РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ ФИЛОСОФИИ

EPISTEMOLOGY & PHILOSOPHY OF SCIENCE

ЭПИСТЕМОЛОГИЯ и ФИЛОСОФИЯ НАУКИ

Т. 63 • № 3

Ежеквартальный научно-теоретический журнал

МОСКВА
2026

ISSN 1811-833X (Print)
ISSN 2311-7133 (Online)

ЭПИСТЕМОЛОГИЯ И ФИЛОСОФИЯ НАУКИ

Ежеквартальный научно-теоретический журнал

2026. Том 63. Номер 3

Главный редактор: *И.Т. Касавин* (Институт философии РАН, Москва, Россия)
Зам. главного редактора: *А.Ю. Антоновский* (Институт философии РАН, Москва, Россия), *И.А. Герасимова* (Институт философии РАН, Москва, Россия), *П.С. Куслий* (Гёттингенский университет имени Георга Августа, Гёттинген, Германия)
Ответственный секретарь: *Л.А. Тухватулина* (Институт философии РАН, Москва, Россия)

Редакционная коллегия:

В.А. Бажанов (Ульяновский государственный университет, Ульяновск, Россия), *В.В. Васильев* (Московский государственный университет им. М.В. Ломоносова), *Джон Греко* (Джорджтаунский университет, Вашингтон, США), *И.Н. Грифцова* (Московский педагогический государственный университет, Москва, Россия), *Д.Н. Дроздова* (Национальный исследовательский университет «Высшая школа экономики», Москва, Россия), *М.С. Киселева* (Институт философии РАН, Москва, Россия), *Н.И. Кузнецова* (Институт истории естествознания и техники РАН, Москва, Россия), *Джоан Лич* (Университет Куинсленда, Брисбен, Австралия), *Дженнифер Лэки* (Северо-Западный университет, Чикаго, США), *И.Б. Микиртумов* (Национальный исследовательский университет «Высшая школа экономики» в Санкт-Петербурге, Санкт-Петербург, Россия), *И.Д. Невважай* (Саратовская государственная юридическая академия, Саратов, Россия), *Ханс Позер* (Берлинский технический университет, Берлин, Германия), *В.Н. Порус* (Национальный исследовательский университет «Высшая школа экономики», Москва, Россия), *Данкан Притчард* (Калифорнийский университет в Ирвайне, Ирвайн, США), *Александр Рузер* (Университет Агдера, Агдер, Норвегия), *В.Э. Терехович* (Национальный исследовательский университет «Высшая школа экономики», Москва, Россия), *Тимоти Уильямсон* (Оксфордский университет, Оксфорд, Великобритания), *А.А. Филатова* (Московский физико-технический институт, Москва, Россия), *Стив Фуллер* (Уорикский университет, Ковентри, Великобритания), *Л.В. Шиповалова* (Европейский университет в Санкт-Петербурге, Россия), *Нико Штер* (Университет Цеппелина, Фридрихсхафен, Германия)

Редакционный совет:

В.А. Лекторский (*председатель*, Институт философии РАН, Москва, Россия), *А.А. Гусейнов* (Институт философии РАН, Москва, Россия), *Джон Дюпре* (Эксетерский университет, Эксетер, Великобритания), *Том Рокмор* (Университет Дюкейн, Питтсбург, США; Пекинский университет, Пекин, Китай), *Стивен Тернер* (Университет Южной Флориды, Тампа, США), *Эндрю Финберг* (Университет Саймона Фрезера, Бернаби, Канада), *Дэвид Хесс* (Университет Вандербильта, Нашвилл, США)

Учредитель и издатель: Федеральное государственное бюджетное учреждение науки Институт философии Российской академии наук

Периодичность: 4 раза в год. Выходит с 2004 г.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации СМИ: ПИ № ФС77-57113 от 3 марта 2014 г.

Подписной индекс каталога Урал-Пресс – 46318

Журнал включен в: Перечень рецензируемых научных изданий ВАК (группа научных специальностей 5.7 «Философские науки»); Российский индекс научного цитирования (РИНЦ); Ulrich's Periodicals Directory; ERIH PLUS; Philosophy Documentation Center; Russian Science Citation Index (Web of Science); Web of Science (Core Collection); SCOPUS; ЕГПНИ («Белый список» научных журналов)

Публикуемые материалы прошли процедуру рецензирования и экспертного отбора

Адрес редакции: Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1, оф. 315

Тел.: +7 (495) 697-95-76; e-mail: journal@iphras.ru; сайт: https://journal.iphras.ru

© Институт философии РАН, 2026

ISSN 1811-833X (Print)
ISSN 2311-7133 (Online)

EPISTEMOLOGY & PHILOSOPHY OF SCIENCE

Quarterly peer-reviewed journal
2026. Volume 63. Number 3

Editor-in-Chief: *Ilya T. Kasavin* (RAS Institute of Philosophy, Russia)
Editorial Assistants: *Alexander Yu. Antonovski* (RAS Institute of Philosophy, Russia),
Irina A. Gerasimova (RAS Institute of Philosophy, Russia),
Petr S. Kusliy (University of Göttingen, Göttingen, Germany),
Liana A. Tukhvatulina (RAS Institute of Philosophy, Russia)

Editorial Board:

Valentin A. Bazhanov (Ulyanovsk State University, Russia), *Daria N. Drozdova* (HSE University, Russia), *Asya A. Filatova* (Moscow Institute of Physics and Technology, Russia), *Steve Fuller* (University of Warwick, UK), *John Greco* (Georgetown University, USA), *Irina N. Griftsova* (Moscow Pedagogical State University, Russia), *Marina S. Kiselyova* (RAS Institute of Philosophy, Russia) *Natalia I. Kuznetsova* (RAS Institute for the History of Science and Technology, Russia), *Jennifer Lackey* (Northwestern University, USA), *Joan Leach* (Queensland University, Australia), *Ivan B. Mikirtumov* (HSE University – St. Petersburg, Russia), *Igor D. Nevvazhay* (Saratov State Law Academy, Russia), *Vladimir N. Porus* (HSE University, Russia), *Hans Poser* (Technical University of Berlin, Germany), *Duncan Pritchard* (University of California, Irvine, USA), *Alexander Ruser* (University of Agder, Norway), *Lada V. Shipovalova* (European University at St. Petersburg, Russia), *Nico Stehr* (Zeppelin University, Germany), *Vladislav E. Terekhovich* (HSE University, Russia), *Vadim V. Vasilyev* (Lomonosov Moscow State University, Russia), *Timothy Williamson* (Oxford University, UK)

Editorial Council:

Vladislav A. Lektorsky (the chairman) (RAS Institute of Philosophy, Russia), *John Dupré* (University of Exeter, UK), *Andrew Feenberg* (Simon Fraser University, Canada), *Abdusalam A. Guseinov* (RAS Institute of Philosophy, Russia), *David Hess* (Vanderbilt University, USA), *Tom Rockmore* (Duquesne University, USA; Peking University, China), *Stephen Turner* (University of South Florida, Tampa, USA)

Publisher: Institute of Philosophy, Russian Academy of Sciences

Frequency: 4 times per year. First issue: 2004

The journal is registered with the Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskomnadzor). The Mass Media Registration Certificate No. FS77-57113 on March 3, 2014

Subscription index in the catalogue of Ural-Press is 46318

Abstracting and Indexing: the list of peer-reviewed scientific edition acknowledged by the Higher Attestation Commission of the Ministry of Education and Science of the Russian Federation; Ulrich's Periodicals Directory; ERIH PLUS; Philosophy Documentation Center; Russian Science Citation Index (Web of Science); Web of Science (Core Collection); SCOPUS

All materials published in the "Epistemology & Philosophy of Science" undergo peer review process

Editorial address: 12/1 Goncharnaya St., Moscow, 109240, Russian Federation

Tel.: +7 (495) 697-95-76; e-mail: journal@iphras.ru; website: https://journal.iphras.ru

© Institute of Philosophy, Russian Academy of Sciences, 2026

CONTENTS



EDITORIAL

Natalia I. Kuznetsova, Lada V. Shipovalova. On the Usefulness and Harmfulness of History for Contemporary Science.....6



PANEL DISCUSSION

Harry Collins. The Bath School and the Strong Programme.....22

Charles Thorpe. An Invitation to Harry Collins' Humanistic Sociology.....33

Tatiana D. Sokolova. Empirical Expansion and Philosophical Autonomy: Reconsidering the Strong Programme through Historical Epistemology.....45

Asya A. Filatova. AI and the Transformation of Scientific Knowledge: Hybrid Expertise and the Limits of Collins's Theory.....56

Harry Collins. Reply to Critics.....67



EPISTEMOLOGY & COGNITION

Olga E. Stoliarova. Background Knowledge in Scientific and Technological Practices: The Dialectic of Determination and Freedom.....78

Konstantin G. Frolov. De-anthropocentrism as a Conceptual Alternative to Anthropomorphization in the Epistemology of AI.....94



LANGUAGE & MIND

Pavel V. Grishin. Towards New Approaches in Machine Consciousness Research.....113



VISTA

Petr D. Abramov. D. Bohm's Realism and Holism: The System of Metaphors and Differences from the Lines of N. Bohr and W. Heisenberg.....132



CASE STUDIES – SCIENCE STUDIES

Daniil Yu. Turkenich. Scientific and Technological Utopianism as the Way of Thinking of Science Alternatives. Notes towards the "Utopian Method".....149



INTERDISCIPLINARY STUDIES

Anna V. Sakharova. Objectivity in the Mirror of Language: From Linguistic Analysis to Epistemological Interpretations.....168



ARCHIVE

Maxim V. Vinarski. Adaptive Response: The Reception of F. Engels's Views on Heredity and Anthropogenesis by Soviet Biologists (1920s–1950s).....189



NEW TRENDS

Ulyana P. Belyaeva. Thinking with Machines: Automatism, Autonomy, and the Historical Epistemology of Intelligence.....207

Anastasia A. Lisenkova, Darya S. Bilyeva. Science in the Digital Age: How AI Is Changing the Understanding of Scientific Knowledge and Its Creation.....220

СОДЕРЖАНИЕ



РЕДАКЦИОННАЯ СТАТЬЯ

Наталия Кузнецова, Лада Шиповалова. О пользе и вреде истории для современной науки.....6



ПАНЕЛЬНАЯ ДИСКУССИЯ

Гарри Коллинз. Батская школа и «Сильная программа».....22

Чарльз Торп. Приглашение к гуманистической социологии Гарри Коллинза.....33

Татьяна Соколова. Эмпирическая экспансия и философская автономия: пересмотр “Сильной программы” в свете исторической эпистемологии.....45

Ася Филатова. Искусственный интеллект и трансформация научного знания: гибридная экспертиза и границы теории Г. Коллинза.....56

Гарри Коллинз. Ответ оппонентам.....67



ЭПИСТЕМОЛОГИЯ И ПОЗНАНИЕ

Ольга Столярова. Фоновое знание в научно-технических практиках: диалектика детерминации и свободы.....78

Константин Фролов. Де-антропоцентризм как концептуальная альтернатива антропоморфизации в эпистемологии искусственного интеллекта.....94



ЯЗЫК И СОЗНАНИЕ

Павел Гришин. К новым подходам в исследовании машинного сознания.....113



ПЕРСПЕКТИВА

Петр Абрамов. Реализм и холизм Дэвида Бома: система метафор и отличия от линий Н. Бора и В. Гейзенберга.....132



СИТУАЦИОННЫЕ ИССЛЕДОВАНИЯ

Даниил Туркенич. Научно-технологический утопизм как способ мышления альтернатив науки. Заметки об «утопическом методе».....149



МЕЖДИСЦИПЛИНАРНЫЕ ИССЛЕДОВАНИЯ

Анна Сахарова. Объективность в зеркале языка: от лингвистического анализа к эпистемологическим интерпретациям.....168



АРХИВ

Максим Винарский. Адаптивный ответ: рецепция взглядов Ф. Энгельса на наследственность и антропогенез советскими биологами (1920–1950-е гг.).....189



ТЕНДЕНЦИИ

Ульяна Беляева. Мыслить с машинами: автоматизм, автономия и историческая эпистемология интеллекта.....207

Анастасия Лисенкова, Дарья Быльева. Наука в цифровую эпоху: как ИИ меняет представление о научном знании и его создании.....220

Эпистемология и философия науки
2026. Т. 63. № 3. С. 6–21
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 6–21
DOI: https://doi.org/10.5840/eps202663336

О ПОЛЬЗЕ И ВРЕДЕ ИСТОРИИ ДЛЯ СОВРЕМЕННОЙ НАУКИ

Кузнецова Наталия Ивановна – доктор философских наук, главный научный сотрудник.
Институт истории естествознания и техники им. С.И. Вавилова РАН.
Российская Федерация, 125315, г. Москва, ул. Балтийская, д. 14;
e-mail: cap-cap@mail.ru

Шиповалова Лада Владимировна – доктор философских наук, профессор.
Центр практической философии «Стасис», Европейский университет в Санкт-Петербурге.
Российская Федерация, 191187, г. Санкт-Петербург, ул. Гагаринская, д. 6/1А;
e-mail: shipovalova@eu.spb.ru

В статье рассматривается значение историко-научных исследований для современной науки. Авторы отталкиваются от двух контекстов проблематичной значимости истории науки – институционально-образовательного и социально-эпистемологического. В первой части рассматриваются три модуса отношения истории и (научной) жизни с опорой на идеи Ф. Ницше. Их эпистемологическое применение позволяет проблематизировать однозначный ответ на вопрос о пользе и вреде истории для современной науки. Во второй части авторы обращаются к уже существующему различию подходов антикваризма и презентизма, отмечают недостаточность презентистского подхода, раскрывают отличие эпистемологического антикваризма от философского, а также определяют, чем может быть полезна презентистская и антикваристская историография как в социально коммуникативном (для формирования и поддержки сообщества), так и в эпистемологическом смысле (для развития науки). В завершение дается ответ на вопрос об условии пересечения этих двух подходов, а также предлагается рассматривать недоопределенность прошлого науки как их общий исток. Делается общий вывод о значении работы историка науки для современных научных исследований.

Ключевые слова: история науки, историческая эпистемология, презентизм, антикваризм, недоопределенность





ON THE USEFULNESS AND HARMFULNESS OF HISTORY FOR CONTEMPORARY SCIENCE

Natalia I. Kuznetsova – DSc in Philosophy, Chief Research Fellow.
S.I. Vavilov Institute for the History of Science and Technology, Russian Academy of Sciences.

The article examines the significance of historical and scientific research for contemporary science. The authors start from two contexts of the problematic relevance of the history of science – the institutional-educational and the socio-epistemological. The first part examines three modes of the relationship between history and (scientific) life, drawing on the ideas of F. Nietzsche. Their epistemological application makes it possible

© Кузнецова Н.И., 2026
© Шиповалова Л.В., 2026



14 Baltiyskaya St., Moscow
125315, Russian Federation;
e-mail: cap-cap@mail.ru

Lada V. Shipovalova –
DSc in Philosophy, Professor.
Stasis Center for Practical Philosophy, European University at St. Petersburg.
6/1 A Gagarinskaya St.,
Saint Petersburg 191187,
Russian Federation;
e-mail: shipovalova@eu.spb.ru

to problematize any unequivocal answer to the question of the usefulness and harmfulness of history for contemporary science. In the second part, the authors turn to the existing distinction between the antiquarian and presentist approaches, note the insufficiency of the presentist approach, reveal the difference between epistemological antiquarianism and philosophical antiquarianism, and also determine how presentist and antiquarian historiography can be useful both in the socio-communicative sense (for shaping and supporting the community) and in the epistemological sense (for the development of science). In conclusion, an answer is given to the question of the condition for the intersection of these two approaches, and it is also proposed that the underdetermination of the past of science be regarded as their common source. A general conclusion is drawn about the significance of the work of the historian of science for contemporary scientific research.

Keywords: history of science, historical epistemology, presentism, antiquarianism, underdetermination

Введение

Вопрос о пользе и вреде историко-научных исследований для современной науки возникает сразу из нескольких контекстов. Первый контекст – институциональный и по преимуществу образовательный. Скоро будет 20 лет введению в высших учебных заведениях России курса «История и философия науки» для подготовки аспирантов и соискателей любых научных специальностей. Это был далеко не первый опыт в международном пространстве. Более того, о конструктивной совместности этих направлений исследования, пересекающихся на науке, свидетельствует и существование соответствующих журналов и обществ. Все это говорит о признании значения истории науки для будущих ученых, а также о связи исторической и философской рефлексии оснований научных исследований. Однако каковы основания этого признания? Не следует ли рассматривать его в качестве случайного? Нельзя ли историю науки в научном образовании просто взять и отменить? Необходима ли история науки современным ученым? Может быть, история науки – лишь предмет чистого незаинтересованного любопытства эпистемологов и историков науки, а будущего ученого, особенно не гуманитария, она только отвлекает от существа его работы? Стоит ли учить истории науки и если да, то какой?

Второй – социально-эпистемологический контекст относится к осмыслению практик научного исследования, связанных (или нет) с историей науки. Можно заметить, что востребованность истории при этом может быть истолкована амбивалентно. С одной стороны, речь идет о том, что формальная и неформальная научная коммуникация, присутствие сформированного, поддерживаемого

и развивающегося научного сообщества могут осуществляться и обосновываться не только в синхроническом, но и в диахроническом контексте. Историческая память науки может служить образованию научного коллектива (Л. Флек), историческая преемственность – формированию стиля научного мышления (А. Кромби, Я. Хакинг), подкреплению консенсуса относительно концептуальных и методологических оснований исследования и определенности научных фактов. С другой стороны, так ли однозначна роль научной коллективной памяти? Не может ли она, напротив, служить догматизации идей и препятствовать прогрессу? Не будет ли парадигмальный консенсус, достигаемый на основе обращения к истории науки, поддерживать консерватизм научного сообщества (о чем говорил Т. Кун) и выступать преградой на пути появления и утверждения научной новизны?

Таким образом, вопрос о значении истории науки может получить ответ через уточнение ее роли и определение ее пользы и вреда. Сама эта роль может быть истолкована в социально-эпистемологическом смысле. При этом можно допустить, что ее социально-коммуникативный аспект, связанный с формированием, поддержкой или «пересборкой» научного сообщества, может и противоречить аспекту эпистемологическому, связанному с развитием научного знания[1].

Допустим, что ответ на поставленные вопросы можно дать опосредованно, через философско-научные и эпистемологические исследования, связанные с историей науки и нуждающиеся в ней в том или ином смысле. Однако связь эта также представляется неоднозначной. Отношения истории и философии науки в контексте установок программы И. Лакатоса уже давно критикуются как неконструктивные, противоречащие возможностям междисциплинарного диалога исследователей науки [Riesch, 2014]. Эти установки рассматривают историю науки по преимуществу в качестве поставщика материала для эпистемологической рефлексии, материала, призванного подтвердить философские тезисы. При этом дело историка науки в его автономии оказывается недооцененным в философии науки, а его значение для современной науки остается непроясненным. Такой же вывод возможен и относительно альтернативных предложений представителей современной исторической эпистемологии, рассматривающих историю как мотивирующую воображение философа, провоцирующую его на постановку проблем [Rheinberger, 2012]. В этом контексте дело историка науки оказывается случай-

[1] Это различие может быть соотнесено с различием и комплементарностью двух контекстов определения науки – предметного и социального [Антоновский, 2024].



ным, не обязательным, потенциально замещаемым исследовательским вниманием к событиям современной науки.

Более того, если предположить, что оправдание значения истории для науки может быть произведено апелляцией к историзации философии науки, то и здесь картина далека от определенности. Так, еще в первой половине XX в. Э. Гуссерль ставит в упрек предшествующей философской теории познания ее неисторический характер. В то же время отцы-основатели традиции аналитической философии провозглашают ее отказ от историчности, защищая тем самым собственное знание от релятивизма и отстаивая его научный характер.

Иначе говоря, апелляция к философии как таковой в решении вопроса о значении историко-научных исследований для современной науки может лишь усложнить нашу задачу и будет оставлена на перспективу. Хотя нельзя не признать, что наше рассуждение оказывается по необходимости философским. И первым его шагом будет обращение к тексту, очевидность которого следует из названия статьи, а релевантность обеспечивается предлагаемым эпистемологическим применением.

История и жизнь: философская рефлексия и ее эпистемологическое применение

Ф. Ницше в своих знаменитых *Несвоевременных размышлениях* раскрывает три способа отношения истории и жизни, которые могут приносить пользу и вред, оказываясь релевантными в различных обстоятельствах. Ницше пишет: «Если человек, желающий создать нечто великое, вообще нуждается в прошлом, то он овладевает им при помощи монументальной истории; кто, напротив, желает оставаться в пределах привычного и освященного преданием, тот смотрит на прошлое глазами историка антиквария, и только тот, чью грудь теснит забота о нуждах настоящего, и кто задался целью сбросить с себя какою бы то ни было ценою угнетающую его тяжесть, чувствует потребность в критической, то есть судящей и осуждающей истории» [Ницше, 1990, с. 174].

В первом случае *монументальной истории науки* в поле внимания историка и ученого попадает только верифицирующее современность прошлое. Только оно оказывается полезным настоящему, служа ему основанием (монументом). Однако неизбежная избирательность позиции настоящего может приносить вред познанию, в том числе научному. Вот как говорит о соответствующем опыте Ницше: «…пока прошлое будет изображаться как нечто, достойное подражания и как доступное подражанию и могущее

повториться еще раз, – до тех пор истории, конечно, грозит опасность подвергнуться некоторому искажению, приукрашиванию и, в силу этого, сближению со свободным вымыслом» [Ницше, 1990, с. 172].

Монументальная история мотивирована, по Ницше, теоретической направленностью в великое будущее, сопровождающейся практическим стремлением его создавать. В этом смысле она может быть соотнесена с прогрессистской вигской историографией, раскритикованной еще Г. Баттерфилдом в его известном эссе 1931 г. *The Wig Interpretation of History* [Butterfield, 1964]. Такая критика представляется для современного гражданского историка почти само собой разумеющейся, так же как и ницшеанская критика гегелевского историзма, поскольку рефлексия исторического познания выявляет очевидность отличия прошлого от настоящего, а также ограничения, связанные с «прочтением» истории в контексте целей текущего момента. Однако неслучайный эпистемологический казус Баттерфилда состоит в том, что сам он, спустя несколько лет, пишет фундаментальную работу о генезисе Новоевропейской науки (*The Origins of Modern Science*), которая трактуется современными эпистемологами как принципиально вигская [Alvargonzalez, 2013]. Это конкретное обстоятельство, признание в целом неустранимости позиции настоящего из исследования прошлого, которое к этому настоящему привело, а также последующее внимание к стратегиям работы историков науки актуализируют современные дискуссии среди исторических эпистемологов.

С одной стороны, они признают, подобно Альваргонзалесу, неустранимость тенденции прогрессистской историографии из внимания к научному прошлому. Польза от такой истории науки состоит как минимум в том, что подчинение прошлого нуждам настоящего способствует поддержке целостности научного сообщества и легитимации современной науки в публичном пространстве. Верное единому Великому Прошлому сообщество может обоснованно надеяться на успехи в настоящем и будущем! Другими словами, польза монументальной, или вигской, истории науки может быть понята как основа сохранения и общественного признания парадигмы в куновском смысле. С другой стороны, исторические эпистемологи критикуют такой подход, указывая на его конкретные эпистемологические следствия, проявляющиеся в том числе в практической сфере. Один из самых ярких примеров такого негативного влияния склонности к вигизму (или к монументализму) в истории науки – осмысление экономического кризиса 2008 г. и неспособности экономистов его предсказать: факт, широко обсуждавшийся в публицистике и в научной литературе. В качестве одной из причин, обусловивших такую неспособность, называется неадекватное историческое образование экономистов, предполагающее гомогенность



исторического процесса и, как следствие, их лишенность широты исторической ретроспективы и отсутствие внимания к многообразию научных подходов [Freeman et al., 2014]. Таким образом, польза монументальной истории в ее применении к истории науки – социально-коммуникативная, а ее возможный вред – эпистемологический и, опосредованно, практический.

*Историк-антиквари*й, согласно Ницше, обладает важными интеллектуальными способностями, в том числе знаниями языков, позволяющими правильно читать и интерпретировать традицию. Перед ним, способным к «проникновению в сокровенный смысл событий» [Ницше, 1990, с. 173], опыт прошлого раскрывается в его аутентичности. Однако ему грозит одна существенная опасность и соответствующий ей вред – уклониться в ситуацию, когда «живая современная жизнь перестает его одухотворять». Тогда благоговейное отношение к Прошлому может замениться «отвратительным зрелищем слепой страсти к собиранию фактов, неутомимого накапливания того, что когда-либо существовало» [Там же, с. 177]. Кажется, что такая история науки, лишенная связи с современностью, не может быть полезной для ученого и может служить в лучшем случае обеспечению его случайного незаинтересованного любопытства. Хотя нельзя не заметить, что историк посредством такой работы с научным прошлым может обеспечить бо́льшую полноту его представления, нарисовать его детальную картину, справившись отчасти с угрозами, которые несет кумулятивная монументальная историография для современного научного познания. Проще говоря, ницшеанский антикварий в области истории науки может способствовать минимизации вреда в эпистемологическом смысле, но может ли он надеяться на пользу от своих исследований? Не вполне понятно, можно ли обосновать или хотя бы продемонстрировать позитивность таких исследований – как для самой современной науки, так и для ее общественной легитимации?

