1, легко заметить, что при переходе индекса Хирша от показателя  $n$  к показателю  $(n+1)$ , число цитирований увеличивается на  $(2n+1)$ .

Так, при переходе с индекса равного 5 на индекс равный 6, число цитирований возрастает на  $2 \times 5 + 1 = 11$ .

А при переходе с индекса 16 на индекс 17, число цитирований возрастает на  $2 \times 16 + 1 = 33$ .

Вывод простой: чем выше индекс Хирша, тем больше необходимо цитирований, причём, мы учитываем, только цитаты, которые влияют на данное повышение индекса.

Для подсчёта числа цитирований при переходе с индекса Хирша от показателя  $n$  к показателю  $(n+k)$ .

Мы составили формулу:  $S = k \times (2n+k)$ ,

Где  $S$  – число цитирований.

$k$  – число, на которое увеличился индекс Хирша.

$n$  – первоначальный индекс исследователя.

Первоначальный индекс исследователя  $n$  должен находиться у руководителя, число  $k$  легко высчитать, взяв данные в базе электронной библиотеки РИНЦ. Подставив данные в формулу, мы получим число активных цитирований  $S$ . Зная показатель  $S$  и установленную надбавку за одно цитирование, мы можем узнать общую сумму за полученные цитирования.

Примеры подсчёта.

### **Пример 1.**

Найдём число цитирований, при переходе от индекса Хирша, равного 5 к индексу Хирша, равного 9.

$$n = 5, \quad k = 9 - 5 = 4.$$

$$\text{Подставим в формулу, получим: } S = 4 \times (2 \times 5 + 4) = 56.$$

### **Пример 2.**

Найдём число цитирований, при переходе от индекса Хирша, равного 31 к индексу Хирша, равного 35.

$$n = 31, \quad k = 35 - 31 = 4.$$

$$\text{Подставим в формулу, получим: } S = 4 \times (2 \times 31 + 4) = 264.$$

В обоих примерах индекс Хирша увеличивался на 4 единицы, но в первом случае для этого потребовалось 56 цитат, а во втором – 264. Разница весьма существенная.

## Вывод

Формула  $S = k \times (2n+k)$  для подсчёта числа цитирований при переходе индекса Хирша с показателем  $n$  к показателю  $(n+k)$  может выполняться как вручную, так и автоматически. Для этого следует воспользоваться, например, компьютерной программой Microsoft приложением Excel, в арсенале которой легко ввести функцию для нахождения нужной нам величины. Руководителю подразделения хорошо бы иметь подобную таблицу.

Таблица 2

**Таблица учёта написанных статей и цитирований учёного за определённый порядок времени**

	Фамилия И.О. учёного	По результатам РИНЦ было			По результатам РИНЦ стало			n	k	S
		Число статей	Число цитат	Индекс Хирша	Число статей	Число цитат	Индекс Хирша			
1	Иванов И.И.	32	172	9	37	204	14	9	5	115
2										
3										
4										
5										
6										
7										
...										

*Источник:* таблица составлена автором.

В приведенной таблице 2:

$n$  – индекс Хирша, который был первоначально;

$k$  – разность индекса Хирша, который стал и индекса Хирша, который был первоначально;

$S$  – находится по предлагаемой нами формуле  $S = k \times (2n+k)$ .

Производя регулярный мониторинг написанных статей и активных цитирований, позволяющих увеличить индекс Хирша за рассматриваемый период времени, год или же полугодие, руководитель легко будет видеть рост научной деятельности своих подчинённых и это будет стимулировать учёных для написания большего количества работ, причём, как можно лучшего качества.