Критический историк также не избегает испытания ницшеанским подозрением. Ему свойственна активность и актуальность, отсутствие угрозы погрузиться в «пыль библиографических мелочей» и остаться без всякой связи с текущими проблемами. Однако стремление изменить настоящее может привести к ситуации, когда «подступают с ножом к его корням и жестоко попирают все святыни» [Там же, с. 178]. Критический историк рискует упустить неизбежную связь настоящего, в том числе научного, с прошлым, редуцировать наш статус наследников прежних поколений, обеспеченный, порой незаметно, их языком, накопленным опытом и инструментарием, но при этом обремененный их заблуждениями и страстями. Критический историк разделяет ограниченность и слепоту с другими модальностями отношения истории и жизни. Слепота антиквария относительно нужд настоящего дополняется слепотой относительно

полноты прошлого, из которого монументальный и критический историки выбирают лишь то, что для науки может быть полезно в будущем, ограничивая свое представление или ради утверждения, или ради изменения настоящего. Критический историк спасает не всех «проигравших», а только тех, кто оказывается его союзником на пути определенным образом понятого освобождения. Практика критического историка в науке, с одной стороны, может быть востребована в ситуации смены парадигм, когда прежние образцы деятельности во многом остаются лишь формальной данью уважения к «гигантам», на плечах которых построен монумент настоящего. Так, историки науки отмечают, что в географии куновская идея смены парадигм повлияла на работу по обоснованию научной новизны, а также на самовосприятие географов как новаторов дисциплины, несущих в нее революционные преобразования [Meeteren, Sidaway, 2020]. Однако очевидно, что рискованным предположениям новых парадигмальных положений не гарантирован успех в обосновании. Последнее требует серьезной работы по привлечению союзников (Б. Латур), работы над признанием, которого еще нет. Таким образом, социально-коммуникативная и эпистемологическая польза практики критического историка гипотетична.

Итак, три модуса отношения истории и (научной) жизни заостряют вопрос о пользе и вреде историко-научных исследований для современной науки. Двойственность ответа – социально-коммуникативная польза и эпистемологический вред – мы наблюдаем в случае монументальной историографии науки, лишь негативную пользу находим в антикварной истории, довольствуясь гипотетической полезностью работы критического историка. Такое проблематичное положение дел мотивирует дальнейший поиск и обращение к современной эпистемологической рефлексии истории науки.

История и современная наука: эпистемологический анализ и его ограничения

В 1965 г. Дж. Стокинг-мл., главный редактор журнала «История наук о поведении», известный американский антрополог, описал методологическую дилемму (альтернативные ориентации, по его выражению), стоящую перед историками науки [Stocking, 1965]. Он четко зафиксировал различие целеполагания двух подходов и, по сути, ввел термин «презентизм» в широкий культурный обиход. Один подход – презентистский – стремится «понять прошлое ради настоящего»; другой – «понять прошлое ради самого прошлого».



Современная историческая эпистемология[2], обращаясь к стратегиям историко-научного исследования, признает относительную полезность и даже неизбежность написания истории прошлой науки с позиции науки современной. При этом она следует тому именованию, которое было предложено Стокингом. Нельзя не отметить различия, существующие между историко-научным и общим историографическим презентизмом[3]. Историк науки – презентист в отличие от антрополога, также учитывающего актуальность настоящего, видит каузальную связь настоящего с прошлыми событиями, что позволяет ему выбирать значимые (для современной науки) объекты для анализа. Однако в отличие от своего коллеги, гражданского историка, историк науки знает, «как можно было бы избежать ошибки» в прошлом, потому что перед ним раскрывается не только каузальность, в контексте которой настоящее – следствие прошлого, но и нормативный характер настоящего, в контексте которого прошлые научные находки – несовершенные предшественники завершенных современных исследований [Loison, 2016]. Историк науки – презентист работает в духе монументализма, заботясь о поиске предшественников или «зародышей» важнейших идей [Puras, 2014; Adrian, 2015; Кузнецова, Розов, 1996]. Вигская историография при этом рассматривается как вариант или как опасная радикализация презентизма, акцентирующая внимание на научном прогрессе, направленном из прошлого через настоящее в будущее [Chang, 2021], трактующая современную науку не только как следствие, но и как цель науки прошлой [Loison, 2016].

Относительная полезность презентизма для современной науки реализуется за счет выстроенной ретроспективым образом истории в ее дисциплинирующей и легитимирующей функции. Полезность может быть дополнена и эпистемологической функцией. Работа историка науки – презентиста служит созданию «базы данных» для современных исследований, что очевидно, например, в вулканологии, географии [Meeteren, Sidaway, 2020], или позволяет подтвердить инвариантность субстрата объекта исследования, сохраняющегося в определенном смысле «тем же самым» в прошлом и настоящем, если речь идет, например, о видах в науках о живом. Необходимым

[2] Используемый нами термин имеет в современности как минимум два смысла: эпистемологической рефлексии базовых понятий и подходов истории как науки, а также критики базовых понятий и подходов эпистемологии посредством их историзации. Представляется, что история науки оказывается на пересечении этих двух версий исторической эпистемологии и потому ее рефлексивная работа имеет особую значимость как для исторической науки, так и для теории познания. Также об актуальных смыслах исторической эпистемологии см.: [Касавин, 2020; Порус, 2021].

[3] Об общем историографическом презентизме см., например: [Wilson, Ashplant, 1988].

условием такой пользы служит предпосылка о том, что наше познание никак не влияет на познаваемый объект, что характерно для радикальной реалистической установки и действенно по преимуществу в естествознании [Loison, 2016]. Относительность пользы определяется в том числе границами применения такого условия. За этими границами – социальные науки с их историей, связанной с изменяемостью предметного поля в процессе познания. За этими границами – конструктивистская установка, признающая активность познающего субъекта, определенного историческим, социально-политическим, психологическим и иным «внешним» контекстом.

Как можно «усилить» пользу историко-научного презентизма в социально-коммуникативном и эпистемологическом смысле, предполагающем не только подтверждение и обоснование существующего знания, но и его развитие? Для этого должно быть реализовано смягчение презентистской установки, позволяющее увидеть настоящее не гомогенным и, соответственно, допускающее сложность, вариативность прошлого, к нему ведущего. В этом контексте становятся актуальными такие практики историко-научного исследования, которые мы предлагаем назвать «вариативный презентизм», предполагающие одновременное существование различных линейных историй. Можно привести два примера таких исследований.

Первый пример принадлежит вариативному презентизму в истории социологии. Известные социологи современности, рассуждая об идентичности собственной дисциплины и определяя ее различным образом, выстраивают три линейные истории, связывая каждую с Робертом Оуэном как предшественником. В одном случае он оказывается основателем современной науки об обществе, исходя из представления о современной социологии как публичной, активно включающейся в решение общественных проблем. Во втором случае на основании современных представлений о социологии как по преимуществу теоретической деятельности Р. Оуэн, будучи, очевидно, практиком, относится к значимой, но все же к предыстории. В третьем – Оуэн представляется классическим авторитетом, обосновывающим альтернативный марксизму проект понимания общества и «помогающим» современным социологам защищать антимарксистские установки [Puras, 2014]. Такой вариативный презентизм, строящийся на основании негомогенного настоящего, использующего (одно и то же) прошлое в различных интересах, оказывается далек от возможной монументализации. Он может быть полезен не только для собирания сторонников той или иной позиции, но, потенциально, для понимания сложности и противоречивости современного социологического знания и, как следствие, для развития, которому эти противоречия могут служить источником.

Второй пример относится к вариативному презентизму в истории психологической науки, которая в рамках традиционного



презентизма возводит свои основания к работам Вундта [Adrian, 2015]. Однако, если принять в расчет то, что человек как предмет психологии культурно и социально дифференцирован, практическая деятельность психолога и ее теоретические основания не могут не учитывать соответствующее различие интеллектуальных традиций. В этом контексте оправданно искать и находить истоки психологической науки не только в Античности, в работах Аристотеля, служащих началом европейской линии, но в арабской или индийской мысли. Такой ход к вариативному презентизму может также служить развитию современной науки, если, учитывая гетерогенность современного объекта исследования, будет поставлен вопрос о пересечении и связи в современности каузальных последовательностей и соответствующих традиций.

Однако что может мотивировать актуализацию такого вариативного презентизма? Что может заставить видеть прошлое сложным, принципиально отличающимся от его однозначной традиционной презентистской трактовки? Вигская историография представляет собой негативный полюс, от которого отталкивается исследование, обнаруживая опасности и вред презентизма. Но есть ли позитивный полюс? Существует ли практика историко-научного исследования, ориентированная на автономию прошлого, выводящая его за границы нормативного и даже вечного характера настоящего?

Отчасти взгляд на автономное научное прошлое присутствует в раскрытии образа нововременной науки как того, что «не было» экспериментальным математическим естествознанием (Б. Латур), предметом которого является «бесконечная Вселенная» (А. Койре). Отчасти работа по «освобождению» прошлого ведется относительно образа научной революции, который раскрывается в системе контекстов, в пространстве пересечения различных интересов участников, раскрывается как событие, имеющее не только и не столько облик радикального разрыва с донаучным периодом истории, к которому мы привыкли по кратким учебникам по истории науки [Shapin, 1996; Schuster, 2013; Дмитриев, 2005]. Такой освобождающий подход применяется и к средневековой научности, в которой могут быть обнаружены отличные от современности стили мышления, впрочем, переводимые на язык современности с осознанием жертвы, приносимой успешности перевода [Рабинович, 1979]. Для такого освобождающего прошлое подхода в отечественной исторической эпистемологии есть термин «антикваризм», впрочем, уже известный нам из контекста философской рефлексии, однако наполняемый дополнительным существенным смыслом[4].

[4] Для описания подобных историко-технологических исследований в англоязычной традиции используется словосочетание «историцистская чувствительность» [Soden et al., 2021].

Подчеркнем это важное дополнение. Антикваризм в смысле, раскрываемом Ницше, предполагает знание прошлого самого по себе, эмпатическое отождествление с ним. Начиная с осознания границы между временами, историк-антикварий достигает того, что завеса, отделяющая его от прошлого, падает. Именно такая понимающая связь с прошлым может привести к опасной утрате связи с современностью. Антикварная и монументальная ницшеанские истории ищут понимания, достигаемого в первом случае посредством погружения в прошлое, во втором – включения прошлого в настоящее. Напротив, подход антикваризма в отечественной рефлексии истории науки предстает как принципиально объясняющий, формирующий определенную картину научного прошлого, наряду и в оппозиции картине презентизма. В рамках этой оппозиции, которая была в свое время охарактеризована М.А. Розовым [Розов, 1994] как дополнительная в смысле квантово-механической интерпретации Н. Бора, презентистский подход предполагает установление значения прошлого для науки настоящего, возможность его использования, *понимающую ассимиляцию* его в современное знание. Причем для такого использования необходимо соблюдение двух условий – однозначная интерпретация содержания знания (научного открытия) и выражение его в терминах современного научного языка. Антикваристский же подход *объясняет* становление событий, идей и концепций прошлого, раскрывает их происхождение и использование в их собственном историко-научном и историко-культурном контексте, оставляя способы современного использования прошлого научного знания вниманию презентиста. Антикваристский подход делает видимым разрыв в истории науки, отличие прошлого от современности. Презентистский – актуализирует связь, акцентируя внимание на современном субъекте, удерживающем эту связь. Два подхода – объясняющий и понимающий – различены и дисциплинарно. Первый реализует чаще всего историк и философ культуры. Второй – специалист в области современной науки и ее истории. Отсутствие внимания к этому принципиальному различию – методологическому и дисциплинарному – приводит к тому, что современные исследователи вписывают подход антикваризма в единую с презентизмом классификацию [Кузин, 2020].

Историки науки, использующие эти подходы, обращаются к прошлому из современности и ради ее пользы: презентист – ради установления и поддержки современного смысла, антикварий – ради уточнения прошлого значения, иначе не очевидного из современности. Работая с прошлым, историк-презентист скажет, что сделал Колумб – открыл Америку, а историк-антикварий опишет, что делал Колумб, как он воспринимал задачи своего путешествия и оценивал его результаты [Кузнецова, 2009]. Историк науки – презентист скажет, что стремление к объективности разделяли философы и ученые, начиная как минимум с начала Нового времени. Он сможет найти



и у Ф. Бэкона, несмотря на отсутствие такого слова, соответствующий современному термину объективности смысл [Zagorin, 2001]. Историк науки – антикварий же уточнит, что ученые (и философы) не употребляли термин объективности в современном нам смысле до XIX в. и это имеет значение, потому что именно в этот период объективность складывается как ценность научного исследования и добродетель ученого, что сопровождает утверждение научной деятельности как профессии [Daston, Galison, 2007].

Будучи полезным в научно-коммуникативном и эпистемологическом смысле, оставляет ли какую-нибудь пользу презентизм так понятому антикваризму? Если проводить параллель между куновским видением истории смены парадигм и антикваристской историографией, мы поймем, что границы современного научного сообщества, несовершенство и относительность его парадигмы видны именно тогда, когда реализуется практика антиквария. Антикваристская историография науки допускает возможность альтернатив и пересборок эпистемологических оснований и коммуникативных научных практик. Есть очевидный пример пользы антикваризма в этом контексте: решимость Коперника публиковать свою работу была обеспечена в том числе усилиями историка-антиквария, сохранившего дискуссии о неоднозначном положении Земли в Космосе в античной философии. В такой своей интенции антикваризм оказывается полезен развитию науки, сохраняя для него альтернативные, на время оставленные в стороне и отчасти фальсифицированные позиции, которые он, повторим, не просто описывает, но объясняет и тем самым сохраняет их возможную значимость! Он оказывается полезен и в качестве основания пересборки научного сообщества.

Таким образом, мы получаем конкретизацию того, чем может быть полезна презентистская и антикваристская историография как в социально коммуникативном (для формирования, поддержки или пересборки сообщества), так и в эпистемологическом смысле (для развития науки). Для презентизма польза обусловлена вниманием, помимо прочего, к вариативности линейных историографий, провоцирующих вопрос о том, как такая вариативность возможна в одной науке, и мотивирующих работу над пересечением перспектив. Для антикваризма она связана с сохранением внимания к некумулятивности научной истории, к объяснению прошлого, отличного от прошлого современной науки, объяснению, которое делается *для* современной науки.

Остается еще один вопрос, требующий ответа и оставляющий сомнения по поводу дополнительности презентизма и антикваризма как двух картин прошлого. Как разным историкам понимать друг друга, как им учить ученых одновременно антикваризму и презентизму? Не будет ли происходить всякий раз уклонение в одну или другую сторону, не будет ли случаться забвение другой стороны,

«ущемление прав» другого и догматизация собственного подхода? Этот вопрос имеет не только образовательный, но и социально-эпистемологический характер, он относится к возможной полноте историко-научной пользы, и подступом к ответу на него может служить предположение возможности перевода, пересечения картин, работы на их границах.

Заключение. О работе на границах истории науки

Историко-научные картины антиквария и презентиста могут пересекаться, если у них есть основание для этого, однако прошлое как объект исследования в двух случаях различно! Что же их может сближать? Создавая по-разному картины прошлого, историки – антиквариста и презентист – стремятся прошлое определить: описать его в значимом отношении к современности или объяснить, погрузив в многообразие прошлых контекстов. Различаясь в объекте, антиквариста и презентист пересекаются в мотиве. *Недоопределенность события прошлого науки* будет их общим исследовательским истоком. Внимание к ней не позволяет считать прошлое тем, что должно быть однозначно объяснено антикварием, забыто в случае несоответствия современности или поднято на флаг в случае соответствия презентистом. Соответственно, все то, что заставляет эту недоопределенность заметить, подчеркнуть с позиции автора одной или другой картины прошлого – следует признать инструментом, актуализирующим возможность перевода. Среди таких инструментов, например, проблематизация гомогенности образа современной науки, раскрытие неоднозначности ее предшественников, описание нелинейного и негарантированного процесса генезиса научных объектов[5].

Между двумя дополняющими друг друга картинами прошлого должна быть возможной работа перевода (подобно тому, как в физике существует возможность изменения экспериментальной ситуации). Эту работу можно осуществить, ее результаты предъявить, но для нее самой нельзя создать картину. Она всегда оставляет что-то неявным, как и любой перевод. На пересечении этих картин находятся

[5] Так, Х.-Й. Райнбергер в своей «биографии научного объекта» описывает сложный и негарантированный путь становления РНК – в различных лабораторных практиках, посредством различных приборов, в разных науках и с разными именами, в спорах и дискуссиях постепенно формируется то, что сейчас считается научным фактом [Rheinberger, 2020].



двойственные объекты[6] или исторические события[7], которые могут быть истолкованы в качестве пограничных объектов (С. Стар и Дж. Гризмер), условий возможного перевода. Поиск таких пограничных объектов, а также их двойное историческое обоснование: объясняющее для антиквария и понимающее для презентиста с их различными дисциплинарными установками – дело историка в его полезности для современной науки. Ведь наука в своем генезисе, в своем развитии и в своей легитимации именно с такими объектами и работает, дискутируя об их значении, задавая его разными способами интерпретации. Историк науки – антиквариий и презентист – сопровождают эту работу, раскрывая и обосновывая ее исток и «конечный» пункт.

Список литературы

Антоновский, 2024 – *Антоновский А.Ю.* Наука в суперпозиции. К коммуникативной семантике понятия науки // Эпистемология и философия науки. 2024. Т. 61. № 4. С. 6–24.

Дмитриев, 2005 – *Дмитриев И.С.* Искушение святого Коперника. Ненаучные корни научной революции. СПб.: Изд-во СПбГУ, 2005. 278 с.

Касавин, 2020 – *Касавин И.Т.* Знание и реальность в исторической эпистемологии // Эпистемология и философия науки. 2020. Т. 57. № 2. С. 6–19.

Кузин, 2020 – *Кузин И.А.* Классификация форм презентизма и антипрезентизма в историографии науки // Наука как общественное благо: сборник научных статей / Науч. ред. и сост. И.Т. Касавина, Л.В. Шиповалова. В 7 т. Т. 1. М.: Русское общество истории и философии науки, 2020. С. 74–77.

Кузнецова, 2009 – *Кузнецова Н.И.* Презентизм и антикваризм – две картины прошлого // Arbor Mundi. 2009. Вып. 15. С. 164–196.

Кузнецова, Розов, 1996 – *Кузнецова Н.И., Розов М.А.* История науки на распутье // Вопросы истории естествознания и техники. 1996. № 1. С. 3–18.

Ницше, 1990 – *Ницше Ф.* О пользе и вреде истории для жизни. Соч.: в 2 т. Т. 1. М.: Мысль, 1990. С. 158–230.

Порус, 2021 – *Порус В.Н.* Историческая эпистемология – триггер реформы философии познания // Вопросы философии. 2021. № 5. С. 47–57. https://doi.org/10.21146/0042-8744-2021-5-47-57

Рабинович, 1979 – *Рабинович В.Л.* Алхимия как феномен средневековой культуры: Гермес трижды Величайший. Изумрудная скрижаль. М.: Наука, 1979.

6 Например, скорость в работах Галилея, которая может быть понята и как объективно существующее качество движущегося тела на языке средневековой научности, и как измеримая через отношение пространства и времени величина на языке науки современной [Розов, 1994].

7 Например, географический путь, который может быть понят и как путешествие Колумба в Западную Индию, и как открытие Америки [Кузнецова, 2009].

Розов, 1994 – *Розов М.А.* Презентизм и антикваризм. Две картины истории // Вопросы истории естествознания и техники. 1994. № 3. С. 13–23.

References

Adrian, 2015 – Adrian, C. “Brock Presentism and Diversity in the History of Psychology”, *Psychological Studies*, 2015, vol. 60 (4), pp. 373–379. https://doi.org/10.1007/s12646-015-0315-9

Alvargonzalez, 2013 – Alvargonzalez, D. “Is The History of Science Essentially Whiggish?”, *History of Science*, 2013, vol. 51, no. 1, pp. 85–99.

Antonovski, A.Yu. “Nauka v superpozitsii. K kommunikativnoj semantike ponyatiya nauki” [Science in Superposition. Towards a Communicative Semantics of the Concept of Science], *Epistemology & Philosophy of Science*, 2024, vol. 61, no. 4, pp. 6–24. (In Russian)

Butterfield, 1965 – Butterfield, H. *The Whig Interpretation of History.* New York: W.W. Norton & Company, 1965.

Chang, 2021 – Chang, H. “Presentist History for Pluralist Science”, *Journal for General Philosophy of Science*, 2021, vol. 52, pp. 97–114. https://doi.org/10.1007/s10838-020-09512-8

Freeman et al., 2014 – Freeman, A., Chick, V., Kayatekin, S. “Samuelson’s Ghosts: Whig History and the Reinterpretation of Economic Theory”, *Cambridge Journal of Economics. L. etc.*, 2014, vol. 38, no. 3, pp. 519–529.

Kasavin, I.T. “Znanie i real’nost’ v istoricheskoj epistemologii” [Knowledge and Reality in the Historical Epistemology], *Epistemology & Philosophy of Science*, 2020, vol. 57, no. 2, pp. 6–19. https://doi.org/10.5840/eps202057216. (In Russian)

Kuzin I.A. “Klassifikaciya form prezentizma i antiprezentizma v istoriografii nauki” [Classification of Forms of Presentism and Anti-presentism in the Historiography of Science], in: I.T. Kasavin, L.V. Shipovalova (eds.) *Nauka kak obshchestvennoe blago* [Science as a Public Good], vol. 1. Moscow.: Russkoe obshchestvo istorii i filosofii nauki, 2020, pp. 74–77. (In Russian)

Kuznetsova, N.I. “Prezentizm i antikvarizm – dve kartiny proshlogo” [Presentism and Antiquarianism – Two Pictures of the Past], *Arbor Mundi*, 2009, is. 15, pp. 164–196. (In Russian)

Kuznetsova, N.I., Rozov, M.A. “Istoriia nauki na rasput’e” [History of Science at the Crossroads], *Voprosy istorii estestvoznaniia i tekhniki*, 1996, is. 1, pp. 3–18. (In Russian)

Loison, 2016 – Loison, L. “Forms of presentism in the history of science. Rethinking the Project of Historical Epistemology”, *Studies in History and Philosophy of Science*, 2016, vol. 60, pp. 29–37. http://dx.doi.org/10.1016/j.shpsa.2016.09.002

Meeteren, Sidaway, 2020 – Meeteren, M., Sidaway, J.D., “History of Geography”, *International Encyclopedia of Human Geography*, 2nd edition, ed. A. Kobayashi, 2020, vol. 7, pp. 37–44. https://doi.org/10.1016/B978-0-08-102295-5.10724-3

Nietzsche F. *O pol’ze i vrede istorii dlya zhizni. Sochineniya* [On the Advantage and Disadvantage of History for Life, Works], vol. 1. Moscow: Mysl’, 1990, pp. 158–230. (Trans. into Russian)

Porus, V.N. “Istoricheskaya epistemologiya – trigger reformy filosofii poznaniya” [Historical Epistemology – the Trigger of the Reform of the Philosophy of Knowledge], *Voprosy filosofii*, 2021, no. 5, pp. 47–57. https://doi.org/10.21146/0042-8744-2021-5-47-57. (In Russian)

Puras, 2014 – Puras, A. “Robert Owen in the History of the Social Sciences: Three Presentist Views”, *Journal of the History of the Behavioral Sciences*, 2014, vol. 50 (1), pp. 58–78.

Rabinovich, V.L. *Alkhimiya kak fenomen srednevekovoj kul’tury: Germes trizhdy Velichajshij. Izumrudnaya skrizhal’* [Alchemy as a Phenomenon of Medieval Culture: Hermes the Thrice Greatest. The Emerald Tablet]. Moscow: Nauka, 1979. (Trans. into Russian)

Rheinberger, 2000 – Rheinberger, H.J. “Cytoplasmic Particles. The Trajectory of a Scientific Object”, in: L. Daston (ed.), *Biographies of Scientific Objects*, Chicago and London: The University of Chicago Press, 2000, pp. 270–294.

Rheinberger, 2012 – Rheinberger, H.-J. “A Plea for a Historical Epistemology of Research”, *Journal for General Philosophy of Science*, vol. 43, no. 1, pp. 105–111.

Riesch, 2014 – Riesch, H. “Philosophy, History and Sociology of Science: Interdisciplinary”, *Studies in History and Philosophy of Science*, 2014, vol. 48, pp. 30–37.

Rozov, M.A. “Prezentizm i antikvarizm. Dve kartiny istorii” [Presentism and Antiquarianism: Two Pictures of History], *Voprosy istorii estestvoznaniia i tekhniki (VIET)* [Studies in the History of Science and Technology], 1994, no. 3, pp. 13–23. (In Russian)

Shapin, 1996 – Shapin, S. *The Scientific Revolution*. Chicago & London: The University of Chicago Press, 1996.

Shuster, 2013 – Schuster, J. *Descartes-Agonistes: Physico-mathematics, Method & Corpuscular-Mechanism 1618–1633*. Springer, 2013.

Soden et al., 2021 – Soden, R., Ribes, D., Avle, S., Sutherland, W. “Time for Historicism in CSCW: An Invitation”, in: *Proceedings of the ACM on Human-Computer Interaction,* vol. 5, CSCW2, article 459, 18 p., https://doi.org/10.1145/3479603

Stocking, 1965 – Stocking, Jr.G.W. “On the Limits of ‘Presentism’ and ‘Historicism’ in the Historiography of the Behavioral Sciences”. *Journal of the History of the Behavioral Sciences*, 1965, vol. 1, no. 3, pp. 211–218.

Wilson, Ashplant, 1988 – Wilson, A., Ashplant, T.G. “Whig History and Present-Centered History”, *The Historical Journal*, 1998, vol. 31, pp. 1–16.

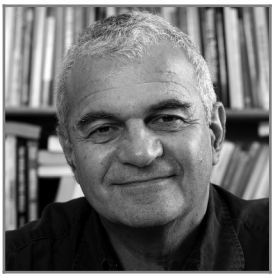
Zagorin, 2001 – Zagorin, P. “Francis Bacon’s Concept of Objectivity and the Idols of the Mind”, *The British Journal for the History of Science,* 2001, no. 34, pp. 379–393.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 22–32
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 22–32
DOI: https://doi.org/10.5840/eps202663337

THE BATH SCHOOL AND THE STRONG PROGRAMME

Harry Collins – FBA.
Honorary Professor.
University Colledge London.
Emeritus.
Cardiff University;
e-mail: CollinsHM@cardiff.ac.uk



I describe the origins of the 'Bath School' with its Empirical Programme of Relativism (EPOR) and its relationship to the Edinburgh Science Studies Unit and its Strong Programme. The Bath work arose independently and did not make contact with the Edinburgh Unit until around 1973. But it then benefitted enormously from the interaction. It is likely that without the Strong Program to 'bounce off' the Bath work would have remained an eccentric curiosity. Nevertheless, the two approaches remained different driven by a strong emphasis on empirical work at Bath, albeit set in a philosophical context, and, in the first instance, a more purely philosophy directed approach from Edinburgh. It is suggested that the apparently coincidental alignment of the approaches of the two groups arose because of the way British sociology had developed. The expansion of sociology in the 1960s drew heavily on the pioneering of the subject at the London School of Economics and that gave rise to a strong philosophical flavour in British sociology, not found in, say, the US. That is probably how both groups found themselves drawing on the later philosophy of Wittgenstein with his theory of meaning depending on use, with 'forms of life' being a precursor to Kuhnian paradigms. The differences between the two groups became less important as Latourian Actor/ant Network Theory came to dominate Science and Technology Studies followed by the determined democratisation of science associated with Jasanoff's 'coproduction' idea, which brought ordinary citizens into the heartland of science. Neither Edinburgh nor Bath ever wanted to doubt the existence of elite scientific expertise even if they did remove science from its claims to certainty, allowing a modicum of social values into the truth finding process. The paper regrets the current state of STS, which appears to support a populism which makes science, not a continuation of politics by other means, but simply politics. The paper invites speculation about how things could have turned out differently.

Keywords: Strong Programme, Bath School, sociology of scientific knowledge, experimenter's regress, Third Wave of Science Studies

БАТСКАЯ ШКОЛА И «СИЛЬНАЯ ПРОГРАММА»

Гарри Коллинз –
член Британской академии.
Почетный профессор.
Университетский колледж Лондона.
Заслуженный профессор.
Университет Кардиффа;
e-mail: CollinsHM@cardiff.ac.uk

В статье рассматривается генезис «Батской школы» и ее Эмпирической программы релятивизма (ЭПР), а также ее соотношение с Эдинбургским отделением исследований науки и его Сильной программой. Батские исследования возникли независимо от эдинбургских; их непосредственный контакт установился лишь около 1973 г. Однако последующее взаимодействие оказалось для батской группы исключительно плодотворным. Можно предположить, что при отсутствии Сильной программы, служившей своего рода «точкой опоры» для критической рефлексии, батская работа так и осталась бы маргинальным интеллектуальным курьезом. При всем при том два подхода сохраняли существенные различия: в Бате

© Harry Collins, 2026



приоритет отдавался эмпирическим исследованиям – пусть и в философской рамке, – тогда как эдинбургская линия изначально тяготела к чисто философской проблематике. Выдвигается гипотеза, что наблюдаемое совпадение траекторий обеих групп не было случайным, а коренилось в особенностях развития британской социологии. Ее экспансия в 1960-е гг. во многом опиралась на традицию, заложенную в Лондонской школе экономики, что сообщило британской социологии выраженный философский компонент, не свойственный, например, американской социологической традиции. Этим, по всей вероятности, объясняется обращение обеих групп к позднему Витгенштейну с его концепцией значения как употребления, где «формы жизни» выступают своего рода предшественником куновских парадигм. Расхождения между двумя школами постепенно утрачивали остроту по мере того, как в исследованиях науки и технологий (STS) все большее влияние приобретала акторно-сетевая теория Латура, а вслед за ней – концепция «совместного производства» (co-production) Ясанофф, ознаменовавшая решительную демократизацию науки и включение рядовых граждан в ее ключевые области. При этом ни Эдинбург, ни Бат никогда не ставили под сомнение само существование элитного научного знания, даже если они и лишали науку притязаний на безусловную достоверность, допуская определенную меру социальных ценностей в процесс установления истины. В заключительной части статьи высказывается сожаление по поводу нынешнего состояния STS, которое, судя по всему, смыкается с популизмом, превращающим науку не в продолжение политики иными средствами, а просто в политику. В связи с этим автором ставится вопрос о том, как могла бы сложиться альтернативная траектория развития данной области, и предлагается поразмышлять над этим читателю.

Ключевые слова: Сильная программа, Батская школа, социология научного знания, регресс экспериментатора, Третья волна исследований науки

I always date the start of the strong programme to David Bloor's 1973 paper, 'Wittgenstein, Mannheim and the Sociology of Mathematics' [Bloor, 1973]. That was the paper that impressed me. But I was doing strong-programme stuff *avant la lettre*, from, according to how you count it, 1967, 1971, or 1972. Actually, I didn't even know there was a field called the 'sociology of science' with Robert Merton as its figurehead until late in 1971. I knew about Robert Merton but as the promulgator of 'Merton's Theory of Anomie', not the science stuff so I was bound to be inventing the subject for myself. That was until I wound up at the University of Bath to do a PhD, with a Mertonian as a supervisor. My origins as a sociologist were very modest – a 'London External Degree' at a local college in Liverpool, with no sociology at a 'proper' university until I did a one-year taught master's in sociology at Essex in 1970–1971.

Going back to those 'according to how you count it' dates, 1967 was when I read, with enormous concentration, Peter Winch's book, *The Idea of a Social Science* – published in 1958 [Winch, 1958]. The reason for the enormous concentration was that, at first, I could not understand it, but when I did understand it, it changed my life. I would eventually realise when, sometime later, I read Kuhn's *Structure of Scientific Revolutions* [Kuhn, 1962], that Winch had prefigured Kuhn's idea of 'paradigm' by four years. Winch described the way surgery worked before and after the establishment of the germ theory of disease: the concept of 'germ' was given meaning by the changed actions of surgeons – shifting from operating in dirty waistcoats to pedantic disinfecting in spotless surgical gowns and gloves etc. This was a change in the social life of surgery. Winch took the idea from the later (eg 1953) philosophy of Wittgenstein [Wittgenstein, 1953] and referred to 'forms of life' and how, if you wanted the meaning of a concept you should ask, not for the meaning, but for the use. Winch thought he was attacking sociology using his exposition of Wittgenstein to say that sociology was 'misbegotten philosophy', but he didn't understand his own book. What he was really doing, and doing it quite brilliantly, was showing how concepts and actions were two sides of the same coin. He could just as easily have said that philosophy was misbegotten sociology. And, of course, Bloor would spend many years trying to get philosophers to accept that Wittgenstein was really a sociologist – it was delightful for someone like me to discover, some years later, that Bloor was doing this [Bloor, 1983]. But this was after I had stumbled upon Kuhn and realised that his 'scientific paradigms' were Wittgensteinian forms of life applied to science.

1971 was when I had to find a little research project to finish off my one-year taught master's degree in sociology at Essex University and wandered round the science labs until I found a likely candidate. This was people trying to build a Transversely Excited Atmospheric Pressure CO_2 laser (TEA-laser), and largely failing. I got in my car and visited all the locations in the UK where the attempt was being made to build one of these recalcitrant devices. I decided I would use my Winch/Wittgenstein/Kuhn perspective to treat the building of such a laser as like trying to learn a language rather than trying to follow a circuit diagram or list of instructions. I called this the 'enculturational' as opposed to the 'algorithmical' model. This was a great idea because I found that no-one could make a TEA-laser work unless someone from their laboratory had spent time in another laboratory where they had a working model (inventor aside, of course). I still knew nothing about the sociology of science nor the Edinburgh group – this was just my obligatory master's research project. I then went to Bath University to do a PhD and that is where I began to find out about the community of which it could be said I was becoming a part because my Bath supervisor was Stephen Cotgrove, who had co-authored a book about how industrial scientists could work within



the Mertonian norms. Of course, the job of a British sociology student in those days – not long after 'the sixties' – was to explain to his or her supervisor how thoroughly wrong they were about almost everything.

1972 was when I found myself driving across America in an old car interviewing scientists. I was completing my TEA-laser study by interviewing the Americans who had tried to build one; I was trying to interview some theoreticians in the domain of the theory of amorphous semiconductors, but that went nowhere except in terms of learning what it meant to understand and not understand a field of science; I was interviewing parapsychologists, which was its own kind of experience and would lead to some papers and a book; and I was interviewing scientists in the controversial field of the detection of gravitational waves, a project that I would stick with for another 45 years.

If I've ever had a 'eureka moment' it was towards the end of this journey across America. Driving through the endless scrub of Nevada I realised that I could not write up the gravitational wave study in the way I had written about the TEA-laser study. In the case of the TEA-laser scientists I knew whether they had succeeded in building a working device because if they had it would produce a powerful beam of infra-red radiation that would make the wall of the laboratory smoke. In the case of gravity wave detectors, however, scientists were still arguing about whether a device of the sensitivity of those currently being built should or should not detect gravitational waves. And if they didn't know that, they didn't know whether their gravitational wave detector was like a TEA-laser that worked or a TEA-laser that didn't work. For about 20-minutes I was devastated as I realised my PhD couldn't be written up in the way I had intended when I started my trans-American journey so I had wasted my time and money. Then it struck me that I had 'discovered' that experimenters in frontier fields never knew if their experiments were working or not and that is what they had to argue about. Frontier science was not a matter of aggregating the results of a series of repeated experiments but a matter of arguing about which set of experiments with competing results were the ones that were working properly. Later it became clear that the same problem beset the fields of parapsychology that I was looking at. This problem would eventually come to be called 'the experimenter's regress' and unknowingly, I had tripped over a vital foundation stone for the 'social constructivist' approach to science. But all this was still before I knew there was a strong program.

It was around this time or in the next year or two that I found out about the existence of the Edinburgh Science Studies Unit and something about the work going on there and found out more about the field of which I was becoming a part – the sociology of science. I remember bumping into Richard Whitley who was in the sociology of science group located in Manchester university and he asked me where I was going to publish my TEA-laser study. Up to that point I had just thought of myself as

a student not someone who publishes stuff. I think it might have been him who told me about the journal *Science Studies* and its editor, David Edge, who was also the head of the Science Studies Centre at Edinburgh. I suppose it must have been 1973 when I sent my TEA-laser paper, now including the American data, to Edge, as a submission for publication in his new journal. Edge wrote back very encouragingly, telling me how good it was that someone was doing empirical work in this domain not just theorising about it. He sent the paper out for refereeing and one referee wrote back saying that what I was writing about was Michael Polanyi's concept of tacit knowledge (which I'd never heard of). So that is how the paper (and a lot of my subsequent career) has come to be about tacit knowledge.

The TEA-set paper was published in 1974 [Collins, 1974] and the paper about experimentation in frontier physics supported by the gravitational wave study was published in 1975 [Collins, 1975] and this led to me being one of most salient researchers in the new version of sociology of science that was building fast.[1] I also won a grant to study the parapsychology that turned on the Israeli showman, Uri Geller, who was puzzling scientists, and that was how I wound up with a job at Bath University. With the grant I was able to recruit Trevor Pinch to be my research assistant. This, along with a graduate student, and a lot of subsequent publications, led to us being referred to as 'The Bath School', which had by now acquired an international reputation. We had our 'Empirical Programme of Relativism' (EPOR) while the Edinburgh Unit had its much better known 'Strong Programme'.

What I was doing and what Edinburgh were doing had common roots in the later philosophy of Wittgenstein. I think this grew out of the way sociology in the UK developed. At the time I was becoming a sociologist, there was a lot of philosophy in UK sociology. The London External degree that I had done was driven by the syllabus of the London School of Economics and that LSE syllabus fed strongly into most British universities. Philosophy of science was a central concern, with Karl Popper being the figurehead, but the LSE also felt it had to teach Winch's philosophical attack on sociology (philosophical attack as Winch saw it) and that was why I found myself reading it with burning intensity. Various syllabuses would also include discussion of what was called 'the rationality debate'. This questioned how anthropologists should treat 'strange' societies such as that of the Azande. The Azande believed that a good test for whether someone was a witch was to use the 'poison oracle'. They asked the question and made-up some poison called 'benge' and fed it to a chicken. If the chicken died the person was a witch, if it lived, they were innocent. But they repeated the test to make sure. This time,

[1] The ideas in these papers were consolidated in my 1985 book *Changing Order*, which would have been written a few years earlier were I not working on parapsychology fieldwork.



if the chicken lived the person was a witch and if it died, they were innocent. The question for the anthropologist was whether the Azande should be treated as 'irrational'. When I was at Essex, I wrote an essay about this arguing that the Azande weren't irrational because if you were inhabiting a magical rather than a scientific paradigm/form of life it made perfect sense that the magic should work different ways on repeated tests. So, this kind of paradigm/form of life kind of thinking was in the air in British sociology in those years. I think that's why I wound up doing something the consequences of which were not unlike the strong programme even though I had never heard of it. But one might say, I had the 'advantage' of not having heard of it because it meant my approach started from empirical studies. I loved getting in a car and interviewing scientists but I also loved putting the results into a philosophical framework because I started with Winch and Wittgenstein.

I had come from 'nowhere' and stumbled into this stuff and without the strong programme I think I would have just disappeared as just another eccentric.[2] New ways of thinking have to be recognised and build alliances if they are to flourish. Ludwik Fleck's analysis of science [Fleck, 1979] published in the 1930s was pretty much the same as Thomas Kuhn's except for the notion of paradigm-revolution but he was not recognised, at least not in the English-speaking world, until long after Kuhn's book caused a splash. The existence of the Edinburgh Science Studies Unit and the controversy associated with the Strong Programme gave me a social identity and a social group to join up with, rather than just being classed as an odd individual. This had a disadvantage too. The downside was that people who came across my work thought I was a member of the Edinburgh team and had come out of the Strong Program. Therefore, I had to spend some time doing 'product differentiation' in respect of the Edinburgh group even as a number of Edinburgh students were taking up my strongly empirical approach. Nowadays, that careful picking apart of differences seems odd because from the 1990s onward Bruno Latour's Actor Network Theory and hyper-symmetry made Bath and Edinburgh seem very close in term of our overall approach and we were also very close in not thinking that our job was to attack science and level it down, but to understand it better while still respecting it. The similarity of Edinburgh and Bath was even more apparent with the growing popularity of full-scale attempts to democratise science associated with Jasanoff's 'co-production' idea and support for the parent rebellion against Measles, Mumps and Rubella vaccine, which transported ordinary citizens into the heartland of science.[3] Neither Edinburgh nor Bath ever wanted to doubt

[2] Though the little TEA-laser study is still being cited and even discussed in the newspapers 50 years on.

[3] Though it has to be said that Wynne, who was central to this democratisation movement, was originally an Edinburgh student.

the existence of elite scientific expertise even if they did remove science from its claims to certainty, allowing a modicum of social values into the truth finding process. This paper regrets the current state of STS, which appears to support a populism which makes science, not a continuation of politics by other means, but simply politics.

Though David Edge, an ex-radio astronomer, was the founder of the Edinburgh Science Studies Unit, the theoretical drive came from David Bloor and Barry Barnes. It was Bloor's concept of 'symmetry' [Bloor, 1976] that delighted me most in so far as it fitted my 'relativistic' approach to things and, of course, I loved the way that Bloor was interpreting Wittgenstein. Though I admired Barry Barnes, I did not get the same warm feelings from his work. But Bloor was a philosopher and wanted to do philosophy. Only later in his career, after he had written two books about the interpretation of Wittgenstein, did he turn to case studies of airplane flight, circling round to his personal enthusiasm, flying – he piloted an old Spitfire at one point. But neither Bloor nor Barnes had any enthusiasm for empirical investigations that took them out of the library and that's why several of the Edinburgh PhD students turned to my work as a model.

In terms of building a community, however, it was David Edge, founding the Unit and the journal and engaging us in foreign conferences who made the growth of the approach possible.

In immediate concrete terms the person who made all the difference was Steven Shapin, a biologist turned historian of science, also employed in the Science Studies Unit. Shapin knew a whole network of historians of science – he would co-author a prize-winning book about Robert Boyle and the airpump with Simon Schaffer, a historian of science at Cambridge – and introduce non-entities like me to these people, and be the driving force behind a series of conferences we organised at Bath. They were great days and for a few years we thought we might succeed in reinventing social studies of science as a genuinely cross-disciplinary field – a true combination of history, philosophy and sociology of science. But the professional roots of history of science and philosophy of science went too deep into the soil for them to join in with the shallow rooted sociology of science. And it is true that history had professional standards that sociology could not match and, according to me anyway, sociology, and sociology of science in particular, have misunderstood their role in the world – we'll come to that – so I doubt it would ever have worked.

Meantime, I went from being a nonentity to an 'entity', becoming the President of the international Society for Social Studies of Science from 1991 to 1993. David Bloor continued doing his individualistic thing, while Steve Shapin did stand for the presidency after my time was over but, scandalously in my view, failed to be elected. I wonder if the three academic professions would have held together longer if that election had worked out differently.



Since then, a lot of blood has flowed under the bridge. From about the 1990s onwards, Latourianism came to dominate the science studies domain to a staggering extent. To me it has never made any sense. The speed with which it 'caught-on' must have had something to do with its polysemic quality. A polysemic slogan is useful in political campaigning because it allows voters to impute to it whichever meaning they prefer. 'Make America Great Again' is like this: it can mean invade Venezuela, Greenland and Canada, or pull out of 'forever wars' such as that in Ukraine and dump your allies; it can mean 'make our streets safe again' or send ICE into Minnesota to murder citizens and strengthen the gun lobby; it can mean bolster employment by imposing tariffs on foreign goods or do away with welfare and support 'the wealth-pump' while ruining the climate for poor nations. With Actor(ant) Network Theory you could not even tell if was the most radical application of social constructivism there had ever been, or a bugle call for collaborationism, or a recipe for returning to old fashioned history of technology with the non-humans regaining their traditional agency as I argued in a 1992 article [Collins, Yearley, 1992, especially pps 316ff]. I also offered a narrower interpretation of the success of Latourianism: that Bruno's brilliance was in giving arts and humanities scholars a way of criticising science without having to understand any science and thus coming out on top of the 'two cultures' debate without any 'heavy lifting'. You can see how confused I am about it because I've been complaining that Latour offered no explanation for his notorious *volte face* of 2004 when we all found him back on the side of science again (leaving the science warriors sprawling). But then I realised after some nudging that in the earlier 'Epistemological Chicken' I had argued that ANT was scientifically conservative. So, it wasn't really a *volte face* after all but just another small step in the usual conservative direction dressed up as *volte-face* so as to be a judo-type sidestep 'performance' with everyone, including me, tumbled over the outstretched leg for dramatic effect.[4]

Anyway, going back to the blood flowing under the bridge, there's been a tsunami's-worth in the last decade. I wonder how much of the STS domain's reaction to Trumpism can be traced back to those early differences. Perhaps the unpredictability of what has happened to science in America suggests there is nothing to say that could link the early days to today's happenings. Or maybe it was not so unpredictable. Remember, there was a period on the 1990s when the 'science warriors' emerged and started warning about the dark ages that would follow from the levelling down of science. And they were right – the dark ages for science have arrived! And they have arrived in the wider realm of national politics, with

[4] Thanks to Nicolas Langlitz for pointing out to me that this wasn't a *volte-face*.

post-modernism serving as a workshop manual for the erosion of truth in entire populations.

Unfortunately, the science warriors, though they saw what was coming, had no idea what to do next. All they could do was put their energy into sniping stupidly at the really good work that was being done in the post-1960s studies of science toward understanding science better.[5] In 2002, prior to Trumpism, Collins and Evans (C+E) tried to do something that they thought was more sensible by inventing the Third Wave of Science Studies; they were really reacting to ambiguity over vaccination. C+E said the First Wave was the high point of science before and after WWII. They said the Second Wave was the much richer analysis of science, probably triggered by Kuhn's book but based more in the philosophy of Wittgenstein which began in the early 1970s which had the effect of eroding the mountain of science down toward the plain of ordinary knowledge; unfortunately this philosophical wave metamorphosed into Latourianism and also into the democratisation of science movement, led by Sheila Jasanaoff and Brian Wynne. C+E said the Third Wave had to be designed to raise the level of science back up again, if not so far as the first wave, at least far enough so that we could still make sensible, science-based, decisions about vaccination and so forth.

The Third Wave was to be based on the analysis of scientific expertise, which could be recognised and analysed in real time even if scientific consensus formation was much slower. The more recent iterations of the Third Wave programme declare that science is 'craftwork with integrity' with aspiration that goes with 'integrity' being the crucial feature: as an institution, science is obsessed with truth in the way that no other institution is obsessed and can therefore act a check and balance. Its political importance is not so different as that of law, or the parliamentary opposition and, unlike these, it could also be an object lesson in how to use truth to make decisions of every type. And then along came Trump, revealing in cartoon form how vital the Third Wave was if STS was to contribute to the preservation of democracy.

But the unexpected thing is the way STS has responded. First, though the Third Wave paper is hugely cited, the term 'Third Wave' is almost never mentioned in STS circles. The citations are mostly coming from Philosophy, Psychology and other disciplines, with those coming from within STS nearly all having to do with the analysis of expertise. STS it seems, is determined not to 'look up' at the huge lump of political rock about to collide with the Earth and destroy all scientific life.[6] Instead, STS is either stuck with the criticisms of the Third Wave paper that

5 Richard Dawkins and others have recently reiterated their tired old stuff in a new book which I imagine they wrote wearing red baseball caps sporting the 'Make Science Great Again'.

6 I am referring to the film, *Don't Look Up*.



greeted its initial publication or with a determination to 'democratise' science and make it the property of the ordinary people or of disadvantaged or marginal groups.[7] There seems to be no realisation that the disadvantaged groups will be destroyed by the rock just as much as the rest of us. The problem has changed: the problem is no longer rescuing the disadvantaged from the power of elite scientists but rescuing elite (expert) scientists from the power of politicians. Truth is stranger than fiction so it is in this case. Sheila Jasanoff who, along with her ally in supporting parents in their opposition to Measles, Mumps and Rubella (MMR) vaccine, have never withdrawn or reconsidered their position on this matter but, on the contrary, Jasanoff has 'doubled down', arguing (supported by Kant so she claims though this is disputable), that citizens have to make their own judgements. Her words are a perfect match for those of Robert F. Kennedy Junior and Donald Trump is uttering the same stuff, attacking not only climate change science, which attack is at least comprehensible in terms of oil interests, but vaccination too, right down to the particular case of MMR.

I am more and more astonished at what the STS community is doing, and not doing, and cannot but wonder if it would have turned out differently if things had been different in the early days.