## Литература

1. *Миронкина А.Ю.* Новый наукометрический показатель – индекс Хирша+ // *Философия науки.* – 2016. – №1 (68). – С. 127-132.
2. *Миронкина А.Ю.* Ра-индекс – индекс публикационной активности и востребованности публикаций ученого // «Философия науки». 2016. №4(71). С. 176-185.
3. *Штовба С.Д., Штовба Е.В.* Sh-индекс – новая дробная модификация индекса Хирша // *Научные труды Винницкого национального технического университета.* – 2011. – №3. – С. 6.
3. *Alonso S., Cabrerizo F., Herrera-Viedma E., Herrera F.* Hg-Index: A new index to characterize the scientific output of researchers based on the hand G-Indices // *Scientometrics.* – 2010. – Vol. 82. – No. 2. – P. 391-400.
4. *Egghe L.* An improvement of the H-Index: the G-Index // *ISSI Newsletter.* – 2006. – No. 2(1). – P. 8-9.
5. *Hirsch J.E.* An index to quantify an individual's scientific research output // *Proc. Natl Acad. Sci. USA.* – 2005. – Vol. 102, No. 46. – P. 16569-16572.
6. *Jin B.H.* The AR-Index: complementing the H-Index // *ISSI Newsletter.* – 2007. – Vol. 3. – No. 1. – P. 6.
7. *Zhang C.T.* The E-Index, complementing the H-Index for excess citations // *PLoS ONE.* – 2009. – Vol. 4. – No. 5.

## References

1. *Mironkina, A.Yu.* (2016) Novyy naukometricheskiy pokazatel – indeks khirsha+ [A new scientometric indicator – the H-index+]. *Philosophiya nauki.* [Philosophy of science], 1 (68), 127-132.
2. *Mironkina, A.Yu.* (2016) The pa-index: the index of a scientist's publication Activity and the relevance of his or her Publications. *Philosophiya nauki.* [Philosophy of science], 4 (71), 176-185.
3. *Shtovba, S.D, and E.V. Shtovba* (2011). Sh-index – novaya drobnaya modifikatsiya indeksa Khirsha [Sh-index – a new fractional modification of the Hirsch-index]. *Nauchnye Trudy Vinnitskogo Natsionalnogo Tekhnicheskogo Universiteta* [Scientific Works of the Vinnitsa Nationa Technical University], 3.
3. *Alonso, S., F.Cabrerizo, E.Herrera-Viedma, and F.Herrera.*(2010). Hg-Index: A new index to characterize the scientific output of researchers based on the hand G-Indices. *Scientometrics* Vol. 82, No. 2, 391-400.
4. *Egghe, L.* (2006). An improvement of the H-Index: the G-Index // *ISSI Newsletter*, 2 (1) 8-9.

5. *Hirsch, J.E.* (2005). An index to quantify an individual's scientific research output // Proc. Natl. Acad. Sci. USA, Vol. 102, No. 46, 16569-16572.

6. *Jin B.H.* (2007) The AR-Index: Complementing the H-Index // ISSI Newsletter. Vol. 3. No. 1. P. 6.

7. *Zhang, C.-T.* (2009). The E-Index, complementing the H-Index for excess citations. PLoS ONE, Vol. 4. No. 5.

### **Информация об авторе**

*Миронкина Алина Юрьевна* – кандидат экономических наук, доцент, Смоленская государственная сельскохозяйственная академия (214000, Смоленск, ул. Большая Советская, д.10/2).  
alina24m@mail.ru

### **Information about the author**

*Mironkina Alina Yuryevna* – the candidate of economic sciences, docent, Smolensk state agricultural Academy (10/2, Bolshaya Sovetskaya str, Smolensk, 214000, Russia).  
alina24m@mail.ru

Дата поступления 10.04.2025

Принята к печати 11.12.2025

## К СВЕДЕНИЮ АВТОРОВ

1. «Философия науки» – официальный научный журнал Сибирского отделения Российской академии наук и Института философии и права СО РАН, в котором бесплатно публикуются:

а) ранее не опубликованные научные статьи, содержащие новые результаты исследований в области философии, методологии, логики и истории естественных наук и математики, а также по фундаментальным проблемам естествознания;

б) ранее публиковавшиеся работы и архивные материалы по философии, методологии, логике и истории естественных науки и математики, содержащие оригинальные и актуальные для развития современной науки идеи, малоизвестные или неизвестные широкому кругу читателей;

в) рецензии на работы по философии, методологии и логике естественных наук и математики, опубликованные в других изданиях;

г) сообщения о проводимых научных конференциях, симпозиумах, конгрессах;

д) краткие научные сообщения, заметки, письма.