References

Bloor, 1973 – Bloor, D. "Wittgenstein and Mannheim on the Sociology of Mathematics", *Studies in the History and Philosophy of Science*, 1973, vol. 4, pp. 173–191.

Bloor, 1976 – Bloor, D. *Knowledge and Social Imagery*. London: Routledge and Kegan Paul, 1976.

Bloor, 1983 – Bloor, D. *Wittgenstein: A Social Theory of Knowledge*. London: Macmillan, 1983.

Collins, 1974 – Collins, H.M. "The TEA Set: Tacit Knowledge and Scientific Networks", *Science Studies*, 1974, vol. 4, no. 2, pp. 165–186.

Collins, 1975 – Collins, H.M. "The Seven Sexes: A Study in the Sociology of a Phenomenon, or The Replication of Experiments in Physics", *Sociology*, 1975, vol. 9, no. 2, pp. 205–224.

Collins, Yearley, 1992 – Collins, H.M., Yearley, S. "Epistemological Chicken", in: Pickering, A. (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press, 1992, pp. 301–326.

Collins, Evans, 2026 – Collins, H., Evans, R. *Veritocracy: Truth, Science, and How to Preserve Democracy*. New York: Bloomsbury Academic, 2026 (forthcoming).

Fleck, 1979 – Fleck, L. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press, 1979 (first published in German in 1935 as *Entstehung und*

[7] The opposite approach is represented in, for example, [Collins, Evans, 2026].

Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv).

Kuhn, 1962 – Kuhn, T.S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Winch, 1958 – Winch, P.G. *The Idea of a Social Science*. London: Routledge and Kegan Paul, 1958.

Wittgenstein, 1953 – Wittgenstein, L. *Philosophical Investigations*. Oxford: Blackwell, 1953.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 33–44
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 33–44
DOI: https://doi.org/10.5840/eps202663338

An invitation to Harry Collins' humanistic sociology

Charles Thorpe – PhD in Sociology, Professor.
Department of Sociology,
University of California
San Diego.
9500 Gilman Drive, La Jolla,
CA 92093-0533, USA;
e-mail: cthorpe@ucsd.edu



Harry Collins has developed a humanistic sociology of science, technology, and expertise. This is based on the insight that humans are distinctive by virtue of having *socialness*. Socialness is made possible by language. But language is only meaningful in a social form of life. Only human beings are capable of being collective representations. Therefore only human beings are capable of being and doing knowledge. Collins shows that expertise is real and only exists in forms of life. So, for society to have scientific knowledge, it has to protect and respect the integrity of the human forms of life, communities of scientific knowledge-making, communities that embody expertise.

Keywords: Collins, STS, socialness, forms of life, scientific knowledge

Приглашение к гуманистической социологии Гарри Коллинза

Чарльз Торп – доктор социологии, профессор.
Кафедра социологии
Университета Калифорнии
в Сан-Диего.
9500 Gilman Drive, La Jolla,
Калифорния 92093-0533,
США;
e-mail: cthorpe@ucsd.edu

Гарри Коллинз разрабатывает гуманистическую социологию науки, техники и экспертизы. Ее исходный пункт состоит в том, что отличительной чертой человека является социальность (“socialness”), которая становится возможной благодаря языку. Однако язык обретает значимость лишь внутри социальной формы жизни (“form of life”). Следовательно, лишь человеческие существа способны выступать в качестве коллективных репрезентаций – и, стало быть, лишь они способны быть и творить знание. Коллинз последовательно показывает, что экспертиза реальна и существует исключительно в формах жизни. Отсюда вытекает нормативное следствие: если общество желает обладать научным знанием, оно обязано защищать и уважать целостность тех человеческих форм жизни, которые представляют собой сообщества производства научного знания – сообщества, воплощающие в себе экспертизу.

Ключевые слова: Коллинз, исследование науки и технологий, социальность, формы жизни, научное знание

© Charles Thorpe, 2026

Harry Collins' sociological project centers on a conception of the human being as, uniquely, a social being. This conception is set out in the pivotal article, in which he introduces the term "socialness" [Collins, 1998]. Among all the beings of which we are aware, only human beings are *capable* of membership in a society. Only human beings are capable of being a collective representation. Therefore, only human beings are capable of being a scientist and doing science; only human beings are capable of being a gravitational wave physicist and doing gravitational wave physics; only human beings are capable of being a sociologist and doing sociology.

Collins' sociology is humanistic sociology. This is the humanistic sociological tradition represented by Peter Berger's *Invitation to Sociology: A Humanistic Perspective* [Berger, 1963], of which Collins acknowledges the influence [Collins, 2019, p. 65]. Berger suggested that sociology was humanistic in its scientific values. He also argued that it should be humanistic in its epistemological approach to the world. While defining the situation of sociology as a "scientific enterprise," Berger wrote that "the discipline has other traits that assign it to the immediate vicinity of the humanities," even though, he still insisted, not belonging "fully" to the humanities. What made sociology epistemologically humanistic, according to Berger, was that it was about "the human condition itself"; that is, "what it means to be a man" and "what it means to be a man in a particular situation" [Berger, 1963, p. 167]. The keyword, here, is "means."

Berger writes that

> One cannot fully grasp the political world unless one understands it as a confidence game, or the stratification system unless one sees its character as a costume party. One cannot achieve a sociological perception of religious institutions unless one recalls how as a child one put on masks and frightened the wits out of one's contemporaries by the simple expedient of saying "boo" [Ibid., p. 165].

There is a potential difference between how Berger characterizes understanding in the three above examples of the political world, stratification, and religion. In the first two examples, understanding of an institutionalized pattern of social action is achieved by understanding of it *as* an instance of something else, understanding that this also *counts* as that, that this game is also that game. This means understanding the definition of the situation and therefore the meaning of what is being done (see also [Ibid., p. 20]). But in describing the example of religion, Berger suggests that sociological understanding requires "recall" and, in particular, "recall how." This could mean that, in order to understand religion sociologically, one must be able to call up in oneself past embodied experience of spooking or being spooked, being frightened by something one cannot see, or by a magic word. Is the main thing about "boo" the onomatopoeia? Or does "recalling how" involve recalling how one went about it



as a practice, and how one learnt to do it? Collins would suggest that what spookiness is, sociologically, can be learned by talking about it. The opposite view is that, if one has never jumped in fright or experienced glee in making someone jump, then one cannot understand "boo" and one can't understand the phenomenology of religion. Berger argues that the social meaning of religion is the protection that it offers from what is beyond sight, or from chaos and death, by its provision of a socially constructed reality. This reality forms a sacred canopy as nomized, "humanly significant," or sacred, cosmos ([Berger, 1969, p. 21, 25, 114]; see also [Berger, Luckmann, 1967]).

The social meaning of awe (or fright) is not the same thing as the feeling of awe, even though the social meaning depends on some people having had that feeling, because it depends on the social practices of "boo" and religion. What is social is what is shared. So the sociologist does not need to know what it is personally, experientially like to be spooked, only what it is socially understood to be like. What is social is meaningful within a form of life. Therefore, what is social can be conveyed in language, even though it can never be rendered entirely explicit in language; it is contained implicitly in the usage of the language. To understand a form of life is to be able to fluently interact with participants using the practice language of the form of life in which they participate.

This is where a gap opens within humanistic sociology, between different conceptions of this project, which were not always, including by Berger and Luckmann, entirely clearly separated. In *Changing Order* [Collins, 1992] Collins notes that "There is a relationship between phenomenology and the ideas of Wittgenstein" [Ibid., p. 25, note 13]. This became apparent with regard to Hubert Dreyfus's phenomenological critique of Artificial Intelligence, which Collins found inspiring, but ultimately misdirected in its emphasis on embodiment rather than social language-use [Ibid., p. 28, note 22; Collins, 2018].

Phenomenology seeks to unify knowledge with being. In this way, phenomenology suggests a sociological task of finding the making of meaning in the flow of everyday social life [Berger, Luckmann, 1967]. Collins seeks to understand the "intermingling of language and social life," conceptual order and social order [Collins, 1992, p. 24, note 7]. Collins treats the phenomenon sociologically at stake not as private experience but social meaning.

Collins demonstrates how Peter Winch's critique of social science can become a positive program for sociology rather than a negative limit. Collins argues that sociology concerns what constitutes forms of life *as* forms of life. Forms of life are made of language. Therefore, social relations are internal relations, logical relations of entailment among meanings within a form of life. *And*, logical relations are made and sustained socially. Collins writes that Winch shows "that social relations are embodied logical relations while logical relations are the abstract

counterpart of social relations" [Collins, 1992, p. 27, note 18]. This is *living* language in use in social life. It is language-use as containing tacit knowledge [Collins, 2011]. The use of the language requires knowledge, contained, but implicit, in use. To understand language is to glean tacit meaning. This tacit, unexplicated, meaning, consists of what is "of course" in a form of life [Berger, 1963, p. 47]. Of-courseness, or the natural attitude, is an important moment in the social process of reification whereby human beings turn action into practice, rule, and institution. Closure of controversy allows knowledge to enter the domain of background assumptions. This domain is contained in a form of life, its practices, and practice-language. The Science & Technology Studies (STS) metaphor of the "black box" is what phenomenology identifies as taken-for-granted reality.

Collins agrees with phenomenologists that social things are ways of being. He further agrees that ways of being, and the practices in which they inhere and which produce them, are intrinsically open. In this way, all practices contrast with a pure theoretical system and with the "internal workings of a developed computer" [Collins, 1992, p. 25, note 11]. Experiment, as practice, is an open system, as shown by the great difficulty of separating 'signal' from 'noise', meaningful order from chaos. The problem is separating out and thereby constructing a phenomenon, and through that, a fact, from the open flow of life. Doing so is "cosmization," the transformation of chaos into cosmos, or a meaningful order of things, in which phenomena are known and there is an order that can be comprehended [Berger, 1969, p. 26–28, 114].

Every event is at the edge of socially constructed reality and introduces uncertainty. This makes for the problem of induction. Collins is interested in people whose job it is to extend cosmos, to produce scientific knowledge, and thereby, socio-cognitive order. The production of knowledge, Collins argues, is a practice, which means it is open not only in the sense of growth but in the sense that it is capable of producing genuine newness. Rules are algorithms but following the rules is not algorithmic [Collins, 1992]. Rule-following as a practice must go beyond the rules. Science is a (multiplicity of) practice(s) and therefore not algorithmic [Ibid.].

The difficulty that science faces as a social practice of cosmization is that it does not provide nomos. Steven Shapin [Shapin, 2008] traces how the secular institutionalization of the role of scientist, associated with the research university and with industry, emerged as a conceptual innovation or innovation in meaning *and* innovation in social being. This was a movement away from claiming religious, or spiritual, and moral authority and from calling to, morally ordinary, job [Ibid., p. 21–91].

Weber's [Weber, 1958] lecture, "Science as a Vocation," according to Shapin [Shapin, 2019], marks a historical moment in which the attempt to connect science to answers to meaningful existence was



declining in significance within the definition of science as a form of life and the role of the scientist. But Weber nevertheless thought that science was worthwhile as a personal commitment, or vocation. That science was not entirely de-moralized [Fevre, 2000], was indicated by Robert K. Merton's notion of "the ethos of science" or of science as having a distinctive "normative structure" [Merton, 1973]. Merton suggested that the scientific ethos had significance beyond the social institutions and professional community of science because there was an affinity between the normative structure of science and that of democracy. But for Merton the key implication of that affinity was that science needed a broader democratic society in which to thrive.

Collins' Empirical Programme of Relativism (EPOR) was different from Mertonian sociology of science not only because of Collins' willingness to pursue the sociology of knowledge but also because of a difference in style. Whereas Mertonian sociology was reverent in its tone toward science, the new British sociology was irreverent, in its ambition to treat scientific knowledge as ordinary. In this way, the sociology of scientific knowledge and the empirical programme in relativism followed the Chicago symbolic interactionists in a deflationary way of describing the practices of professional experts and high-status or official organizations. Alvin Gouldner illustrated the difference between Chicago and Harvard/Columbia sociologists' orientation to the professions by the contrast between book titles of their respective studies of medical education: "Harvard and Columbia have soberly called two of their most important works *The Student-Physician* and *Experiment Perilous*, while the Chicagoans have irreverently labelled their own recent study of medical students the *Boys in White*" ([Goulder, 1973, p. 17]; see [Fox, 2019; Merton et al., 1957; Becker et al., 1961]).

Berger observed in his 1963 book that "science is an almost sacred entity among Americans in general and American academicians in particular." Berger thought that it was pursuit of this symbolic status that motivated many social scientists' "fixation on methodology", or fetishism of the natural sciences, out of which he hoped sociology would mature [Berger, 1963, p. 13]. Collins also criticizes "magical scientism" which equates being scientific with applying statistical techniques [Collins, 2019, p. 22]. Collins defends the claim of sociology to be a science. He acknowledges that statistics have an important place in sociology in illustration of relations within society. But he argues that statistical correlations are not the foundation of sociological knowledge. While statistics may indicate a relationship, they do not in themselves constitute the relationship. This is true of natural phenomena. Epidemiology does not render laboratory studies unnecessary. But there is also the inherent difference between social and natural phenomena, which is the constitutive meaningfulness of social phenomena. As Winch argued, social relations are internal because they are relations of ideas: "social relations

between men exist only in and through their ideas" [Winch, 2008, p. 115]. Winch explicates the internal relationship between the command and the act of obedience. The meaning of the act as an act of obedience is internal to the act and to the command, since command and act are only comprehensible in terms of each other [Ibid., p. 116–117]. So sociological knowledge is the comprehension of relations of entailment of meanings within a cultural system of ideas or concepts.

As logical relations, social relations can be known with great certainty. If this is not absolute certainty, this is because forms of life undergo change. Forms of life are incomplete totalities, in the process of being continually remade, and this change is informed by the fact that forms of life exist in dynamic relation to other forms of life. Change in forms of life necessarily involves changes in meaning.

Winch says, "Part of the opposition one feels to the idea that men can be related to each other through their actions in at all the same kind of way as propositions can be related to each other is probably due to an inadequate conception of what logical relations between propositions themselves are" [Ibid., p. 117]. Here, Winch opens the door to Collins' project of a sociological epistemology: "it is only from their roots in this actual flesh-and-blood intercourse that those formal systems draw such life as they have" [Ibid., p. 118]. Collins realized the meaning of this statement insofar as it suggested a completely new way of thinking about, and going about, epistemology. This involves understanding relations of meaning within a form of life, and so it must be necessary to gain knowledge of the form of life. It must be necessary to study the form of life as, and in, life, in order to understand how its forms are practiced, or, rather, the practical meaning of forms.

So whereas analytical, or logical empiricist, philosophers formalized and hypostatized scientific method, including notions of experiment and replication, Collins argued that a practice, such as experiment and replication of experiments, could never be completely understood algorithmically or through the elaboration of rules. Instead, he sought to show experiment as practiced by an experimental community as a form of life. The same insight that enabled him to understand Winch's implications for a sociological epistemology enabled Collins to identify the experimenter's regress precisely as a problem that arises in practice. When the question of what has been found about the physical world becomes a question of the criteria of evaluation of the validity of the experiment itself, this is a kind of regress, or reflexivity. This reflexivity reflects the internality of social relations as social meanings. The result of the experiment is internally, in its meaning, related to the experiment itself.

Winch distinguishes the example of obedience and command from the example of hearing 'thunder.' The term "thunder" is "theory-impregnated" because the ability to use it to identify the cause of a sound depends on holding beliefs such as in its relation to lightning and electricity,



or, one could alternatively say, fire and gods. The idea of thunder is connected with other ideas. But Winch draws the distinction that "An event's character as an act of obedience is *intrinsic* to it in a way which is not true of an event's character as a clap of thunder" [Winch, 2008, p. 116–117. Emphasis in original]. Collins shows how "thunder," or a gravitational wave, exists in internal relations with other ideas and is therefore constituted within a form of life. What is "intrinsic" to whatever a gravitational wave is, apart from the idea, is not the business of sociology. Constructivism is intrinsic to sociology.

Gods and gravitational waves exist in different kinds of form of life, different in their concept-practices. An individual may inhabit multiple forms of life, and therefore alternate between their concept-practices, while these forms of life remain, in themselves, separate and incommensurable [Berger, 1963, p. 51–52]. It is the possibility of alternation that makes sociology possible and shows the necessity of methodological relativism for sociological comprehension of the phenomenon of society ([Collins, 2019, p. 65–67], citing [Berger, 1963]). The notion relates to the concept of participant comprehension [Collins, 1984; Collins, 2019, p. 54], as a formulation of what Collins and Evans would later call "interactional expertise" [Collins, Evans, 2007]. Methodological relativism corresponds to the ontology of society as a complex, fractal, multiplicity of forms of life alongside and within forms of life [Collins, 2019, p. 8–13]. The difference between the concept of thunder as god and the concept of thunder as electricity is also the differences of chains of meanings, concepts, practices that are internally related within each of these forms of life. This difference therefore corresponds to vast differences in ways of living and being and vast differences in the "of courses" of each social way of living.

Sociology knows social life only to the extent that the sociologist is able to immerse themselves in that social life, which means in the socialization processes of that form of life, thereby becoming socialized. But, Collins argues, it is possible to be socialized in a form of life to the extent that one understands that form of life, without becoming a participant in it. This is what makes alternation possible and the complex fractal of forms of life possible.

And the way in which meanings, and forms of life, are practico-logically interrelated may create resistance to change because of widening and intersecting stakes and may create a certain brittleness or fracturability in social relations, due to ways in which "of courses" interdepend. Science is a form of solidarity. This solidarity involves expertise or know-how and cultural fluency. It also involves a moral ethos or collective aspiration. This suggests that the acquisition of expertise is also, tacitly, socialization into a moral order [Ibid., p. 101]. Trust occurs within moral order [Shapin, 1994]. It is the fact that practices, including science, cannot be fully stated algorithmically that makes practices

trust-dependent. The practice of science contains tacit knowledge and therefore is trust-dependent. Collins' ethnographic studies and Steven Shapin's historical studies are both concerned with the moral fabric of scientific community and role or being. Both Collins and Shapin emphasize what science shares with other forms of life, that is its ordinary socialness.

What demarcates science distinguishes it as a form of life and makes sociology a science, even though its way of knowing obedience is in many ways completely other to the way of knowing thunder as caused by electricity. Except of course it isn't completely different because both involve internal relations of meaning and therefore knowing as interpreting. But sociology is methodologically different from physics because the things that it studies are ontologically different from those of physics. This is because social things are *constituted by* their meanings, because their very existence depends on language and meaning [Winch, 2008, p. 117]. Scientific facts are social things, and this is how they are designated by sociologists, but not by physicists, because sociology and physics have different constitutive goals. They are different language-games. But this does reflect the real difference between social things and physical things, or things-understood-as-social not things-understood-as-physical.

There is a real difference between social things and physical things (*qua* physical) because only human beings have socialness. This means that the existence of social things depends on human beings. This is because only human beings use language as part of their form of life. So only human beings can do concept-practices and concept-being. Only human beings can be concept beings. Only human beings can be a concept.

Being a concept requires expertise. One must know how to do it in order to be it. This is participatory expertise. It is knowing how to use concepts in the course of a practice. The key expertise that a sociologist must acquire, through methodological relativism, alternation, and immersion, is interactional: knowing how to use concepts in order to understand a practice and its practitioners [Collins, Evans, 2007; Collins, 2011; Collins, 2020].

Practices are normative. Normative judgment is inherent in practices. This is because practices are teleological. Forms of life, and institutions, have constitutive goals, or what Collins [Collins, 2023, p. 12] calls "formative aspirations." Since they have purpose, practices are value-laden. This is part of their theory-ladenness and knowledge-ladenness which may be tacit, i.e. assumed in the practice and contained in the practice-language. Collins argues that science is value-laden, and that at its core is a moral "formative aspiration" to tell the truth [Ibid., p. 145]. This seems to be close to Merton's conception of the "ethos of science" and norms of science (cf. [Collins, 1982, esp. 301]). Merton and Collins both argue that there is a connection between the moral order of science and that



of democracy. Collins emphasizes the dependency of democracy on science. These connections are interdependencies of meaning within and across intersecting forms of life, morphing to different polities. Shapin argues that experiment emerged as a practice meaningfully related to moral ways of being, a key such example being the early modern English gentleman [Shapin, 1994; Shapin and Schaffer, 1986]. Yaron Ezrahi [Ezrahi, 1990] traces intersecting meanings of visual witnessing and transparency between science and political democracy.

In 1963 science was "almost sacred" in America ([Berger, 1963, p. 13]; see also [Welsh, 2000]). Yet sociology is ambivalent toward sacredness. It may be argued, sociologically, that social order depends on sacralizing, because this is a crucial moment of social world-making, protecting core values that provide the focus of allegiance and commitment for a form of life, or society, and for giving anchoring legitimacy to chains of "of courses" that extend through social life [Berger, 1969; Berger, Luckmann, 1967; Shils, 2016]. If science is to receive social trust, so that assent is collectively forthcoming to social acts associated with this core institution, and people participate in the actions it legitimizes, such as childhood vaccinations, then science must be sacred.

Acts like vaccinating one's children contain meaning as what John Locke called "tacit consent." As Carole Pateman [Pateman, 1985, p. 72] puts it "Tacit consent can be inferred from the fact that people are going peacefully about their everyday lives" including using government-provided infrastructure such as roads. The meaning of such everyday acts contains allegiance to a polity. Getting vaccinated is polimorphic action (see [Collins, Kusch, 1999]). It is associated with social identity and membership of a form of life.

But not everything we trust, or ways we give tacit consent, involves the sacred. I trust that water will come out of my kitchen tap but I do not hold the Water Board sacred. It may be said that I have confidence rather than trust, or system trust rather than personal trust [Luhmann, 1990; Giddens, 1990, p. 83–88]. But is it true that this kind of system trust can be entirely decoupled from personal trust and de-moralized? Or, does my trust in public utilities have to do with ethics of bureaucratic impartiality, technical competence as a value, and public service that I tacitly ascribe to the people working in the utility? Is it possible that my confidence in such background technologies and routines of everyday life is weakened if I feel that the privatized water company is mostly interested in lining its shareholders' pockets or if I suspect that the government is corrupt? It would seem that the normative commitments that I ascribe to the Water Board, professions, and bureaucracies, that underly my confidence in technological systems, are meaningfully related to the normative commitments that I ascribe to science and scientists (see also [Collins, 1982, p. 301]). And those meanings may be mutually determining so that my trust in science might be informed by my opinion of other

meaningfully-related concepts and institutions such as the Water Board. If so, then, what is sacred about science and the Water Board has to do with the structure of value that is part of the socially constructed reality of a form of life.

If science must be sacred in order to be trusted and must exist within a sacred order of specially elevated and protected values, then this poses a problem for science, if science is disenchanting and disenchanted (cf. [Weber, 1958]). Science drew sacredness from other social sources. The Natural History Museum in London is a cathedral of science. Scientific institutions drew legitimacy from religion even while differentiating from religious institutions. Scientific institutions, including universities, are often associated with the state and so draw upon the symbolic authority of the nation-state. Berger [Berger, 1963, p. 44] noted that sociology is rendered "respectable" by its "affiliation in university, business and government".

Humanistic sociology is disenchanting. Berger gave the analogy between religion and a children's game of "boo" within his argument that the sociological attitude should be comic. According to Berger, "the social scientist who does not perceive this comic dimension of reality is going to miss essential features of it" [Ibid., p. 165]. Sociology is a comic form because it is deflationary. It exposes the ordinariness of all social action, and therefore provides a profane view of the sacred. In this way, it provides a version of Hans Christian Anderson's story of "The Emperor's New Clothes." Ezrahi [Ezrahi, 1990, p. 108–127] argues that Anderson created a key allegory of modernity, expressing the new mode of public witnessing and holding accountable, shared across science and democracy. The connection between laughter and humanism to which Berger alludes is symbolically and historically close [Hietalahti, 2024; Berger, 2014]. This laughter is part of the language-game of humanistic sociology and is part of its honesty [Berger, 1963, p. 165–166].

The Empirical Program of Relativism may be seen as developing and extending the Chicago symbolic interactionists' deflationary exposure of the professions. The kind of solidarity within which this deflationary humanistic laughter is not treated as a threat and as destructive is the kind of solidarity in which the two moments of sociology – comedic disenchantment and the necessity of the sacred – can be more easily alternated between. But humanistic sociology also poses the question of whether social order can mature sufficiently to dispense with sacredness and engage in "a mellowing in self-understanding" and an "ironical skepticism concerning its own undertakings" [Ibid., p. 164–165].

References

Becker et al., 1961 – Becker, H.S., Geer, B., Hughes, E.C., Strauss, A.L. *Boys in White: Student Culture in Medical School*. Chicago: University of Chicago Press, 1961.

Berger, 1963 – Berger, P.L. *Invitation to Sociology: A Humanistic Perspective*. New York: Anchor Books, 1963.

Berger, 1969 – Berger, P.L. *The Sacred Canopy: Elements of a Sociological Theory of Religion*. New York: Anchor Books, 1969.

Berger, 2014 – Berger, P.L. *Redeeming Laughter: The Comic Dimension of Human Experience*. Berlin: Walter de Gruyter, 2014.

Berger, Luckmann, 1967 – Berger, P.L., Luckmann, T. *The Social Construction of Reality*. New York: Anchor Books, 1967.

Collins, 1982 – Collins, H.M. "Knowledge, Norms and Rules in the Sociology of Science", *Social Studies of Science*, 1982, vol. 12, pp. 299–309.

Collins, 1984 – Collins, H.M. "Researching Spoonbending: Concepts and Practice of Participatory Fieldwork", in: Bell, C., Roberts, H. (eds.), *Social Researching: Politics, Problems, Practice*. London: Routledge, 1984, pp. 54–69.

Collins, 1992 – Collins, H.M. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press, 1992. (First published 1985)

Collins, 1998 – Collins, H.M. "Socialness and the Undersocialized Conception of Society", *Science, Technology, & Human Values*, 1998, vol. 23, no. 4, pp. 494–516.

Collins, 2011 – Collins, H. "Language and Practice", *Social Studies of Science*, 2011, vol. 41, no. 2, pp. 271–300.

Collins, 2018 – Collins, H.M. *Artifictional Intelligence: Against Humanity's Surrender to Computers*. Cambridge: Polity Press, 2018.

Collins, 2019 – Collins, H.M. *Forms of Life: The Method and Meaning of Sociology*. Cambridge, MA: MIT Press, 2019.

Collins, 2020 – Collins, H.M. "Interactional Imogen: Language, Practice and the Body", *Phenomenology and the Cognitive Sciences*, 2020, vol. 19, pp. 933–960.

Collins, 2023 – Collins, H.M. "Science as a Counter to the Erosion of Truth in Society", *Synthese*, 2023, vol. 202, article 145. https://doi.org/10.1007/s11229-023-04357-2

Collins, 2024 – Collins, H.M. "Establishing Veritocracy: Society, Truth and Science", *Transcultural Psychiatry*, 2024, vol. 61, no. 5, pp. 783–794.

Collins, Evans, 2007 – Collins, H.M., Evans, R. *Rethinking Expertise*. Chicago: University of Chicago Press, 2007.

Collins, Evans, 2017 – Collins, H.M., Evans, R. *Why Democracies Need Science*. Cambridge: Polity Press, 2017.

Collins, Kusch, 1999 – Collins, H.M., Kusch, M. *The Shape of Actions: What Humans and Machines Can Do*. Cambridge, MA: MIT Press, 1999.

Ezrahi, 1990 – Ezrahi, Y. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy*. Cambridge, MA: Harvard University Press, 1990.

Fevre, 2000 – Fevre, R. *The Demoralization of Western Culture: Social Theory and the Dilemmas of Modern Living*. London: Continuum Books, 2000.

Fox, 2019 – Fox, R.C. *Experiment Perilous: Physicians and Patients Facing the Unknown*. London: Routledge, 2019. (First published 1959)

Giddens, 1990 – Giddens, A. *The Consequences of Modernity*. 1990.

Gouldner, 1973 – Gouldner, A.W. *For Sociology: Renewal and Critique in Sociology Today*. New York: Basic Books, 1973.

Hietalahti, 2024 – Hietalahti, J. "Humour and Humanism: The Role of Funniness in an Imperfect Life", *The European Journal of Humour Research*, 2024, vol. 12, no. 4, pp. 1–20.

Luhmann, 1990 – Luhmann, N. "Familiarity, Confidence, Trust: Problems and Alternatives", in: Gambetta, D. (ed.), *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, 1990, pp. 94–107.

Merton, 1973 – Merton, R.K. "The Normative Structure of Science", in: Storer, N.W. (ed.), *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press, 1973, pp. 267–278. (First published 1942)

Merton et al., 1957 – Merton, R.K., Reader, G.G., Kendall, P.L. *The Student-Physician: Introductory Studies in the Sociology of Medical Education*. Cambridge, MA: Harvard University Press, 1957.

Pateman, 1985 – Pateman, C. *The Problem of Political Obligation: A Critique of Liberal Theory*. Cambridge: Polity Press, 1985.

Shapin, 1994 – Shapin, S. *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: University of Chicago Press, 1994.

Shapin, 2008 – Shapin, S. *The Scientific Life: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press, 2008.

Shapin, 2019 – Shapin, S. "Weber's Science as a Vocation: A Moment in the History of 'Is' and 'Ought'", *Journal of Classical Sociology*, 2019, vol. 19, no. 3, pp. 290–307.

Shapin, Schaffer, 1986 – Shapin, S., Schaffer, S. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, NJ: Princeton University Press, 1986.

Shils, 2015 – Shils, E. "Centre and Periphery", in: Polanyi Festschrift Committee (ed.), *The Logic of Personal Knowledge: Essays Presented to Michael Polanyi*. London: Routledge, 2015, pp. 117–130. (First published 1961)

Weber, 1958 – Weber, M. "Science as a Vocation", in: Gerth, H.H., Mills, C.W. (eds.), *From Max Weber: Essays in Sociology*. New York: Oxford University Press, 1958, pp. 129–156. (First published 1917)

Welsh, 2000 – Welsh, I. *Mobilising Modernity: The Nuclear Moment*. London: Routledge, 2000.

Winch, 2008 – Winch, P. *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge, 2008. (First published 1958)

Эпистемология и философия науки
2026. Т. 63. № 3. С. 45–55
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 45–55
DOI: https://doi.org/10.5840/eps202663339

Empirical Expansion and Philosophical Autonomy: Reconsidering the Strong Programme through Historical Epistemology

Tatiana D. Sokolova –
PhD in Philosophy,
Research Fellow.
Institute of Philosophy,
Russian Academy of Sciences.
12/1 Goncharnaya St.,
Moscow 109240,
Russian Federation;
e-mail: sokolovatd@gmail.com



The article offers a philosophical response to Harry Collins's account of the Strong programme through the framework of historical epistemology. The first part situates the Strong programme within a longer transformation in the relation between philosophy and science. From positivism to the emergence of sociology and science studies, philosophy gradually loses its privileged role as the discipline defining the conditions of knowledge, while science itself becomes an object of empirical investigation. The second part reconstructs the historical a priori of the Strong programme. Drawing on Hélène Metzger's conception of historically variable a priori structures, the article shows how Collins's appeal to Winch, Kuhn, and the indeterminacy of experimental practice makes possible an empirical treatment of meaning, rationality, and knowledge. Collins's gravitational-wave studies and the experimenters' regress serve as the central case of this transformation. The third part argues that the Strong programme reveals a structural tension in contemporary philosophy: empirical expansion increases philosophy's engagement with scientific practice, but at the same time weakens its disciplinary autonomy.

Keywords: Strong Programme, Harry Collins, historical epistemology, historical a priori, sociology of scientific knowledge, philosophical autonomy

Эмпирическая экспансия и философская автономия: пересмотр "сильной программы" в свете исторической эпистемологии

Соколова Татьяна Дмитриевна – кандидат философских наук,
научный сотрудник.
Институт философии РАН.
Российская Федерация,
109240, Москва,
ул. Гончарная, д. 12, стр. 1;
e-mail: sokolovatd@gmail.com

Статья рассматривает Сильную программу, предложенную Гарри Коллинзом, в рамках исторической эпистемологии. В первой части Сильная программа вписывается в более широкую трансформацию отношений между философией и наукой – от позитивизма к возникновению социологии и science studies. На этом пути философия постепенно утрачивает статус дисциплины, задающей нормативные условия знания, тогда как наука сама превращается в объект эмпирического исследования. Во второй части реконструируется историческое априори Сильной программы. Опираясь на концепцию исторически изменчивых априорных структур Элен Мецжер, я показываю, что обращение Коллинза к Уинчу, Куну и неопределенности экспериментальной практики открывает возможность эмпирического анализа значения, рациональности и знания. Ключевым примером этой трансформации служат

© Tatiana D. Sokolova, 2026

коллинзовские исследования гравитационно-волновых экспериментов и феномен регресса экспериментаторов. В третьей части обосновывается тезис о том, что Сильная программа обнажает структурное напряжение в современной философии: эмпирическая экспансия углубляет вовлеченность философии в научную практику, но одновременно подрывает ее дисциплинарную автономию.

Ключевые слова: Сильная программа, Гарри Коллинз, историческая эпистемология, историческое априори, социология научного знания, философская автономия

In one of his earlier papers Harry Collins opens his reflection on the Strong programme with a question that has accompanied science studies for decades: "Is what I do here philosophy or sociology?" His answer is: "the question is misconceived" [Collins, 2018, p. 67]. At first glance, this remark concerns disciplinary classification. More fundamentally, however, it points to an uncertainty regarding the status of philosophical inquiry itself and its relation to empirical investigation. Collins's account can therefore be read not only as a history of the Strong programme but also as a contribution to a broader discussion about the changing place of philosophy in the production of knowledge.

My perspective on this discussion differs from Collins's. I write as a philosopher rather than a sociologist, and as a member of a later generation for whom the emergence of the Strong programme is already a solid historical fact rather than an intellectual novelty. For Collins, the Strong programme appears as a project in formation. For me, it appears primarily as part of the history of science studies, whose conditions of possibility themselves call for explanation. Rather than revisiting these debates from within the tradition that produced them, I approach them through questions that arise from my own work in epistemology: how do particular forms of knowledge become possible, what historical conditions shape epistemic concepts, and how do disciplinary boundaries themselves emerge and transform? These questions become especially important if, as Timothy Williamson argues, "contemporary philosophy lacks a self-image that does it justice" [Williamson, 2007, p. 8]. The emergence of the Strong programme can be read as one episode in a history of philosophy's attempts to redefine its relation to science and empirical research.

To analyse this episode, I draw primarily on Hélène Metzger's conception of the historical a priori. Unlike conceptions of the a priori as fixed and universal conditions of knowledge, Metzger treats them as historically dynamic structures. As she writes, a priori principles "are to a certain extent plastic; they may change in order to harmonise with the totality of our knowledge, or, on the contrary, they may transform our spontaneous system of the world" [Metzger, 1937, p. 215]. While historical epistemology was originally developed through the study of the natural sciences, I propose to apply it here to the study of science studies



itself. My aim is to describe the Strong programme as the product of a particular configuration of philosophical, sociological, and institutional factors that emerged in British academia during the 1960s and 1970s and to reconstruct the historical a priori underlying it.

As points of departure for this reconstruction, I take the three theoretical sources identified by Collins himself: the Wittgensteinian conception of meaning as use, largely mediated through Peter Winch; the Kuhnian conception of paradigms and scientific communities; and the recognition of the fundamental indeterminacy of experimental practice. Together, these elements challenge fixed meanings, universal standards of rationality, and the assumption that scientific outcomes can be understood independently of the practices through which they are produced. In doing so, they make possible the translation of philosophical questions into empirical investigations. Collins's gravitational-wave study is especially important in this respect because it shows how a classical philosophical problem of induction and the status of replication could be transformed into an empirical investigation of scientific practice.

I argue that the Strong programme forms part of a broader transformation in which philosophy gradually loses both its privileged epistemic status and its autonomy from empirical inquiry. The result is a structural tension: the expansion of empirical investigation increases philosophy's engagement with scientific practice while simultaneously threatening the disciplinary distinctiveness on which its authority traditionally depended.

From Positivism to the Sociology of Science

The intellectual developments described by Collins can be situated within a much longer process concerning the changing relation between philosophy and science. One of its earliest and most influential formulations can be found in the positivist project associated with Auguste Comte. Positivism did not merely propose a new philosophical doctrine: it attempted to redefine philosophy itself by grounding legitimate knowledge in the methods and achievements of the empirical sciences. This reform had important consequences for the status of philosophy. Rather than serving as a discipline that provides foundations for scientific knowledge, philosophy increasingly came to justify itself by reference to science. Scientific inquiry became the model of epistemic legitimacy, while philosophy was expected to clarify, systematise, or generalise results produced elsewhere. The authority to determine what counts as knowledge gradually shifted from philosophy to the sciences themselves.

The tendency to regard empirical inquiry as the privileged route to knowledge remains influential: the significance of positivism lies not

in any specific doctrine but in the disciplinary realignment it initiated. Once science becomes the normative model of knowledge, philosophy finds itself in a derivative position. The question is no longer how science should be grounded philosophically, but how philosophy should relate to scientific knowledge. This reversal constitutes one of the historical conditions that later make possible the emergence of sociology of science and, eventually, the Strong programme itself. The transformation was often carried out not by philosophers adapting science but by scientifically trained thinkers entering philosophy.

The development of sociology introduced a new way of approaching science. Whereas philosophy traditionally investigated science from within – through questions concerning knowledge, rationality, truth, and justification – sociology approached it as one social institution among others. The significance of this development lies in the fact that science could now be studied without direct philosophical mediation. Questions that had traditionally belonged to epistemology could increasingly be reformulated as questions about institutions, communities, professional practices, and social organisation. The study of science gradually acquired its own empirical methods and explanatory frameworks. Yet this separation was never complete. From the perspective of historical epistemology, empirical investigations of science continue to rely on conceptual assumptions that they do not themselves establish. Hélène Metzger emphasised this point and argued, "although the history of science cannot replace philosophy, neither can it entirely abandon the a priori without becoming inconsistent and vague" [Metzger, 1935, p. 5]. Historical and empirical investigations may provide evidence, but they cannot entirely determine the conceptual framework within which that evidence becomes meaningful.

This observation is particularly important for understanding the later emergence of science studies. The sociology of science appears to externalise the study of knowledge by treating science as an empirical object. Yet the very possibility of such an object presupposes conceptual distinctions inherited from philosophy. The result is not a replacement of philosophical inquiry by empirical investigation, but a new and increasingly unstable relationship between them. It is within this space that the Strong programme will later emerge.

The Strong programme is usually discussed as a theoretical position within the sociology of scientific knowledge. Here, however, I treat it as an object of historical epistemology. This approach draws on Hélène Metzger's notion of the historical a priori. For Metzger, the a priori should not be understood as a set of timeless principles preceding experience. Rather, "the a priori also consists in the fundamental tendencies that generate these notions" [Metzger, 1936, p. 33]. The task of historical epistemology is therefore not to identify universal foundations of knowledge but to reconstruct the historically specific tendencies that shape particular



forms of thought. Different intellectual formations may therefore rest upon different historical a priori, generating distinct conceptions of knowledge, rationality, and scientific practice. The Strong programme can be interpreted in precisely these terms. Rather than treating its central assumptions as self-evident or purely theoretical commitments, I suggest understanding them as expressions of a particular historical configuration. From this perspective, its emergence becomes a case study in the historical transformation of epistemic concepts and in the changing relationship between philosophy and empirical inquiry.

The Historical A Priori of the Strong Programme

If the Strong programme is approached as a historically specific epistemic formation, the next task is to identify the conceptual resources that made it possible. Taken together, these elements constitute what may be described as the historical a priori of the Strong programme.

The first component concerns the relation between meaning and action. Collins explicitly identifies Winch as a decisive influence, summarizing his position as follows: "What Winch argues, totally convincingly as I see it, is that peoples' actions and peoples' concepts are two aspects of the same thing" [Collins, 2018, p. 67]. Consequently, "you cannot understand the actions without the concepts and you cannot understand the concepts without the actions" [Ibid.]. Meaning is not something that exists independently of practice and is subsequently applied to it. Rather, meanings and practices are mutually constitutive. This position ultimately derives from Wittgenstein's analysis of rule-following. As Winch emphasises, "establishing a standard is not an activity which it makes sense to ascribe to any individual in complete isolation from other individuals" [Winch, 1990, p. 33]. Rules and meanings are therefore social phenomena sustained by collective practices. This provides one of the central premises of the Strong programme.

The second component is provided by Kuhn's analysis of scientific paradigms. Kuhn's significance for the Strong programme lies in his conception of scientific communities as historically specific traditions of practice: "some accepted examples of actual scientific practice – examples which include law, theory, application, and instrumentation together – provide models from which spring particular coherent traditions of scientific research" [Kuhn, 2012, p. 10]. Scientific communities are therefore organised around shared exemplars, practices, and standards rather than abstract methodological rules alone.

The third component concerns the indeterminacy of experimental practice. Collins repeatedly emphasises that experimental outcomes do not automatically determine their own interpretation. Experimental work

depends upon tacit skills, local judgments, and practical decisions that cannot be completely formalised. Although this theme emerges from studies of scientific practice rather than from philosophy in the narrow sense, it reinforces the implications of both Winch and Kuhn. Just as meanings depend on shared practices and paradigms depend on scientific communities, experimental results depend on forms of practical competence that exceed explicit methodological prescriptions.

Viewed separately, these three elements originate in different intellectual traditions. Viewed together, however, they perform a common epistemic function. Each challenges the assumption that knowledge can be grounded in fixed meanings, universal standards of rationality, or self-interpreting empirical facts. The Strong programme emerges at the intersection of these developments, transforming them into a research programme centred on the empirical investigation of knowledge itself. In this configuration, meaning, rationality, and knowledge become not only philosophical categories but also empirical objects of investigation. Meaning becomes inseparable from social practices, standards of rationality from scientific communities, and empirical facts from practical competences.

Collins's study of gravitational-wave detection gives this indeterminacy its most precise form. In the controversy over Joseph Weber's detectors, the issue was not simply whether particular experimental results were correct. The more fundamental problem was that the criteria of correctness themselves were unsettled. First, Collins concludes that "no set of 'scientific' criteria" could establish the validity of the findings in this field [Collins, 1975, p. 215]. He later formulates the point as the "experimenters' regress": to decide whether gravitational waves have been detected, one needs a good detector; yet to decide whether a detector is good, one needs to know what result it ought to produce [Collins, 1985, p. 84]. Finally, he extends this case into a long-term study of scientific seeing: in gravitational-wave detection, the signal passes through a long chain of instruments, transformations, statistical procedures, and acts of trust. Therefore, as Collins writes, "to 'see' in a science like gravitational wave detection is to agree, at some point, to cease to doubt the chain of inference – to see, then, is to agree to see" [Collins, 2004, p. 7]. The case is therefore not merely an example of experimental error or technical uncertainty. It shows how evidence, replication, and competence become mutually dependent within a scientific community. Collins makes this explicit when he writes that the definition of a good gravitational-wave detector and the question of whether gravitational waves exist are "congruent social processes" [Collins, 1985, p. 89]. From the perspective of historical epistemology, this case shows how a broader intellectual tendency – the empirical treatment of meaning, rationality, and knowledge – makes possible a research programme in which classical epistemological questions are investigated through concrete scientific



practices. Read through this lens, the problem of evidence in Collins's work is no longer addressed as a purely philosophical problem of justification, but as an empirical problem concerning the practices through which scientific communities stabilise facts.

The originality of the Strong programme lies in translating philosophical questions into the empirical study of practices. The movement toward practice was realised not primarily by philosophers but by sociologists, historians, and scholars of science studies. The study of scientific activity became increasingly empirical, while philosophy gradually lost its privileged position as the discipline responsible for defining the conditions of knowledge. This development is particularly striking because many of the concepts mobilised by the Strong programme have unmistakably philosophical origins. Yet arguments originally formulated within philosophical discussions of meaning and rule-following eventually became resources for empirical investigations of scientific communities and expertise.

The result is a peculiar reversal. Philosophical concepts remain indispensable, but philosophy itself no longer monopolises their interpretation or application. Questions that were once treated as internal to philosophy increasingly become objects of empirical inquiry. Philosophy continues to provide conceptual resources, yet it no longer controls the terms under which those resources are deployed. The empirical turn therefore does not eliminate philosophy; rather, it displaces it from its previously privileged position.

Empirical Expansion and Philosophical Autonomy

If the Strong programme translates philosophical questions into empirical investigations, it also transforms the conditions under which philosophy operates. Neither Collins nor Williamson regards empirical inquiry as a threat to philosophy. Yet once questions concerning meaning, rationality, and knowledge become empirical questions about practices and institutions, philosophy becomes increasingly dependent on forms of investigation whose authority it can neither fully appropriate nor entirely ignore.

I propose to describe this situation as an inverse relation between empirical expansion and philosophical autonomy. The more philosophy incorporates empirical methods and adopts the investigative strategies of the social sciences, the more successfully it participates in the analysis of scientific knowledge and expertise. Collins explicitly endorses such an orientation, arguing that "there is also nothing wrong with getting out of the armchair and doing experiments and what are, in regular academic

parlance, referred to as 'sociological' investigations of one group of experts or another" [Collins, 2018, p. 68]. He further insists that "we must be careful that our loyalty to our disciplines, and especially our loyalty to particular corners of our disciplines, does not divert us from using any tool that helps us learn about, in this case, expertise" [Ibid.]. From this perspective, disciplinary boundaries appear secondary to the pursuit of knowledge.

Yet this movement has a consequence. As philosophy increasingly adopts empirical methods, its distinctively philosophical status becomes more difficult to specify. The investigation of expertise, scientific practice, or technological systems may remain philosophically motivated, but its procedures increasingly resemble those of sociology, anthropology, or science studies. Philosophy gains empirical relevance at the cost of disciplinary distinctiveness.

The opposite movement generates a symmetrical difficulty. If philosophy seeks to preserve its autonomy by maintaining a primarily conceptual mode of inquiry, it risks distancing itself from precisely those domains in which knowledge is being produced and studied. The preservation of conceptual inquiry therefore remains necessary, but conceptual autonomy alone no longer guarantees epistemic authority in the study of science. The resulting tension should not be understood as a temporary methodological disagreement or as a problem that can be solved through a better balance between disciplines. Rather, it constitutes a structural condition of contemporary philosophy. The expansion of empirical inquiry simultaneously increases philosophy's access to scientific practice and weakens the autonomy that traditionally distinguished philosophical reflection from empirical research. Conversely, the preservation of philosophical autonomy tends to reduce philosophy's direct involvement in the empirical study of knowledge. What is gained on one side is partially lost on the other.

The significance of the Strong programme lies precisely in making this condition visible: it exemplifies a larger transformation in which philosophy becomes increasingly dependent upon empirical investigations while remaining unable to dissolve itself entirely into them. This condition emerged historically through the positivist subordination of philosophy to science, the sociological externalisation of science as an object of inquiry, and the Strong programme's operationalisation of empirical approaches to knowledge. Kuhn's characterisation of scientific revolutions is instructive in this regard. Reflecting on major conceptual transformations, he writes that each of them "transformed the scientific imagination in ways that we shall ultimately need to describe as a transformation of the world within which scientific work was done" [Kuhn, 2012, p. 111]. The developments discussed here may be understood in a similar way. They do not merely introduce new theories of knowledge; they alter the intellectual environment within which philosophical reflection on knowledge takes place.



This situation gives rise to two responses, each of which may be understood as a form of philosophical courage. The first consists in leaving the ivory tower and embracing empirical investigation. Philosophy enters into closer cooperation with sociology, science studies, anthropology, and related disciplines. It seeks to understand knowledge through the study of actual practices rather than through conceptual analysis alone. The advantage of this strategy is clear: it grants direct access to the empirical realities of scientific life. The cost is less obvious but equally significant. As philosophy adopts the methods and objects of empirical disciplines, its own disciplinary distinctiveness becomes increasingly difficult to define. The second response moves in the opposite direction. Philosophy preserves its conceptual autonomy and insists upon forms of inquiry that cannot be reduced to empirical investigation. It continues to ask questions concerning meaning, rationality, justification, and normativity that exceed the scope of descriptive research. Yet this strategy also has a cost. By limiting its engagement with empirical developments, philosophy risks weakening its authority in discussions concerning scientific knowledge itself. Neither position eliminates uncertainty. The difficulty therefore does not arise from a failure to choose the correct methodological strategy. It arises from the fact that philosophy itself occupies a position between conceptual reflection and empirical inquiry.

For this reason, the opposition between these two responses should not be interpreted as a choice between alternative strategies. They represent two poles of a single condition. The movement toward empirical investigation increases philosophy's relevance to the study of scientific practice while simultaneously eroding the boundaries that distinguish philosophy from neighbouring disciplines.

The tension identified here should therefore be understood not as an accidental difficulty but as a constitutive feature of contemporary philosophy. It emerges from the historical transformations that have redefined the relationship between philosophy, science, and empirical research since the nineteenth century.

Conclusion

Reconstructing the historical a priori of the Strong programme reveals more than the origins of a particular research tradition. It reveals a shift in the intellectual status of philosophy itself. The Strong programme appears not as an isolated episode in the sociology of scientific knowledge but as one manifestation of a wider tendency toward the empirical reconfiguration of philosophical questions. Seen in this context, the Strong programme constitutes a historically specific epistemic regime in which philosophical questions concerning meaning, rationality, knowledge, and

expertise are reformulated as empirical questions concerning practices, communities, and institutions. This transformation was prepared by a longer historical trajectory extending from positivist attempts to align philosophy with science, through the emergence of sociology as an empirical study of social forms, to the development of science studies as a field in its own right.