Основные разделы журнала:

«Общие вопросы истории и философии науки»,

«Проблемы логики и методологии науки»,

«Из истории науки»,

«Из архивов»,

«В помощь изучающим историю и философию науки»,

«Научная жизнь. Обзоры. Рецензии»,

«Из писем в редакцию».

2. Журнал включен в перечень ведущих рецензируемых научных журналов, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, доктора наук по философии.

3. Статьи, поступившие в редакцию, проходят экспертизу членов редколлегии и при необходимости направляются на внешнее рецензирование. О принятом решении автор извещается по адресу электронной почты, указанной в статье. Если решение отрицательное, редколлегия не вступает в переписку и теоретические дискуссии.

Недопустимо представлять в редакцию ранее опубликованные работы, а также статьи, скомпилированные из цитат и/или изложения ранее опубликованных работ, если при этом отсутствует новизна, связанная с развитием идей, изложенных в предыдущих работах.

4. Материалы должны быть тщательно подготовлены к публикации (см. образец оформления):

а) обязательным условием публикации материалов является наличие УДК, отвечающего основным разделам журнала;

б) в заглавии должны быть указаны инициалы и фамилия автора (авторов), название статьи;

в) после заглавия следует аннотация (на русском и английском языках, до 200 слов), где должна быть сформулирована основная идея работы и приведены ключевые слова (до 10 слов).

Название статьи и инициалы и фамилия автора (авторов) в аннотации представляются на русском и английском языках.

г) в конце статьи (после списка цитируемой литературы) в информации об авторе (авторах) необходимо полностью указать фамилию, имя, отчество, полное наименование места работы автора (авторов), должность, ученую степень и звание (при наличии), почтовый адрес, номер телефона, факса, электронную почту.

Информация об авторе (авторах) предоставляется на русском и английском языках.

#### 4.1. Библиографические ссылки (см. образец оформления):

а) в тексте статьи в квадратных скобках указывается порядковый номер источника из списка ссылок, если приводится цитата – указываются страницы.

б) список ссылок оформляется в конце статьи в двух обязательных вариантах:

«Литература». Библиографическое описание ссылки включает: фамилию и инициалы автора (авторов), полное название публикации, издания (журнал, сборник, в котором опубликована цитируемая статья), город, название издательства или издающей организации, год издания, том (для многотомных изданий), номер, выпуск (для периодических изданий). При отсутствии цитирования, но при наличии ссылки, указывается объем публикации (для статей – первая и последняя страницы).

В описании ссылки на источник из Интернета кроме URL также обязательно указывается дата обращения (пример ссылки: *Михалевская А.С.* Миф и наука: проблемы взаимодействия в современном обществе: Автореф. дисс. ... канд. филол. наук. – URL: <http://cheloveknauka.com/mif-i-nauka-problemy-vzaimodeystviya-v-sovremennom-obschestve#ixzz4iF3XIO1> (дата обращения 30.08.2016)

Сначала в алфавитном порядке помещается список цитируемых источников на русском языке, за ним – список иностранных источников (в порядке расположения букв латинского алфавита). Нумерация источников сплошная.

«References». Русскоязычные источники должны быть представлены в транслитерации и, в квадратных скобках, переводе авторов и названий на английский язык (нумерация сохраняется).

в) комментарии автора статьи оформляются в виде сноски (текст сноски располагается внизу страницы). При наличии в комментарии ссылки на источник последний оформляется как в тексте.

4.2. Переводы иностранных статей должны предоставляться с указанием полных выходных данных источника и текста оригинала. Переводчик должен получить у издателя или автора разрешение на публикацию перевода.

5. Материалы должны быть направлены только в электронной форме по E-mail: [sasha\\_khl@mail.ru](mailto:sasha_khl@mail.ru).