The consequences of this transformation extend beyond the sociology of scientific knowledge. As philosophical problems become increasingly accessible through empirical investigation, philosophy finds itself operating under a structural condition that it cannot fully overcome. The expansion of empirical inquiry strengthens philosophy's engagement with scientific practice while simultaneously weakening its disciplinary autonomy. Conversely, the preservation of conceptual autonomy risks reducing philosophy's authority in domains increasingly shaped by empirical forms of knowledge. The inverse relation between empirical expansion and philosophical autonomy is therefore not a temporary methodological difficulty but a constitutive feature of philosophy's contemporary situation.

Metzger's observation here remains highly relevant. She suggested that "the study of the history of science would cure the philosopher of the strange habit of positing, whether a priori or a posteriori, definitive concepts upon which the mind seeks to ground its desire for certainty" [Metzger, 1935, p. 13]. The history of the Strong programme suggests that philosophy itself may be subject to the same lesson. From this perspective, the significance of the Strong programme lies not only in what it says about science but also in what it reveals about philosophy itself. Its history makes visible the changing relationship between conceptual reflection and empirical investigation and, more broadly, the change of philosophy's place within contemporary knowledge production. Its concepts, methods, and disciplinary boundaries are not fixed foundations of knowledge but historically changing formations whose transformation remains an object of philosophical reflection.

References

Collins, 1975 – Collins, H.M. "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics", *Sociology*, 1975, vol. 9, no. 2, pp. 205–224.

Collins, 1985 – Collins, H.M. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage, 1985.

Collins, 2004 – Collins, H.M. *Gravity's Shadow: The Search for Gravitational Waves*. Chicago: University of Chicago Press, 2004.

Collins, 2018 – Collins, H. "Studies of Expertise and Experience", *Topoi*, 2018, vol. 37, pp. 67–77. DOI: 10.1007/s11245-016-9412-1.



Collins, 2026 – Collins, H. "The Bath School and the Strong Programme", *Epistemology & Philosophy of Science*, 2026, vol. 63, no. 3, pp. 22–32.

Kuhn, 2012 – Kuhn, T.S. *The Structure of Scientific Revolutions* (50th Anniversary Edition). Chicago: University of Chicago Press, 2012.

Metzger, 1937 – Metzger, H. "La méthode philosophique dans l'histoire des sciences", *Archeion*, 1937, vol. 19, pp. 204–216.

Metzger, 1936 – Metzger, H. "L'a priori dans la doctrine scientifique et l'histoire des sciences", *Archeion*, 1936, vol. 18, pp. 29–42.

Metzger, 1935 – Metzger, H. "Tribunal de l'histoire et théorie de la connaissance scientifique", *Archeion*, 1935, vol. 17, pp. 1–14.

Williamson, 2007 – Williamson, T. *The Philosophy of Philosophy*. Malden; Oxford; Carlton: Blackwell Publishing, 2007.

Winch, 1990 – Winch, P. *The Idea of a Social Science and Its Relation to Philosophy*. London: Routledge, 1990.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 56–66
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 56–66
DOI: https://doi.org/10.5840/eps202663340

AI AND THE TRANSFORMATION OF SCIENTIFIC KNOWLEDGE: HYBRID EXPERTISE AND THE LIMITS OF COLLINS'S THEORY*

Asya A. Filatova – PhD in Philosophy,
Senior Researcher.
Moscow Institute of Physics and Technology (MIPT).
9 Institutskiy Per.,
Dolgoprudny, Moscow Region,
141701, Russian Federation;
e-mail: filatova.aa@mipt.ru



This article proposes a rethinking of the role of artificial intelligence in contemporary science through an analysis of Harry Collins's concept of expertise. The point of departure is the observed tension between normative accounts of expertise – grounded in ideas of socialisation, tacit knowledge and membership in a form of life – and the empirical practices in which AI systems increasingly perform epistemically significant functions. The article contends that traditional deficit-oriented critiques of AI, which focus on the absence of human properties, are losing their explanatory power in the context of the widespread adoption of large language models. The article further argues that with the integration of AI, expertise ceases to be primarily an individual property and increasingly emerges as a feature of hybrid socio-technical systems, whose epistemic capacities arise from the interaction between human and non-human agents. In conclusion, the paper demonstrates the need to reconfigure the anthropocentric conceptual apparatus of science studies to enhance its explanatory adequacy. It proposes a shift in research focus toward the empirical analysis of AI's actual use in scientific practice and the reconstruction of the emergent normative orders within these systems.

Keywords: science studies, expertise, artificial intelligence, large language models, hybrid expertise

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ТРАНСФОРМАЦИЯ НАУЧНОГО ЗНАНИЯ: ГИБРИДНАЯ ЭКСПЕРТИЗА И ГРАНИЦЫ ТЕОРИИ Г. КОЛЛИНЗА

Филатова Ася Алексеевна – кандидат философских наук, старший научный сотрудник.

В статье предлагается переосмысление роли искусственного интеллекта в современной науке через анализ концепции экспертизы Гарри Коллинза. Отправной точкой выступает

* The author acknowledges the support of this research by the Ministry of Science and Higher Education of the Russian Federation under agreement No. 075-03-2026-305 (January 16, 2026), associated with project "Applied Research on the Implementation of Artificial Intelligence Technologies in Higher Education" (project code: FSMG-2025-0086).

© Asya A. Filatova, 2026

Московский физико-технический институт.
Российская Федерация,
141701, Московская область,
г. Долгопрудный,
Институтский пер., д. 9;
e-mail: filatova.aa@mipt.ru

наблюдаемое расхождение между нормативными представлениями об экспертизе, укорененными в идеях социализации, неявного знания и принадлежности к форме жизни, и эмпирическими практиками, в которых системы ИИ все чаще выполняют эпистемически значимые функции. Показано, что традиционная «дефицитарная» критика ИИ, сосредоточенная на отсутствии у него человеческих свойств, утрачивает объяснительный потенциал в условиях распространения больших языковых моделей. В статье утверждается, что с внедрением ИИ экспертиза перестает быть преимущественно индивидуальным свойством и все в большей степени проявляется как характеристика гибридных социотехнических систем, в которых эпистемические способности возникают из взаимодействия человеческих и нечеловеческих агентов. В заключение обосновывается необходимость пересборки антропоцентричного теоретического аппарата исследований науки для повышения его объяснительной адекватности. Предлагается сместить исследовательский фокус на эмпирический анализ практик взаимодействия с ИИ и реконструкцию возникающих нормативных порядков.

Ключевые слова: исследования науки, экспертиза, искусственный интеллект, большие языковые модели, гибридная экспертиза

Harry Collins's article can hardly be described as polemical; rather, it is a programmatic text that takes the reader back to the history of the formation and development of science studies through an autobiographical narrative. The author consistently reconstructs the conceptual shifts that social studies of science have undergone over more than half a century, from Ludwik Fleck and Thomas Kuhn to contemporary theories of scientific expertise. The text refers to the local disputes of the early 1990s between the sociology of scientific knowledge (SSK) and the 'French school' as represented by B. Latour and M. Callon – disputes documented in what have since become classic contributions [Collins, Yearley, 1992; Callon, Latour, 1992]. It also revisits the reasons for the separation of the Bath school's Empirical Programme of Relativism (EPOR) from the Strong Programme of the Edinburgh Science Studies Unit. This first-person reconstruction makes it possible to revisit well-known debates on methodological and ontological symmetry. It also allows us to trace the influence of philosophy on sociological theories, particularly that of the later Wittgenstein. Furthermore, we can follow the rise and decline of constructivism and relativism. Finally, it returns us to the enduring problem of the crisis of scientific expertise and its implications for the democratic order [Tukhvatulina, 2025].

In the concluding part, Collins turns to the contemporary context and formulates a thesis about new threats that may lead modern democracies into a 'dark age for science'. For many defenders of scientific knowledge, Donald Trump has come to symbolize this threat: right-wing populism

and the ideology of post-truth. These forces undermine science as a truth-oriented professional community more profoundly than they undermine any other social institution. Unlike contemporary populists and the 'merchants of doubt' (Oreskes and Conway), Collins emphasizes that the aim of science studies has never been to 'attack science and level it down', but rather 'to understand it better' [Collins, 2026, p. 27]. Another issue, however, is that parts of the scientific community perceive attempts to open the black boxes of science as a form of iconoclasm. As Latour noted, modern science in this respect resembles the church: the establishment of scientific truth is presented as a process in which 'no human hand has ever touched the image it has produced', for any suggestion of human involvement would risk profaning the sanctity of objectivity [Latour, 2002, p. 7]. The tension between two risks, being accused of relativism and undermining truth on the one hand, and appearing socially uncritical and epistemically naive in the interpretation of scientific practices and institutions on the other hand, continues to define the core tension of science studies today.

Against the background of this retrospective and rearticulation of science studies, one is struck by a conspicuous absence. The missing topic is one that not only commands the attention and sociotechnical imagination of virtually all actors – researchers, policymakers, and laypeople alike – but also foreshadows profound transformations both in scientific practices and in their study. That topic is artificial intelligence (hereafter AI). The omission is especially striking because the critique of AI has been central to Collins's work for several decades, particularly in the development of his theory of expertise. One might assume that it reflects the editorial framing of the present discussion. However, when considering the factors that shape transformations in our understanding of science, the populism associated with Trump and other political actors appears no more significant than the rapid, pervasive spread of AI technologies – in particular, large language models (hereafter LLMs) – even if the normative valence of these influences differs radically. In this context, it seems justified, if not entirely 'legitimate' within the given discussion, to expand its scope: to turn to Collins's interpretation of AI as a potential co-actor in scientific research and to examine the normative constraints on attributing expert status to such systems. This direction offers a productive line of inquiry, as it speaks both to the transformation of scientific practices and to the broader question of how existing forms of scientific life may be reconfigured or extended through human – AI collaboration.

The argument is organized around three main claims. First, AI is transforming the production of scientific knowledge in ways that go beyond quantitative gains in speed or scale: it reconfigures the rules, epistemic regimes, and hierarchies that organize research practices and communities. Second, AI increasingly functions as a component of hybrid expert systems, as evidenced by a growing body of empirical research.

This raises a set of questions: how such configurations are formed at infrastructural and normative levels, what grounds their epistemic advantage, and what makes such systems count as experts – questions that cannot be adequately addressed within anthropocentric models of expertise. Third, science and technology studies must rethink their conceptual apparatus: core notions such as linguistic competence, socialisation, and understanding require revision when applied to heterogeneous socio-technical systems that produce, evaluate, and stabilize knowledge.

In the early 2000s, after a prolonged period of the so-called science wars, marked by accusations of relativism and an excessive focus on the 'deficiencies of scientific objectivity', Latour, not without a degree of self-critique, suggested that in science studies 'critique has run out of steam' [Latour, 2004]. A similar diagnosis applies to another contemporary situation: the wave of AI critique emerging from the humanities. Drawing on what has been described as the so-called 'religion of justaism' [Aaronson, 2024], scholars devote significant attention to demonstrating that AI is merely a 'stochastic parrot', a next-token predictor, a gigantic autocomplete, and so on. Accordingly, critics argue that AI is not X (an agent, an expert, an authority, etc.) because it lacks Y (free will, understanding, communicative intentions, and so forth).

These responses are not entirely unjustified. They arise from a striking tension between theoretical commitments and empirical observations. On the one hand, language models are increasingly capable of performing tasks that, in human contexts, presuppose the presence of mental states, intentions, or desires: they sustain conversations, generate original ideas, and provide critical feedback. On the other hand, these systems do not possess the very capacities that are traditionally taken to ground such performances. The need to account for this behavior – uncannily similar to, yet fundamentally distinct from, that of humans – produces, for some researchers, a kind of 'uncanny valley' effect.

However, if this reaction remains at the level of debunking, it risks becoming counterproductive. Framing AI outputs exclusively in terms of fiction, illusion, or simulation may lead not only to underestimating their advantages but also to failing to recognize emerging risks. Conceptual biases toward non-human systems may prevent adequate preparation for their societal and epistemic implications, as long as we continue to reassure ourselves that such systems are merely 'automatic', 'algorithmic', or 'computational' – lacking genuine existence and incapable of producing meaning [Shumer, 2026]. As Marie Theresa O'Connor aptly observes, 'there may be unforeseen consequences to emergent tension between theory and practice – that is, between people holding that AIs cannot make meaning and sense by definition and, at the same time, acting in daily life as if they do' [O'Connor, 2025, p. 5]. This mismatch between critique and practice is not merely theoretical; it reflects deeper transformations in what counts as knowledge and how it is – or ought to be – produced.

Does science itself change with the advent of generative AI? The preceding discussion has pointed to a growing tension between established theoretical frameworks and emerging empirical practices. I argue that this tension is already reflected in the organization of scientific work and is likely to intensify. New technologies are beginning to reshape how research groups are structured, how they operate, and how authority is distributed within them. Collins describes scientific communities, in his account, as 'families or small communities: they know each other well, speak the same language, share common assumptions, and are cautious about outsiders' [Collins, 2025, p. 1255]. Neural networks may disrupt this familiar and psychologically intuitive picture.

These transformations unfold unevenly. Their trajectory depends on existing epistemic cultures and is often accompanied by resistance, self-deception, or strategic masking. Yet their cumulative effect is becoming increasingly difficult to ignore. The first claim I advance is that AI is transforming established research practices, decision-making processes, and the distribution of authority within scientific communities. Critiques of AI have a long history [Dreyfus, 1972; Searle, 1980; Collins, 1990]. However, many of these arguments have lost some of their force in light of recent developments in deep learning and LLMs, which Collins refers to as 'NEWAI' [Collins, 2025, p. 1253]. These systems are capable of acquiring and adjusting rules through interaction and, compared to earlier generations of AI, perform more convincingly in tasks traditionally associated with human intelligence, including variants of the Turing test. As suggested in recent work by DeepMind researchers, future AI agents are expected to address current limitations by integrating world models, more effective simulation environments, and forms of embodied learning, thereby enhancing their situational awareness and ability to respond to dynamic contexts [Marchal et al., 2026, p. 3].

To illustrate the claim that AI can become part of contributory expertise, consider the joint project of Google DeepMind and Lawrence Berkeley National Laboratory. In this project, researchers integrated the GNoME (Graph Networks for Materials Exploration) system with the autonomous robotic platform A-Lab, forming a closed-loop research system that combines hypothesis generation, experimental synthesis, and their iterative validation [Merchant, Cubuk, 2023]. In Collins's sense, contributory expertise presupposes the capacity to participate directly in the production of new knowledge rather than merely to interpret or reproduce it. In this respect, the GNoME – A-Lab configuration is particularly revealing, as it does not simply assist human researchers but actively contributes to the generation and validation of novel scientific results. GNoME is a deep learning model based on graph neural networks, in which crystalline structures are represented as graphs of atoms and learned from large-scale datasets such as the Materials Project. It operates in an iterative active learning regime, generating candidate structures,



evaluating their stability, and using the resulting data to further refine the model. Importantly, the architecture of GNoME brings together two types of heuristics, structural and compositional, allowing the system to explore the space of materials not merely as a set of possible combinations, but as a domain of potentially significant scientific objects.

Prior to GNoME, approximately 20 000 experimentally identified crystals and around 48 000 computationally stable materials were known. Within a relatively short period, the system generated 2,2 million candidate structures, of which approximately 381 000 are considered stable. Crucially, a substantial proportion of these are not variations of known compounds but new structural prototypes that had not previously entered the field of scientific attention, thereby expanding not only the empirical base but also the very space of possible objects of inquiry in solid-state chemistry.

For the experimental validation of these candidates, the team employed the A-Lab platform, an autonomous robotic laboratory capable of performing the full synthesis cycle: from selecting target compounds and generating synthesis protocols to conducting reactions and analyzing the results via X-ray diffraction, including the adjustment of procedures in cases of failure. Trained on more than 30 000 published synthesis recipes, A-Lab is able not only to reproduce existing practices but also to adapt them in the course of operation. In one key experiment, the system successfully synthesized 41 out of 58 proposed materials within 17 days, significantly compressing the temporal horizon of research.

This case reveals a shift in the organization of 'laboratory life'. Where the search for new materials had previously relied on extended trial-and-error processes and depended on the embodied experience of researchers and their tacit knowledge, the combination of AI and robotics now replaces not only the 'mechanical' but also the cognitive components of the research process. As a result, this reconfigures the distribution of epistemic functions: the human actor is increasingly displaced from the position of a direct producer of knowledge toward that of a goal-setter and operator that oversees the parameters and outputs of the system.

The second claim follows from the first. If AI performs epistemic functions in the production of scientific knowledge and is capable of sustaining professional-level interaction, then it reshapes not only our understanding of expertise but also the procedures through which expertise is recognized and enacted. Collins has devoted several books and numerous articles to the critique of AI and its alleged capacity to qualify as an expert [Collins, 1990; Collins, Evans, 2002; Collins, 2018]. His position warrants careful attention. It has never been alarmist, has remained methodologically sophisticated, and has evolved in response to technological change. The emergence of generative AI has not gone unnoticed in his work. Yet Collins remains consistent in his central claim. In his 2025 paper, he continues to argue that despite significant advances

in the AI cognitive and communicative capacities, including the ability to adjust communicative rules in interaction, such systems cannot share a form of life with human scientists. This is because they do not undergo primary socialisation, through which moral commitments and a corresponding 'sense of truth' and 'sense of stability' are formed [Collins, 2025, p. 1256]. According to Collins, what might be termed as the 'socialisation' of AI begins at the level of secondary socialisation – that is, the acquisition of specialist skills. However, this process remains confined to textual corpora and thus lacks the tacit social knowledge necessary for full membership in a linguistic community. Consequently, on normative grounds, Collins argues that AI systems do not qualify as experts [Voevodenko, 2025].

At the same time, empirical evidence suggests that users increasingly turn to LLMs precisely as experts. This is particularly evident in domains such as medicine, psychological counseling, law, and economic advice [Seabrooke et al., 2024; Cross et al., 2024]. Several explanations can be proposed. In some domains, expertise is institutionally structured in ways that constrain or formalize tacit knowledge, for example, by requiring strict adherence to standardized protocols. Under such conditions, expert performance becomes more readily reproducible by AI systems, and users may not perceive a substantive difference between human- and machine-delivered expertise. Moreover, AI-based consultants may be preferred due to their constant availability, low cost, and capacity to sustain extended, responsive interaction. The success of AI, however, cannot be reduced to the prior algorithmization of expert practices. Contemporary AI systems also introduce novel affordances. For instance, they enable forms of personalization in domains such as medicine that were previously unattainable [Mishra et al., 2025].

The growing reliance on AI in these contexts further intensifies the tension between normative accounts of expertise and actual practices of its attribution. Empirical observations indicate that, even if 'socialisation is far more mysterious' [Collins, 2017, p. 142] than what is exhibited by large language models, users nonetheless treat these systems as interactional experts, effectively engaging them in what Collins conceptualizes as the Imitation Game.

An anthropocentric bias leads us to evaluate non-human systems according to criteria that are not always appropriate to them. Neural networks do not learn in the same way as humans; their ontology is different. Their 'training' consists in the acquisition of patterns of linguistic and epistemic behavior encoded in large-scale textual data. Collins, discussing GPT-4, notes that its pre-training was completed in 2021 and concludes that such systems inevitably lag behind the current state of knowledge. This argument becomes less decisive when we consider actual usage practices. Contemporary models are continuously updated, fine-tuned, and embedded in more complex socio-technical configurations.



In particular, retrieval-augmented generation (RAG) architectures allow systems to access external, updatable data sources and integrate retrieved information into the generative process.

Although AI systems do not undergo human-like socialization and likely lack mental states, intentions, or moral sensibilities in the human sense, this does not mean that they operate outside the normative domain. On the contrary, alignment processes, including reinforcement learning from human feedback (RLHF) and various forms of fine-tuning, systematically incorporate normative expectations into their behavior. For example, Anthropic has designed its Claude model with an explicit 'persona' structured around epistemic and ethical virtues, enabling it to demonstrate competence not only in matters of factual accuracy but also in morally sensitive communication [Anthropic, 2024].

This argument does not deny that AI systems fail to meet certain human-centered criteria of epistemic authority or testimonial reliability [Ferrario et al., 2024; Fricker, 2025]. However, such criteria may not be required for the performance of epistemically significant actions or the attainment of reliable outcomes. In light of this, expertise should be understood not as an individual property but as a feature of hybrid socio-technical systems. Following Croce, one can attribute epistemic capacities to a collective agent if (1) its members jointly pursue a given epistemic goal, and (2) the system demonstrates a stable ability to achieve that goal [Croce, 2019, p. 7]. Reformulating Collins's question, one might ask whether all hybrid systems qualify as experts. The answer is no. Expertise requires regularity, reproducibility, and a commitment, on the part of each component, to a certain level of epistemic virtue. In the case of AI, this commitment can be understood as the degree to which systems reliably sustain user-oriented functions that are epistemically beneficial, whether designed in advance or emerging through interaction [Ferrario et al., 2024]. Trust within such hybrid collectives depends on the alignment of AI systems with human epistemic goals and on the strengthening of socio-epistemic infrastructures that support virtues such as honesty, truthfulness, and truth-seeking [Marchal et al., 2026].

The limited effectiveness of applying concepts developed to account for individual and social human behavior, such as understanding, communicative intention, or desire, to the analysis of hybrid expert systems leads directly to the third claim. As suggested by Latour's well-known metaphor, social scientists who continue to rely on an inherited conceptual vocabulary risk preparing for the previous war rather than the present one. Traditional critiques of AI no longer appear fully convincing, neither to those who design such systems nor to those who engage with them in everyday practice. At the very least, they are not compelling enough to prevent users from interacting with LLMs in contexts where these systems are experienced as meaningful and useful.

What may be described as a deficit-oriented mode of critique, in which the primary task is to demonstrate what AI systems lack in comparison to humans, loses much of its explanatory power under current conditions. In this sense, Latour's well-known question becomes newly relevant: 'Are we not like those mechanical toys that endlessly make the same gesture when everything else has changed around them?' [Latour, 2004, p. 225]. A continued reliance on familiar rhetorical strategies and conceptual oppositions not only limits our ability to grasp the emerging possibilities of socio-technical assemblages but may also obscure the risks they entail. There is no straightforward solution to this problem. The development of a new conceptual language is itself a demanding task. One possible strategy is to reconceptualize existing terms by reducing or carefully qualifying their anthropocentric assumptions. Another strategy is to draw on conceptual resources from adjacent fields. A third strategy is to return to classical theories and identify elements that can be productively reinterpreted in light of current transformations. Each of these approaches has its own strengths and limitations.

What appears more immediately necessary for science and technology studies is to return, in a revised form, to the empirical orientation that originally defined the field, including in Collins's own work. This entails close attention to what scientists actually do with AI systems and to how hybrid expert configurations are assembled, stabilized, and evaluated in practice. Such an approach makes it possible to move beyond inherited conceptual constraints and to develop a more adequate account of expertise under contemporary socio-technical conditions.

The transformations observable in contemporary science suggest that the central problem posed by AI cannot be adequately reduced to the question of whether machines can become experts in the traditional sense. As AI systems become embedded in professional practice, they do not merely imitate or approximate human expertise but actively participate in the reconfiguration of the epistemic field itself. Recognizing the hybrid character of expertise does not, in itself, resolve the problem; rather, it marks the starting point for further analysis. In particular, it becomes necessary to examine how 'mutual understanding' is established between agents of different kinds; how 'alien content' [Cappelen, Dever, 2024] produced by AI can be interpreted in human terms and incorporated into knowledge practices; how the temporal regimes of different agents shape the coordination of their actions; how authority is distributed within such systems; and through which reference points agreement in definitions and judgments between humans and machines is achieved. These questions, I suggest, are likely to define the agenda of science and technology studies in the near future.

References

Aaronson, 2024 – Aaronson, S. "The Problem of Human Specialness in the Age of AI", *Shtetl Optimized*, 2024. [https://scottaaronson.blog/?p=7784, accessed on: 10.04.2026]

Anthropic, 2024 – Anthropic. "Claude's Character", 2024. [https://www.anthropic.com/news/claude-character, accessed on: 10.04.2026]

Callon, Latour, 1992 – Callon, M., Latour, B. "Don't Throw the Baby Out with the Bath School! A Reply to Collins and Yearley", in Pickering, A. (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press, 1992, pp. 343–368.

Cappelen, Dever, 2024 – Cappelen, H., Dever, J. "AI with Alien Content and Alien Metasemantics", *arXiv:2405.19808*, 2024.

Collins, 1990 – Collins, H.M. *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: MIT Press, 1990.