В электронной форме материалы должны быть набраны в программе Word for Windows (любая версия). Основной шрифт – Times New Roman. В случае необходимости использования других, нестандартных, шрифтов (вавилонских, иероглифов, древнегреческих, латыни и других начертаний) необходимо приложить файл шрифтов. Размер шрифта – 14 пт. Межстрочный интервал –

1,5 строки. Все страницы должны быть пронумерованы. Другие стилевые оформления, включая применение табуляции, не допускаются.

Рисунки, схемы и графики – в форматах, допускающих редактирование. Рисунки, схемы и графики должны быть только черно-белыми, без цветных элементов и мелких (сплошных) заливок. Желательно – отдельными файлами.

Таблицы оформляются в текстовом редакторе Microsoft Word. Номер таблицы выравнивается по правому краю (выделяется курсивом), заголовок таблицы по центру (выделяется жирным шрифтом).

Формулы (математические, физические, логические) должны быть набраны в редакторе MathType. Нумерация формул дается по правому краю. Недопустимо использование в тексте статьи автоматической нумерации формул.

6. Размер научных статей, публикаций – до 30000 знаков по статистике Word'a, включая пробелы и знаки препинания; размер рецензий – до 15000 знаков, кратких рецензий – до 3000 знаков; размер научных сообщений, заметок, писем – до 15000 знаков.

Публикации, превышающие указанные объемы, допускаются к рассмотрению только по согласованию с редколлегией.

7. Корректурa авторам не высылаeтся.

8. Если статья возвращается автору для доработки, исправления или сокращения, то датой представления ее в журнал считается день получения редакцией окончательного текста.

9. Гонорар за публикуемые статьи, доклады, сообщения и рецензии не выплачивается.

10. Не принятые к публикации материалы авторам не возвращаются.

## **ОБРАЗЕЦ ОФОРМЛЕНИЯ РУКОПИСИ И СПИСКА ЛИТЕРАТУРЫ**

УДК XXX.XX (Вместо X-ов вставить УДК, соответствующее статье)

**И.И. Иванов**

### **НАТУРАЛИСТИЧЕСКИЙ ПОВОРОТ: ПЕРВИЧНОСТЬ МЕТАФИЗИКИ<sup>1</sup>**

В работе приведен ряд примеров, демонстрирующих преимущества рассуждений с точки зрения натуралистических представлений при анализе классических постановок скептического аргумента: неопределенность указания в рамках семантического приоритета, неопределенность и релятивизм онтологических допущений как следствие принятия во внимание проблемы недоопределенности теории данными и конструктивистских представлений об онтологии.

*Ключевые слова:* натуралистический поворот, скептицизм, семантический приоритет, недоопределенность, конструктивизм, реализм.

---

<sup>1</sup> Статья подготовлена при финансовой поддержке XXXX фонда (проект № xx-xx-xxxx).

I.I. Ivanov

## NATURALISTIC TURN: PUTTING METAPHYSICS FIRST

The paper aims to illustrate one of the main features of the naturalistic turn – overcoming of the sceptic’s challenge by «putting metaphysics first» (M. Devitt). The sceptic’s argument is viewed as a consequence of the wrong direction of thought from a priori epistemology and/or semantics to a priori metaphysics. The naturalistic turn gives us a chance to give up the sceptic’s argumentation by allowing to argue from empirical metaphysics to empirical epistemology and/or semantics. Various examples of the advantage of the naturalized view are given. In particular, it is shown how to avoid scepticism, which is a follow-up of the indeterminism and underdetermination problems and constructivist ideology.

*Keywords:* naturalistic turn, scepticism, semantic priority, underdetermination, constructivism, realism.

## Примеры оформления постатейного списка литературы

1. *Горан В.П.* Онтологические и ценностные предпосылки доминирования рационализма и доминирования иррационализма в древнегреческой философии эпохи классического полиса / Материалы V Российского философского конгресса «Наука. Философия. Общество». Т. 2. Новосибирск: НГУ, 2009. С. 15–16. **(Если это ссылки на материалы семинаров, конференций, конгрессов и т.п.)**.