Collins, 2017 – Collins, H.M. "Interactional Expertise and Embodiment", in: Sandberg, J., Rouleau, L., Langley, A., Tsoukas, H. (eds.), *Skillful Performance: Enacting Expertise, Competence, and Capabilities in Organizations*. Oxford: Oxford University Press, 2017, pp. 125–146.

Collins, 2018 – Collins, H.M. *Artifictional Intelligence: Against Humanity's Surrender to Computers*. Cambridge: Polity Press, 2018. 232 p.

Collins, 2025 – Collins, H. "Why Artificial Intelligence Needs Sociology of Knowledge: Parts I and II", *AI & Society*, 2025, vol. 40, pp. 1249–1263.

Collins, 2026 – Collins, H.M. "The Bath School and the Strong Programme", *Epistemology & Philosophy of Science*, 2026, vol. 63, no. 3, pp. 22–32.

Collins, Evans, 2002 – Collins, H.M., Evans, R. "The Third Wave of Science Studies: Studies of Expertise and Experience", *Social Studies of Science*, 2002, vol. 32, no. 2, pp. 235–296.

Collins, Yearley, 1992 – Collins, H.M., Yearley, S. "Epistemological Chicken", in: Pickering, A. (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press, 1992, pp. 301–326.

Croce, 2019 – Croce, M. "For a Service Conception of Epistemic Authority: A Collective Approach", *Social Epistemology*, 2019, vol. 33, no. 2, pp. 172–182.

Cross et al., 2024 – Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., Alvarez-Jimenez, M. "Use of AI in Mental Health Care: Community and Mental Health Professionals Survey", *JMIR Mental Health*, 2024, vol. 11.

Dreyfus, 1972 – Dreyfus, H.L. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper & Row, 1972

Ferrario et al., 2024 – Ferrario, A., Facchini, A., Termine, A. "Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems", *Minds and Machines*, 2024, vol. 34, article 30.

Fricker, 2025 – Fricker, E. "On the Metaphysical and Epistemic Contrasts between Human and AI Testimony", *Inquiry*, 2025, pp. 1–26.

Latour, 2002 – Latour, B. "What Is Iconoclash? Or Is There a World beyond the Image Wars?" in|: Latour, B., Weibel, P. (eds.), *Iconoclash: Beyond the Image Wars in Science, Religion and Art*. Cambridge, MA: MIT Press, 2002, pp. 14–37.

Latour, 2004 – Latour, B. “Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern”, *Critical Inquiry*, 2004, vol. 30, no. 2, pp. 225–248.

Marchal et al., 2026 – Marchal, N., Chan, S., Franklin, M., Revel, M., Keeling, G., Fischli, R., Chandra, B., Gabriel, I. “Architecting Trust in Artificial Epistemic Agents,” *arXiv:2603.02960*, 2026.

Merchant, Cubuk, 2023 – Merchant, A., Cubuk, E.D. “Millions of New Materials Discovered with Deep Learning”, *Google DeepMind Blog*, 2023. [https://deepmind.google/blog/millions-of-new-materials-discovered-with-deep-learning, accessed on: 10.04.2026]

Mishra et al., 2025 – Mishra, A., Majumder, A., Kommineni, D. et al. “Role of Generative Artificial Intelligence in Personalized Medicine: A Systematic Review”, *Cureus*, 2025, vol. 17, no. 4, e82310.

O’Connor, 2025 – O’Connor, M.T. “AIs as Fellow Participants in the Language Game”, *AI & Society*, 2025, pp. 1–11.

Seabrooke et al., 2024 – Seabrooke, T., Schneiders, E., Dowthwaite, L., Krook, J., Leesakul, N., Clos, J., Maior, H.A., Fischer, J.E. “A Survey of Lay People’s Willingness to Generate Legal Advice Using Large Language Models (LLMs)”, *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems (TAS ’24)*, Austin, TX, September 16–18, 2024. New York: ACM, 2024.

Searle, 1980 – Searle, J.R. “Minds, Brains, and Programs”, *Behavioral and Brain Sciences*, 1980, vol. 3, no. 3, pp. 417–457.

Shumer, 2026 – Shumer, M. Something Big Is Happening, 2026. [https://shumer.dev/something-big-is-happening, accessed on 24.04.2026]

Tukhvatulina, 2025 – Tukhvatulina, L.A. “Ekspertiza posle pandemii” [Expertise after the Pandemic], *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 3, pp. 229–236. (In Russian)

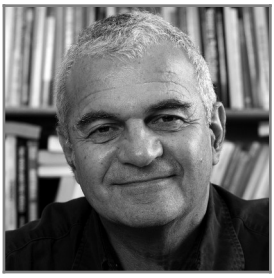
Voevodenko, 2025 – Voevodenko, A.V. “Strategii opredeleniya neyavnogo znaniya: M. Polani, G. Kollinz, S. Terner” [Strategies for Defining Tacit Knowledge: M. Polanyi, H. Collins, S. Turner], *The Digital Scholar: Philosopher’s Lab*, 2025, vol. 8, no. 3, pp. 94–111. (In Russian)

Эпистемология и философия науки
2026. Т. 63. № 3. С. 67–77
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 67–77
DOI: https://doi.org/10.5840/eps202663341

Reply to critics

Harry Collins – FBA.
Honorary Professor.
University College London.
Emeritus.
Cardiff University;
e-mail: CollinsHM@cardiff.ac.uk



I begin my suggesting that the instincts that accompanied early, 1970s, work in the sociology of scientific knowledge – that we were doing something important – have turned out to be sound. Fifty years later responses to an account of those early years lead to deep discussions about artificial intelligence, the nature of sociology and the nature of philosophy. Thereafter, my responses to the discussants are separated into two parts. I disagree with Filatova that recent work in artificial intelligence has produced devices with human-like intelligence and I explain why. She is certainly right in saying that the output of large language models is more readily mistaken for human-like responses than that of previous generations, but this means only that we have to be more ingenious in exploring the ways they differ from us. Thorpe and Sokolova invite a discussion of the trajectories of sociology and philosophy and their relationship to each other. I find that I no longer know whether I am a sociologist or a philosopher though nowadays I have more sympathy with the concerns of philosophy than with what lies at the centre of attention of contemporary sociologists. This invites a discussion of the quintessential sociology and quintessential philosophy, topics which are very unpopular within the constructivist way of thinking. Nevertheless, such an approach is vital if we are to resist the erosion of science and other academic disciplines and, indeed, the erosion of truth itself under contemporary political pressures. I introduce the notion of 'formative aspirations' as a way of defining the essence of science and of activities like sociology and philosophy. I suggest that sociology and philosophy collide in terms of their formative aspirations because both are concerned with the ultimate grounds of truth.

Keywords: Strong Programme, sociology of scientific knowledge, interactional expertise, formative aspirations, philosophy and sociology

Ответ оппонентам

Гарри Коллинз –
член Британской академии.
Почетный профессор.
Университетский колледж Лондона.
Заслуженный профессор.
Университет Кардиффа;
e-mail: CollinsHM@cardiff.ac.uk

Настоящая статья представляет собой ответ Гарри Коллинза на комментарии к работе «Батская школа и Сильная программа». Интуиция, сопутствовавшая ранним работам по социологии научного знания в 1970-е гг., – ощущение значимости предпринятого – оказалась верной: спустя полвека отклики на описание тех первых лет оборачиваются глубокими дискуссиями об искусственном интеллекте, о природе социологии и о природе философии. Ответы участникам обсуждения разделены на две части. Коллинз не согласен с Асей Филатовой в том, что новейшие разработки в области искусственного интеллекта привели к созданию устройств, сравнимых с человеческим разумом, и поясняет почему. Автор признает, что продукты больших языковых моделей ныне гораздо чаще, нежели ранее, можно принять за высказывания человека, однако это лишь означает, что нам следует проявлять больше

© Harry Collins, 2026

изобретательности в поиске того, чем они от нас отличаются. Чарльз Торп и Татьяна Соколова предлагают обсудить траектории развития социологии и философии, а также их взаимоотношения. Коллинз приходит к тому, что более не может сказать определенно, кто он – социолог или философ, хотя сегодня испытывает бóльшую близость к философской проблематике, нежели к тому, что составляет центр внимания современных социологов. Это естественным образом подводит к вопросу о том, что есть подлинная социология и подлинная философия, – вопросу, крайне непопулярному в рамках конструктивистского подхода. И все же подобная постановка вопроса необходима, если мы намерены противостоять эрозии науки и других академических дисциплин, а также – что еще важнее – эрозии самой истины в условиях современных политических реалий. Автор вволит понятие «формативных устремлений» как инструмент для определения сущности науки и таких видов деятельности, как социология и философия. Его гипотеза состоит в том, что социология и философия сталкиваются именно в своих формативных устремлениях, ибо обе обращены к предельным основаниям истины.

Ключевые слова: Сильная программа, социология научного знания, интеракциональная экспертиза, формативные устремления, философия и социология

I thank the three commentators. It is marvellous that a discussion of the early days of science studies could elicit responses that have to do with recent developments in artificial intelligence, with the deep features of sociology, and with the future of philosophy. It felt in those early days that something big was going on and. given the breadth of the responses, it looks like that instinct was sound.

I'll take the responses in the order in which I received them. Filatova sent hers in early and has been, perhaps, unlucky because I disagree with her and she gave me plenty of time to say why. Thorpe and Sokolova invite a different kind of discussion.

Filatova

> Traditional critiques of AI no longer appear fully convincing, neither to those who design such systems nor to those who engage with them in everyday practice. At the very least, they are not compelling enough to prevent users from interacting with LLMs in contexts where these systems are experienced as meaningful and useful [Filatova, 2026, p. 63].

May I say that Filatova's reply to my piece is not really a *reply* because she talks about artificial intelligence (AI) which I don't mention. But that's OK – I have written about AI since the 1980s (three books and quite a few articles) and I do mention expertise in my piece. What she



says is that recent AI has rendered my analysis of expertise out of date since today's computers must be treated as social actors with their own expertise.

Filatova is unlucky in that the field of AI is moving fast and even since she wrote what she wrote another paper of mine on Large Language Models (LLMs) has come out (in *Synthese*), another has been accepted for publication and a third is under submission but posted on arXiv [Collins. Thorne, 2026; Collins et al., 2026a; Collins et al., 2026b]. These papers all deal with the difference between LLMs and human intelligence showing how LLMs fail or are potentially dangerous. All three of these papers are accessible via the internet. I cannot go into the detailed content here except to say that the 'alignment' with human values that Filatova talks about is as much a danger as a benefit since, as we show, it is easy for malign persons to align these devices in malign ways (see the third paper listed above). Instead, I'll take a more distanced look at our differences in approach.

Going to the quotation at the head of this response, I cannot think of a time when the opposite of what Filatova says would be true. Taking the first sentence, AI is an industry depending on sales to the public and stories for venture capitalists and it is never going to welcome criticism (unlike proper sciences). As for the second sentence, it is true that people do continually find themselves treating AI as expert actors but to claim that shows they have finally achieved human-like intelligence is to look through rose-tinted spectacles. When we look at the philosophy/sociology of knowledge, it doesn't matter what properties humans impute to computers because humans are capable of attributing all kinds of properties to all kinds of things, not only AIs but all kinds of non-humans, such as scallops, as Latour and Callon exemplify [Collins, Yearley, 1992]. Worse, in wartime, groups of humans often attribute non-humanness to other groups of humans to make killing them less stressful. In the field of AI, we can go at least as far back as 1976 and Weizenbaum's Eliza to find humans attributing psychotherapeutic abilities to the simplest of programs. Distinguishing LLMs from humans is harder than ever because of their extraordinary fluency with native language but that's not a philosophical argument, it's an observation of how we normally attribute intelligence to others. Latour, the most thoroughgoing of the attributors, has shown that we can attribute agency to anything, so of course there is going to be a space for exploring how the latest generation of AI affects society – and affect it, it will. But to investigate how computers affect society rather than to investigate their intelligence is Filatova's choice. My view is that the interesting thing for a sociologist of knowledge rather than a sociologist of labour, who might just as easily be looking at the introduction of the spinning jenny, is whether the latest generation have really solved the scientific problem of reproducing human intelligence – something that has been claimed for every new generation of AI from the beginning.

In my 1990s book I spend some time asking whether pocket calculators are, really, better arithmeticians than humans and conclude they are not because they don't know how to approximate. Of course, each new generation of AI is better than the last and LLMs *can* approximate. I use ChatGPT all the time but only in the way I use a pocket calculator which, as I explained in 1990, involves having a good idea of the answer first if you are not to make terrible mistakes. My complaint about Latourianism is that it doesn't allow you to ask the critical questions about non-humans' abilities since it attributes the qualities to non-humans that ordinary humans attribute to them in ordinary social life, so it is impossible to say that anyone is getting anything wrong. That is not a club I wish to join but if that's what you want to do then fine. But it won't advance our understanding of what humans and machines can and can't do while making the wrong attributions (note the realist talk), can really muck up social existence for everyone.

Finally, turning to AI-aided science in particular, it is important to understand the science being done. If it is 'hypernormal science' such as molecular biology [Collins et al., 2023] then it is much more open to computerisation than a science like gravitational wave detection (though there is a way of using computers to simulate and explore the terrain of experimental potentialities there too). But that computers can aid science has been obvious since the invention of the slide rule [Collins, Thorne, 2026] distinguish the use of computers in different kinds of science.

Thorpe and Sokolova

I owe a big debt to Charlie Thorpe. Like, I suspect, many adventurous academics I spend quite a bit of time wondering if I am the crazy one, especially when I write papers that I think are important but hardly anyone notices them. One such paper is the one on 'socialness' [Collins, 1998] which Thorpe discusses. This, paper, published in 1998, has been cited 46 times according to Google Scholar, and seven of those are self-citations. For comparison, my most cited item *Changing Order*, is slowly nearing 5000 citations, my tenth most cited paper is approaching 1000 and my twentieth is approaching 500. I'd say the socialness paper should rank in the top ten or twenty of my publications but in terms of citations it's around rank 100. Still worse, is the case of the paper called 'Interactional Imogen', published in 2020, which I consider a really crucial follow-up to the socialness one. It has been cited 11 times, of which 6 are self-citations! In it I try to work out the limits on one social group coming to understand another social group. Again, I believe it comes up with enough important insights to rank among the top 10 or 20 of my publications but is around rank 200 in terms of citations. Please read it!



So, Thorpe's discussion of 'Socialness' offers the possibility that in spite-of these works, which I consider a couple of my very best, being close to invisible in the world I work in, I may not be completely mad. And Thorpe is one the five 'proper' citers of 'Imogen'. So, you see the size of my debt!

To get to the relationship between philosophy and sociology, the socialness paper is a plea for a sociology as I (and Thorpe, and Peter Berger) understand it. This sociology goes back at least to Durkheim and his 'treat social facts as things'. As Thorpe explains, I believe humans are unique because they have 'socialness' – the ability to take part in and understand social groups whereas animals and other non-humans do not; human-like language is central to this capacity and the term 'interactional expertise' captures the ability to understand, through language, the practical face of social capacity without actually being able to execute the practices. Acquiring interactional expertise in the world of an initially unfamiliar social group is hard but, in the unread 'Imogen' paper I argue that it is possible under the right circumstances.

All this is philosophy as much or more than it is sociology, as I see it, albeit I did come up with it as a result of living with scientists during my gravitational wave study. Furthermore, if this is philosophy we are going to have to call Durkheim a philosopher. But why shouldn't we? We also call the later Wittgenstein a sociologist and agree that while Winch was calling sociology 'misbegotten philosophy' he might equally have been calling philosophy (at least, of the late Wittgenstein type which was his topic) 'misbegotten sociology'.[1] That's why I am very pleased to find Sokolova reminding me of what I wrote in 2018 after a somewhat bad tempered encounter with certain philosophers of expertise: "Is what I do here philosophy or sociology?" – "the question is misconceived".

This makes Sokolova's concern about whether philosophy is 'abandoning its birthright', as we might say, by conceding to sociology's empirical approach, very difficult to answer. We might be able to agree that philosophy has changed, but if it has, we can locate the beginning of the change with a philosopher – Wittgenstein. But, in any case, all academic subjects change and not necessarily by conceding ground to other subjects. Let us start with sociology.

[1] [Bloor, 1983] is a good place to find the claim that Wittgenstein was as much sociologist as philosopher: what he was doing was describing the way his society worked, being pretty knowledgeable about it as all citizens are without getting out of their armchairs to do empirical studies.

Sociology has changed

What I have come to realise is that I wrote the 'socialisation' paper as a response to a change in sociology – one that I didn't fully understand at the time. But now that time has passed, and Thorpe has helped me see that this is a topic worth pursuing, the change that I was responding to was from what I still consider to be sociology's main concern, understanding the world through the concept of the social group. Sociology as practised by sociologists was changing to the defence of those who live at the interstices of social groups – those marginal to social groups; sociology's topic had become diversity. The new topic was no longer the science of the social but a political movement, often referred to as 'identity politics'. Identity politics took over from sociology as a science because the main feature of sociological science was the study of groups and the idea of groups seemed to suppress the rights of individuals who were marginal to this group or that group. The study of the old sociology was how groups worked and it concentrated on those who were full members. I'm no kind of historian but I think the driving forces behind the transformation were civil rights and gay liberation because marginality was the dominant political experience to those who were discriminated against, and the idea of social groups seemed to emphasise the marginality of the marginal. Sociology turned into the politics of those in the spotlight of discrimination.

As far as I can see, this is where sociology still finds itself and that is why we should not be surprised that sociologists are not going to read a paper which asks the question about whether one social group can learn to understand another social group and the conditions that make it is possible; they certainly aren't going to cite such a paper in their own work. The question is outlawed by the off-the-shelf political answer: marginal groups can be represented only by members of those groups. What now has to be added, though no-one seems anxious to add it, is that this means anthropology, ethnography and sociology understood as the study of unfamiliar groups by academics specially trained in understanding initially unfamiliar cultures, is impossible: at best, it is disguised colonialism.

As an aside, though the experience of marginality is real enough (my early life was lived as a marginal person), the marginal are usually as much members of social groups as anyone else, which is how they function as members of society. The social identity of being gay or being black is enormously powerful and keeps identity politics going. To understand how socialness works one needs the fractal model of society which locates socialness all the way down to the smallest groups.[2] That is the scientific concept, but it is obvious that it doesn't stop groups being

[2] For the fractal model see, for example, [Collins, 2023, Figure 1].



in conflict and smaller groups being discriminated against by dominant groups. But that's the politics, not the science.

Collins (and Thorpe) are still trying to do the older kind of sociology and the idea of 'the fractal model of society', and 'interactional expertise', and its empirical counterpart – the 'Imitation Game' – are central to it. My typical critique of artificial intelligence grows out of it. AIs are not members of social groups in the way humans are; even large language models, which are hugely successful because they are closer to be being socialised in the way humans are socialised than previous generations of AIs, fail where their socialisation departs from the human experience described by the fractal model and which our recent papers on the topic try to demonstrate.

Becoming a philosopher?

I have never been trained in philosophy by anyone who has a philosophy degree, nor do I have any philosophical qualifications but, in continuing to do sociology in the way described above, and given what's happened to sociology, there are indications that I am now ceasing to be a sociologist in the way most sociologists understand the term nowadays, and becoming a philosopher. The idea of interactional expertise and of the Imitation Game have been criticised by sociologists because they reify the idea of the social group, but that, to me, is what sociology is about. I don't know if this was the reason, but I was astonished when our first paper, empirically exploring the idea of interactional expertise using the Imitation Game was rejected by *Social Studies of Science* (it was the first time I had ever had a paper rejected by that journal); it was then accepted by philosophy journal, *Studies in History and Philosophy of Science*![3] And, nowadays, a large proportion of my work is published in philosophy outlets, including the Imogen paper, even though, as I see it, it deals with a quintessentially sociological question – it explains the circumstances under which sociology (as well as ethnography and anthropology) can work. I have cited philosophy outlets for my work in this paper wherever I have had a choice.

It is also the case that my work is deeply rooted in philosophy. The framing concept of the book that defines my early work – *Changing Order* – is the problem of induction: I describe myself there as setting out a sociological solution to the problem of induction. And I think the philo-

[3] This paper is [Collins et al., 2006]. A comprehensive book on Imitation Games is in press (2027, forthcoming). It is called *Imitation Games and Turing Tests: New methods for exploring societies*, Cambridge, MA: MIT Press. The book has an invitation to consult an open access web source: https://github.com/igwebstore. The last item there is the letter Collins wrote to the editor of *Social Studies of Science* asking for a careful explanation of why it had been desk-rejected.

sophical roots of the idea of interactional expertise lie with Nelson Goodman's 'new riddle of induction' since Goodman claims that our grasp of practical things is found in the degree of entrenchment in the language of our descriptions of reality [Goodman, 1955]. And my philosophy of sociological reality is based on philosopher Wittgenstein. For about two decades, until he died (in 2017), I had a continual and very friendly argument with phenomenological philosopher Hubert Dreyfus, about the relationship between language and practice, Dreyfus insisting that practice was all, and me insisting that language could capture practical understanding in the absence of practice itself. Dreyfus is famous for insisting that computers could not be intelligent unless they had bodies and he wrote the definitive (1973 – 2nd edition 1992) book-length critique of AI (which, as I often explain, would have stopped my own critique of AI in its tracks if I had not been lucky enough *not to* read his book before I had worked out my own position – see [Collins, 2019]).

At this point let us turn aside to explain, if it is not already understood, that philosophy itself can have practical consequences of the kind that science produces. Dreyfus thought he was producing a result for artificial intelligence research of what I am calling. a 'scientific' type; he was telling AI researchers that they had better start working on bodies for their machines if they wanted to reproduce human-like intelligence. But Dreyfus was getting it wrong (as scientists are entitled to). What he was missing (as pointed out by computer scientist, Doug Lenat), was that the human world included people without proper bodies (the severely physically challenged) whose intelligence was unaffected. At this point the notion of 'body' in philosophers' arguments became somewhat diffuse: unfortunately, no-one took responsibility for describing the limits to what counts as a body such as would render the need for body in an intelligent being open to falsification.

I think the 'Imogen' paper brings the practical implications of the Dreyfus-like philosophical argument back via the table of levels of AI, reproduced in, for example, [Collins, 2021, Table 1]. The lower levels of the table are concerned with individual AIs but Level 5 is autonomous societies of AIs. Imagine some thousands of *tabula rasa* AIs – AIs without any cultural experience – dropped onto a desert island with conditions for survival provided (the kind of experiment we might once have envisaged with newborn infants instead of AIs in order to resolve the nature/nurture problem). It seems obvious that for such an artificial society to develop human-like language the machines would need human-like bodies. To make the point as obvious as possible, if they didn't have bodies capable of playing tennis, they would never develop tennis-related language nor the capacity to understand it when spoken by a human visitor to their society. But this seems wrong for a single *tabula rasa* AI dropped into an already functioning human society – indeed, existing LLMs can already handle tennis-related language (I am not saying they



understand what they are talking about but, as Filatova points out, the talk is fluent enough for us not to worry about their understanding when we engage with them unless we are trying to investigate their abilities). So, the practical implication of this paragraph's-worth of philosophy is 'don't worry about the body when dealing with AIs dropped into human societies but do worry about bodies when thinking about autonomous societies of AIs'. A lot of this kind of thing is worked out in the 'Imogen' paper (published in a philosophy journal, by the way).

Formative aspirations

But there is another problem: not every 'philosopher' is really a philosopher! To worry about Sokolova's problem – is philosophy losing its special warrant in taking on sociology's empirical practices? – we have to know what philosophy really is. I'm pretty sure that it is not the sum total of the activities of all those who are referred to as 'philosophers' (though it has become very hard to say such things since constructivist sociology of science began to flourish). For example, I know there are a lot of highly trained and highly respected philosophers whose work involves their becoming experts on the writings of other philosophers and publishing the results. I think this is philosophical scholarship not philosophy. On the other hand, I am pretty sure that the later Wittgenstein is a philosopher even though David Bloor describes him as a sociologist – Bloor's not wrong it is just that in his work philosophy and sociology are not distinguishable.

So, what is philosophy? What are the 'formative aspirations' of a philosopher.[4] I think that is a hard question and I really hope philosophers will have a crack at it. I think the nearest I can get is the old story about the sage and novice who asks what supports the Earth (the Earth being a proxy for truth). The question-and-answer session in which the novice continually asks 'but what does that stand on', goes through the shoulders of Atlas, via an elephant, to a turtle, culminating in the sage saying 'from there on its turtles all the way down'. I think philosophy's formative intention is to get to the bottom turtle in so far as it is possible. And that, it seems to me, is where sociology and philosophy bump into each other because the sociology of knowledge is a very philosophical idea that is as

[4] This uncomfortable question relates to what I have been doing in trying to find a definition for 'science'. Whatever that definition is, it does not include the activities of cheats and liars who work in the scientific profession, it refers only to those who conform to the 'formative aspirations' of science as set out in, for example, [Collins, 2023, Table 1], or for a slightly improved version which will form the basis for future work see Collins and [Evans, 2026, Figure 1.2]. For the origin of the notion of 'formative aspiration', or 'formative intention', see [Collins, Kusch, 1998].

terrifying as philosophical scepticism – we only know what we think we know because of how we have been socialized.