2. *Гемпель К.* Функция общих законов в истории / Розов Н.С. (ред.) *Время мира*. Вып. 1: Историческая макросоциология в XX веке. Новосибирск: НГУ, 2000. С. 13–26. **(Если это ссылки на статью в сборнике)**.

3. *Карпович В.Н.* Термины в структуре теории. Новосибирск: Наука, 1978.

4. *Методологическое обоснование структурных исследований в космологии* / И.И. Иванов, П.П. Петров, С.С. Сидоров и др. М.: Ин-т философии РАН. – 2000. – 315 с. **(Если это издание имеет трех и более авторов)**.

5. *Симанов А.Л., Сторожук А.Ю.* Историко-философские проблемы представлений о пространстве. Новосибирск: Изд-во Омега–Принт. – 2014. – 250 с. **(Если издание имеет двух авторов)**.

6. *Целищев В.В.* Рационалистический оптимизм и философия Курта Геделя // *Вопросы философии*. 2013. № 8. С. 12–23. **(Если это ссылка на журнал)**.

7. *Davidson D.* On the Very Idea of a Conceptual Scheme // *Proceedings and Addresses of the American Philosophical Association*. 1974. Vol. 47. P. 5–20.

## References

1. *Goran V.P.* Ontologicheskie i tsennostnye predposylki dominirovaniya ratsionalizma i dominirovaniya irratsionalizma v Drevnegrecheskoi filosofii epohi klassicheskogo polisa [Ontological and Axiological Background of the Prevalence of Rationalism and of the Prevalence of Irrationalism in the Ancient Greek Philosophy of the Classical City-State Epoch] / *Materialy V Rossiiskogo filosofskogo kongressa «Nauka. Filosofiya. Obshchestvo»*. Tom 2. [Proceedings of the Vth Russian Philosophical Congress «Science. Philosophy. Society». Vol. 2] Novosibirsk: NGU, 2009. P. 15–16. (In Russ.)

2. *Hempel C.* Funtsciya obshchih zakonov v istorii [The Function of General Laws in History] / Rozov N.S. (red.) Vremya mira. Vyp. 1: Istoricheskaya makrosotsiologiya v XX veke. [N. Rozov (Ed.) Vremia Mira. Vol. 1. The Historical Macrosociology on XXth Century] Novosibirsk: NGU, 2000. S. 13–26. (In Russ.)

3. *Karpovich V.N.* Terminy v strukture teorii. [Terms in the Theory Structure] Novosibirsk: Nauka, 1978. (In Russ.).

4. *Metodologicheskoe obosnovanie strukturnykh issledovaniy v kosmologii* [Methodological substantiation of structural studies in cosmology] / I.I. Ivanov, P.P. Petrov, S.S. Sidorov i dr. M., In-t filosofii RAN. – 2000. (In Russ.)

5. *Simanov A.L., Storozhuk A.Yu.* Istoriko-filosofskie problemy predstavlenii o prostranstve. [Historico-philosophical problems of ideas about the space] Novosibirsk: Isd-vo Omega-Print, 2014. (In Russ.).

6. *Tselishchev V.V.* Ratsionalisticheskii optimizm i filosofiya Kurta Gedelya [Rationalist Optimism and the Philosophy of Kurt Godel] // Voprosy filosofii. [Questions of Philosophy] 2013. № 8. S. 12–23. (In Russ.)

7. *Davidson D.* On the Very Idea of a Conceptual Scheme // Proceedings and Addresses of the American Philosophical Association. 1974. Vol. 47. P. 5–20.

#### Пример информации об авторе

Иван Иванович Иванов

Китежградский государственный университет, должность, ученая степень и звание (при наличии),

ул. Тупик, 2, Китежград, 630120, Россия

xxx@xxx.xxx (Вместо x-ов вставить соответствующий E-mail)

Ivan Ivanovich Ivanov

Kitezhgrad's State University, xxxxx.

2, Tupik st., Kitezhgrad, 630120, Russia

xxx@xxxx.xxx

**Таблица транслитерации символов**

Русский алфавит	Транслитерация	Русский алфавит	Транслитерация	Русский алфавит	Транслитерация
А	A	К	K	Х	KH
Б	B	Л	L	Ц	TS
В	V	М	M	Ч	CH
Г	G	Н	N	Ш	SH
Д	D	О	O	Щ	SHCH
Е	E	П	P	Ы	Y
Е	E	Р	R	Ь	'
Ж	ZH	С	S	Э	E
З	Z	Т	T	Ю	YU
И	I	У	U	Я	YA
Й	I	Ф	F		

## СОДЕРЖАНИЕ

---

### АНАЛИТИЧЕСКАЯ ФИЛОСОФИЯ: ТРАЕКТОРИИ ИСТОРИИ И ВЕКТОРЫ РАЗВИТИЯ-2025

<i>Salehi S.</i> A Bit of a Note on Halting Probabilities .....	3
<i>Lobovikov V.O.</i> A principle of equivalence of modalities de-dicto and de-re under the condition of a-priori-ness of knowledge, formulated by an artificial language of formal multimodal axiomatic theory $\Phi^{DR}$ .....	10
<i>Барышников П.Н.</i> Семантика интенсифицированных контекстов и генеративный искусственный интеллект .....	62
<i>Лисанюк Е.Н.</i> Гарри и пингвины в формальной теории аргументации .....	73
<i>Борисова И.И.</i> Табличная теория доказательств для гибридной модальной логики первого порядка .....	86
<i>Борисов Е.В.</i> Логическая репрезентация познаваемости .....	96
<i>Берестов И.В.</i> Проблема в теории изначальной апперцепции Канта .....	106
<i>Шевченко А.А.</i> Неявное знание: от М. Полани до <i>AlphaZero</i> .....	121
<i>Зайкова А.С.</i> О соизмеримости вопросов «Может ли машина мыслить?» и «Может ли машина победить в имитационной игре?» .....	134
<i>Хлебалин А.В.</i> Обосновывая практику логики .....	145

### ФИЛОСОФСКИЕ ПРОБЛЕМЫ ЛОГИКИ И МЕТОДОЛОГИИ НАУКИ

<i>Резников В. М.</i> Мешающие и способствующие факторы философской активности в науке, и некоторые перспективы участия философов в науке .....	156
<i>Головкин Н.В.</i> Пример Б. Уэзерсона, прагматические соображения и эпистемическое отношение укоренения .....	174
<i>Сторожук А.Ю.</i> Проблема измерения в квантовой механике с термодинамической точки зрения .....	190
<i>Власова О.А.</i> Диалог наук в энактивистской психиатрии: нейрофеноменология Т. Фука и вопрос о междисциплинарности .....	204
<i>Гуц А.К.</i> Модельный реализм, способы порождения возможных миров, критерий физичности возможных миров .....	219
<i>Елхова О.И.</i> «Человек в контуре» и пределы рациональности искусственного интеллекта .....	227
<i>Ускова Е.В.</i> О метафизических основаниях натурализма в отношении сознания .....	242
<i>Миронкина А.Ю.</i> Формула нахождения числа цитирования статей ученых при изменении индекса Хирша .....	252
К сведению авторов .....	260

---

*Научное издание*

**ФИЛОСОФИЯ НАУКИ**

**№ 5 (109)**

Оригинал-макет подготовлен Институтом философии и права СО РАН

---

Дата выхода в свет 29.05.2026 Формат бумаги 60×84<sup>1</sup>/<sub>16</sub>. Цифровая печать.  
Усл. печ. л. 15,5. Уч.-изд. л. 15,6. Тираж 100 экз. Заказ № 83.

---

Сибирское отделение РАН. 630090, г. Новосибирск, просп. Акад. Лаврентьева, 17.  
Отпечатано в Сибирском отделении РАН. 630090, г. Новосибирск, Морской просп., 2  
Тел. 330-84-66, e-mail: e.lyannaya@sb-ras.ru