I think that before Sokolova worries too much about whether philosophy is losing its special warrant she needs to find a way to define philosophy and that is itself a terrifyingly difficult question.

It seems to me that Thorpe and I are in a position to answer the question in respect of sociology if we are brave enough: sociology's formative aspiration is to understand the special nature of human social groups. I say this requires us to be brave because it excludes what most sociologists actually do these days. Maybe we should keep our mouths shut! But for me, I think the collision of sociology and philosophy at the level of the bottommost turtle we can find explains why I still don't know whether I'm a sociologist or a philosopher but I would say that Sokolova should not worry too much because in terms of what sociologists and philosophers do, philosophy is in the better shape.

Afterword

I consider, above, that I have become a philosopher and I separate myself from the majority of sociologists, whose academic life is primarily driven by the politics of diversity. But in recent years, as result of political events in 'the West', my work has also become primarily driven by politics. What I am trying to do these days is reform science and science-related education in a way that will inform the citizenry of the need for truth in politics if democracy is to be preserved and in trying to explain the role of science as a bulwark against the erosion of truth. The political battle to be fought is not for the preservation of underprivileged minorities (though this remains a worthy aim) but for the preservation of expert elites against the predations of would-be dictators. One does not always know where an academic career will take one.[5]

References

Bloor, 1983 – Bloor, D. *Wittgenstein: A Social Theory of Knowledge*. London: Macmillan, 1983.

Collins, 1998 – Collins, H. "Socialness and the Undersocialised Conception of Society", *Science, Technology and Human Values*, 1998, vol. 23, no. 4, pp. 494–516.

Collins, 2018 – Collins, H. "Studies of Expertise and Experience", *Topoi*, 2018, vol. 37, pp. 67–77. DOI: 10.1007/s11245-016-9412-1.

[5] The latest expression of this is [Collins, Evans, 2026], which has the title, *Veritocracy: Truth, science, and how to preserve democracy,* but [Collins, Evans, 2017], *Why Democracies need Science,* is a predecessor.



Collins, 2019 – Collins, H. "Remembering Bert Dreyfus", *AI and Society*, 2019, vol. 32, no. 2, pp. 373–376.

Collins, 2020 – Collins, H. "Interactional Imogen: Language, Practice and the Body", *Phenomenology and the Cognitive Sciences*, 2020, vol. 19, no. 5, pp. 933–960. DOI: 10.1007/s11097-020-09679-x.

Collins, 2021 – Collins, H. "The Science of Artificial Intelligence and Its Critics", *Interdisciplinary Science Reviews*, 2021, vol. 46, no. 1–2, pp. 53–70. DOI: 10.1080/03080188.2020.1840821.

Collins, 2023 – Collins, H. "Science as a Counter to the Erosion of Truth in Society," *Synthese*, 2023, vol. 202, no. 145, pp. 1–23. DOI: https://doi.org/10.1007/s11229-023-04357-2.

Collins, Evans, 2017 – Collins, H., Evans, R. *Why Democracies Need Science*. Cambridge: Polity Press, 2017.

Collins, Evans, 2026 – Collins, H., Evans, R. *Veritocracy: Truth, Science, and How to Preserve Democracy*. New York, NY: Bloomsbury Academic, 2026.

Collins et al., 2026 – Collins, H., Hall, M., Evans, R., Weinel, M. *Imitation Games and Turing Tests: New Methods for Exploring Societies*. Cambridge, MA: MIT Press, 2026 (forthcoming).

Collins, Kusch, 1998 – Collins, H., Kusch, M. *The Shape of Actions: What Humans and Machines Can Do*. Cambridge, MA: MIT Press, 1998.

Collins et al., 2006 – Collins, H.M., Evans, R., Ribeiro, R., Hall, M. "Experiments with Interactional Expertise", *Studies in History and Philosophy of Science*, 2006, vol. 37A, no. 4, pp. 656–674.

Collins et al., 2023 – Collins, H., Shrager, J., Bartlett, A., Conley, S., Hale, R., Evans, R. "Hypernormal Science and Its Significance", *Perspectives on Science*, 2023, vol. 31, no. 2, pp. 262–292.

Collins, Thorne, 2026 – Collins, H., Thorne, S. "Large Language Models and Scientific Discourse: Where's the Intelligence?", *Synthese*, 2026 (accepted, February 2026). DOI: https://doi.org/10.1007/s11229-026-05531-y. Open access: https://rdcu.be/fbpKm. Preprint: *arXiv:2603.23543*.

Collins et al., 2026a – Collins, H., Grote, H., Newbury, P., Sutton, P., Thorne, S. "Large Language Models Eroding Science Understanding: An Experimental Study", *arXiv preprint arXiv:2604.25639*, 2026 (under revision for resubmission to *AI and Ethics*).

Collins et al., 2026b – Collins, H., Thorne, S., Newbury, P., Evans, R. "More Problems of Large Language Models in Comparison with Human Knowledge", *AI & Society*, 2026. DOI: http://doi.org/10.1007/s00146-026-03021-w.

Collins, Yearley, 1992 – Collins, H.M., Yearley, S. "Epistemological Chicken", in Pickering, A. (ed.), *Science as Practice and Culture*. Chicago: University of Chicago Press, 1992, pp. 301–326.

Dreyfus, 1972 – Dreyfus, H.L. *What Computers Can't Do*. New York: Harper and Row, 1972 (2nd edition: *What Computers Still Can't Do*, 1992).

Goodman, 1955 – Goodman, N. *Fact, Fiction, and Forecast*. Harvard University Press, 1955 (1st edition).

Weizenbaum, 1976 – Weizenbaum, J. *Computer Power and Human Reason: From Judgement to Calculation*. San Francisco: W.H. Freeman, 1976.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 78–93
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 78–93
DOI: https://doi.org/10.5840/eps202663342

Фоновое знание в научно-технических практиках: диалектика детерминации и свободы*

Столярова Ольга Евгеньевна – доктор философских наук, исследователь. Межрегиональная общественная организация «Русское общество истории и философии науки». Российская Федерация, 105062, г. Москва, Лялин пер., д. 1/36, стр. 2; e-mail:olgastoliarova@mail.ru



В статье рассматривается фоновое, или неявное, знание как объяснительный ресурс эпистемологии, философии и социологии науки и техники. Исходная проблема состоит в двойственном статусе этого понятия: с одной стороны, оно описывает скрытые условия формирования убеждений, навыков, методов и решений и потому несет детерминистские импликации; с другой – выступает источником ситуативной ориентации, профессиональной интуиции, новизны и свободы действия. Для преодоления этого затруднения предлагается развести две оппозиции: эпистемическую оппозицию «явное/неявное» и практико-нормативную оппозицию «автономное/гетерономное». Граница между свободой и детерминацией проходит не между явным и неявным, а между креативным освоением унаследованных структур и их механическим воспроизведением. Устанавливается корреляция гетерономного и автономизируемого модусов неявного знания, с одной стороны, с эпистемическим принуждением и сопротивлением, с другой стороны. Особое внимание уделяется технонаучным практикам, в которых трансформируется кантовская дихотомия теоретического и практического разума. Концепции фонового знания выявляют ограниченность схемы, в которой теоретическое знание связывается с описанием необходимости, а практическое измерение – со свободой. В научно-технических практиках это распределение меняется: знание воплощается в приборах, коде, интерфейсах, моделях и протоколах, превращаясь в инфраструктуру действия. Показано, что формализация не устраняет неявное знание: субъект оценивает границы применимости протокола, различает эффект и шум, соотносит автоматизированный вывод с ситуацией. На микроуровне действий индивидов и коллективов свобода проявляется как ситуативная рекомбинация навыков, интуиции и нормативных рамок, позволяющая актору отклоняться от предзаданных технологических и цифровых сценариев. На макроуровне неявное знание выступает объяснительным ресурсом исторической трансформации практик, делая видимой ограниченность прежних схем рациональности через ретроспективный сдвиг. Обосновывается ретроспективный характер различения свободы и детерминации: то, что в прошлой исторической конфигурации действует как необходимость, в последующей раскрывается как изменяемая форма организации опыта.

Ключевые слова: фоновое знание, неявное знание, артикулированное знание, ретроспективный анализ, свобода и детерминация, наследование прошлого, новизна, научно-технические практики

* Исследование выполнено при финансовой поддержке РНФ, проект № 25-28-00120 «Фоновое знание как эпистемологическая проблема». The research was supported by Russian Science Foundation, project No. 25-28-00120 “Background Knowledge as an Epistemological Problem”.

© Столярова О.Е., 2026

BACKGROUND KNOWLEDGE IN SCIENTIFIC AND TECHNOLOGICAL PRACTICES: THE DIALECTIC OF DETERMINATION AND FREEDOM

Olga E. Stoliarova –
Dsc in Philosophy, Researcher.
Inter-Regional
Non-Government Organization
"Russian Society of History
and Philosophy of Science".
1/36 Lyalin Lane, bd. 2,
Moscow 105062,
Russian Federation;
e-mail: olgastoliarova@mail.ru

The article examines background, or tacit, knowledge as an explanatory resource in epistemology, philosophy, and sociology of science and technology. The initial problem lies in the dual status of this concept: on the one hand, it describes the hidden conditions for the formation of beliefs, skills, methods, and decisions, and therefore carries deterministic implications; on the other hand, it serves as a source of situational orientation, professional intuition, novelty, and freedom of action. To overcome this difficulty, the article proposes to distinguish two oppositions: the epistemic opposition between "explicit/tacit" and the practical-normative opposition between "autonomous/heteronomous". The boundary between freedom and determination does not run between the explicit and the tacit, but rather between the creative appropriation of inherited structures and their mechanical reproduction. A correlation is established between the heteronomous and autonomizing modes of tacit knowledge, on the one hand, and epistemic constraint and resistance, on the other. Particular attention is paid to technoscientific practices, in which the Kantian dichotomy between theoretical and practical reason is transformed. Conceptions of background knowledge reveal the limitations of the scheme in which theoretical knowledge is associated with the description of necessity, while the practical dimension is associated with freedom. In scientific-technical practices, this distribution changes: knowledge is embodied in instruments, code, interfaces, models, and protocols, turning into an infrastructure of action. It is shown that formalization does not eliminate tacit knowledge: the subject evaluates the boundaries of applicability of a protocol, distinguishes between effect and noise, and relates automated output to the situation. At the micro-level of actions of individuals and collectives, freedom manifests itself as a situational recombination of skills, intuition, and normative frameworks, allowing the actor to deviate from predetermined technological and digital scenarios. At the macro-level, tacit knowledge serves as an explanatory resource for the historical transformation of practices, making visible the limitations of previous schemes of rationality through a retrospective shift. The retrospective character of the distinction between freedom and determination is substantiated: what acts as necessity in a past historical configuration is revealed in hindsight as a changeable form of organizing experience.

Keywords: background knowledge, tacit knowledge, explicit knowledge, retrospective analysis, freedom and determination, inheritance of the past, novelty, scientific and technological practices

Фоновое знание как источник объяснения

Фоновое, или неявное, знание с недавних пор стало ключевым объектом внимания философов и социологов науки, специалистов в области исследований науки и технологии. Традиционно основной интерес этих специалистов был сосредоточен на явном, артикулированном знании, но впоследствии ситуация изменилась. Явное знание оказалось не в состоянии удовлетворить запросы исследователей. Если задача исследователей состоит в том, чтобы объяснять наблюдаемые феномены, то неявное знание может послужить продуктивным источником объяснения.

Принято считать, что знание указывает на такой уровень содержания когнитивного опыта, который доступен субъекту в форме отчета, формулировки и обоснования. Знание транспортируется в язык, оно представлено как высказывание, правило, инструкция, концепт или аргумент. Оно подразумевает способность субъекта занять по отношению к собственному когнитивному опыту рефлексивную позицию, сделать этот опыт предметом осознания и тем самым принять за него концептуальную и логическую ответственность. Вместе с тем предполагается, что существует такой слой когнитивного опыта, который остается вне отчетливой рефлексии переживающего данный опыт субъекта. Это слой интуитивного понимания или своего рода пред-знания, избегающий артикуляции. Он не может быть выражен ни в словах, ни в формулах, ни в инструкциях. Невозможно вербализировать и целенаправленно передать его другому в виде информации. Однако он выступает предметом анализа дистанцированного наблюдателя, который способен увидеть то, что остается принципиально невидимым для самих носителей неявного знания. Для внешнего исследователя этот фоновый, или имплицитный, слой превращается в объяснительный ресурс.

Сдвиг внимания к неявному знанию был не случаен. Классическая эпистемология, которая рассматривала познавательную деятельность в виде отчетливых идей, сформулированных правил, выраженных суждений и поддающихся экспликации процедур, обнаружила свою ограниченность [Audi, 2010; Williamson, 2000]. Было выявлено, что формально обоснованное истинное убеждение еще не гарантирует знания, поскольку успешность познания нередко зависит от факторов, избегающих формализации [Bengson, Moffett (eds.), 2011]. Эта особенность присутствует как в обыденной жизни, где мы ориентируемся ситуативно и зачастую безотчетно, так и в науке, где лабораторная работа опирается на инсайты и навыки, не сводимые к методологическим регламентам.

Анализируя представления, лежащие в основе теоретических и практических феноменов, исследователь реконструирует неявное



знание, т.е. выявляет скрытые механизмы организации представлений и действий. Эти механизмы, не будучи сами полноправной частью артикулированного и осознаваемого знания, все же каким-то не вполне рациональным образом в нем участвуют. С точки зрения Майкла Поланьи, который первым попытался концептуализировать неявное знание в систематической форме [Polanyi, 1966], они образуют фон фокуса внимания, сформированный предшествующим жизненным опытом, как личностным, так и социальным. К этому фону относятся телесные привычки и умения, инструментальные средства, невербальные подсказки, усвоенные нормы, негласные критерии надежности – все то, что передается традицией, воспитанием, совместной деятельностью и т.п. Можно назвать такой фон *практическим контекстом* как структурным условием знания, и с этой базовой идеей согласилось бы, вероятно, большинство современных исследователей неявного знания. Но нужно иметь в виду, что идея «практического контекста» чрезвычайно размыта (как и сама идея *контекста*), и в процессе ее конкретизации начинаются принципиальные разногласия [Collins, 2010; Dreyfus, 1992; Thomson (ed.), 2023; Turner, 1994].

Оттолкнемся тем не менее от этой, пусть даже сильно размытой концепции, чтобы двинуться дальше. Допустим, что практически усвоенное неявное знание влияет на формирование убеждений и действий его носителей, оставаясь скрытым от них, и открывается внешнему, дистанцированному взгляду. Данная модель выглядит детерминистической. Познающий субъект, который отдает себе отчет в своих мыслях, убеждениях и действиях, руководим не поддающимися осознанию движущими силами, причем эти последние непосредственно участвуют в формировании его самосознания. Можно ли при таких условиях защитить рациональным образом внутреннюю свободу субъекта?

Я попытаюсь ответить на этот вопрос посредством двух шагов. Во-первых, я рассмотрю фоновое знание через призму инспирированного Кантом различения теоретического и практического разума. И во-вторых, я помещу этот феномен в пространство сегодняшних научно-технических практик. В оптике кантовского различения фоновое знание обнаруживает двойственный характер. Кант соотносит теоретическое применение разума (в последующей традиции часто обозначаемое как позиция «теоретического субъекта») с областью природы. Природа понимается как порядок явлений, подчиненный законам причинной необходимости. Этот порядок воспринимается и мыслится посредством универсальных форм чувственности и категорий рассудка. С эмпирической, чувственно воспринимаемой стороны субъект и сам принадлежит природе и включен в ее причинно-следственные связи. Практическое применение разума, напротив, раскрывает способность субъекта действовать, исходя из собственного целеполагания, т.е. реализовывать особую неприродную

причинность через свободу. Согласно Канту, эта способность не определяется внешними обстоятельствами или чувственными влечениями, а руководствуется идеей морального закона [Кант, 1964]. Таким образом, практический разум выражает автономию воли, т.е. ее независимость от порядка природы и способность к спонтанному действию.

Фоновое знание, по-видимому, относится к области практического разума – хотя бы в силу того, что его вряд ли возможно отнести к теоретическому разуму. Концепция неявного знания аккумулирует те признаки, которые противоположны признакам теоретического знания. Когда исследователи говорят о неявном знании, они имеют в виду следующее: способность действовать уместно без обращения к эксплицитному правилу, использовать методы, которые не требуют вербального отчета, мгновенно оценивать ситуацию, формулировать высказывания и аргументы в рамках того, что подразумевается, но не получает выражения, всегда остается «за кадром», не имеет репрезентационального содержания [Ryle, 1949; Collins, 2010]. Неявное знание – это нечто вроде телесно-культурного режима детерминации явного знания. Соответственно, оно соединяет в себе две идеи – идею детерминации и идею практической вовлеченности субъекта в культуру и общество. Отсюда возникает трудность. В рамках кантовского различения теоретического и практического разума, которое, хотя и с существенными трансформациями, сохраняет философскую значимость до сегодняшнего дня [Brandom, 1998; Korsgaard, 1996], практическое измерение знания следует связать со свободой.

Артикулированное знание, репрезентирующее факты, процессы или положение вещей, обращено к определенному состоянию дел, которое и в природном, и в социальном мире подчинено физическим законам. В науке формализованное знание выражает необходимость, поскольку опирается на фиксированные правила, логические связи и воспроизводимые процедуры, исключающие произвол и случайность. Такого рода знание отсылает к установленным закономерностям, причинным зависимостям или нормативным конвенциям, т.е. к тому, что может быть удержано в общей форме. Фоновое знание, напротив, функционирует как источник ситуативной адаптации и творческого инсайта, оставаясь за пределами артикуляции и формализации. Если воспользоваться различием, которое авторитетный физик и исследователь формальных систем Роджер Пенроуз проводит между алгоритмическими и неалгоритмическими составляющими сознания и мышления [Пенроуз, 2004], то неявное знание окажется на неалгоритмической стороне. Согласно Пенроузу, алгоритмическое мышление по своей структуре детерминировано; оно следует заданной последовательности операций и заключается в перестановке и комбинации наличествующих элементов. Оно

не способно породить ничего принципиально нового. Неалгоритмическое мышление хотя и принадлежит в каком-то смысле сознанию, протекает подспудно и выступает источником новизны, вдохновения и спонтанности. Фоновое знание по своим характеристикам относится именно к неосознанному, имплицитному мышлению. Соответственно, в этой конфигурации понятий оно отвечает за инициативу и свободу.

Между тем, как мы утверждали выше, в современных эпистемологических исследованиях, а также в философии и социологии науки и техники фоновое, или неявное, знание описывается как скрытая основа действия. Предполагается, что оно незаметно направляет поведение, оформляет концепции, методы и ограничивает диапазон возможного. Детерминистские импликации концепта неявного знания хорошо осознаются исследовательским сообществом, о чем свидетельствует оживленная дискуссия последних лет о проблеме ответственности агента, чье поведение обусловлено имплицитными допущениями и предубеждениями [Baston, 2023; Bonicalzi, 2025; Holroyd et al., 2017]. Соответственно, один и тот же феномен может быть истолкован двояко: либо как носитель практической инициативы, либо как фоновая детерминанта, действующая «за спиной» субъекта и служащая источником объяснения для дистанцированного наблюдателя. Таким образом, неявное знание парадоксально выступает в двух противоположных модусах – как каузальная сила и как источник спонтанности. К тому же, если мы согласимся с тем, что неявное знание есть источник свободы, и одновременно допустим, что артикулированное знание, которое мы связываем с естественной и социальной необходимостью, обусловлено неявным знанием, то мы получим еще один парадокс: необходимость выраженного и сформулированного знания окажется вызвана к жизни спонтанностью имплицитно протекающих процессов. Не лучше ли в таком случае полностью развести эти два вида знания, приписав их двум взаимоисключающим областям, наподобие того, как это сделал в свое время Кант в отношении теоретического и практического разума, и рассматривать их отдельно друг от друга?[1]

[1] Подобным образом поступает Х. Дрейфус, предлагая считать, что знание-как и знание-что не только несводимы друг другу, но, более того, исключают друг друга. Дрейфус убежден: чем больше рефлексивного содержания в практических диспозициях, тем меньше в них умения, т.е. экспертного действия [Dreyfus, 2005]. Дальнейшее развитие точки зрения Дрейфуса в направлении радикального габитуализма см.: [Cappuccio, 2023]. Мне ближе позиция Дж. Макдауэлла, который защищает концепцию дорефлексивного, но при этом концептуального содержания неявного знания. Дискуссию Макдауэлла и Дрейфуса о природе дорефлексивного опыта и статусе концептуального знания см.: [Dreyfus, 2007a; Dreyfus, 2007b; McDowell, 2007a; McDowell, 2007b].

Однако такое решение нельзя признать удовлетворительным. Во-первых, неявное знание обретает существование и смысл только в своем отношении к выраженному знанию. Или следует вообще отказаться от концепции неявного знания, или придется сохранить за ним роль скрытого обуславливающего фактора, который реализуется внешним, явным образом [Воеводенко, 2025][2]. Во-вторых, влиятельная линия современной эпистемологии и посткантовской философии настаивает на том, что жесткое разведение навыка и рефлексии, рецептивности и спонтанности, восприятия и суждения, действия и знания искажает структуру самой рациональной практики. Их изоляция друг от друга искусственно разрывает «пространство оснований» [McDowell, 1994], в котором все эти элементы функционируют во взаимной связи [Касавин, 2013; Лекторский, 2001]. Наконец, жесткая дихотомия навыков и формулировок, парадигмы и опыта не выдерживает встречи с реальной научно-технической деятельностью: артикулированные правила без фоновых навыков не работают, а имплицитные знания и умения без как минимум частичной артикуляции не могут быть ни переданы, ни критически скорректированы.

[2] Попытка избежать этой дилеммы, выводя неявное знание за пределы концептуального порядка, предпринята в работе Х. Дрейфуса и Ч. Тейлора *Retrieving Realism*. Авторы предлагают отказаться от репрезентационалистской картины мира в пользу модели практического вовлечения, где первичным оказывается телесное умение (*coping*) и дорефлексивная ориентация в среде [Dreyfus, Taylor, 2015]. Однако, если практический фон действительно первичен, он должен быть включен и в объяснение исторической смены концептуальных схем: почему репрезентационализм приобрел статус самоочевидности и как позднее открылась возможность его критики. Без такого механизма подход сохраняет преимущественно синхронный характер: дорефлексивная практика описывает первичный контакт с миром, но не диахронию рациональности. Чтобы избежать дуализма между статичным фоном и историей концептуальных артикуляций, необходимо показать, как практический фон трансформируется и каким образом прежняя самоочевидность опыта делается доступной ретроспективной рефлексии и критическому присвоению. Здесь обнаруживается, что полное разведение практики и концептуального порядка вряд ли возможно. Практический фон не является статичным внешним условием; он участвует в историческом изменении картин мира. То, что изначально задавало горизонт самоочевидности, впоследствии раскрывается как одна из альтернатив и превращается в материал для критики. В таком случае сам отказ от репрезентационализма указывает на темпоральную связь между телесным вовлечением и рациональной артикуляцией, а также на то, что историческая трансформация рациональности укоренена в способности когнитивного опыта переосмыслять собственные неявные условия.



Практические режимы фонового знания

Затруднение, вызванное тем, что фоновому знанию может быть приписана способность порождать одновременно свободу и необходимость, преодолевается за счет различения двух оппозиций: эпистемической оппозиции «явное/неявное» и практико-нормативной оппозиции «автономное/гетерономное». Эти различия пересекаются, но не совпадают. Иначе говоря, и неявное, и явное знание могут функционировать как в автономном, так и в гетерономном режиме. Неявное знание не является свободным только потому, что оно имплицитно; оно может быть условием умелого и осмысленного действия, но может также выступать формой привычки или социальной инерции. Точно так же явное знание не является детерминирующим только потому, что оно эксплицировано: правило, норма или теория могут быть как внешним протоколом, так и осознанным основанием действия.

Условием свободы, таким образом, выступает не неявное знание само по себе и не его исчерпывающая рефлексивная подотчетность, которая по определению невозможна, а способность субъекта осваивать и актуализировать унаследованные структуры в конкретной ситуации. Свобода проявляется там, где субъект не просто воспроизводит наличную схему, но способен уместно применить ее, скорректировать или преобразовать в ответ на требования новых обстоятельств. Детерминация, соответственно, проявляется не в самой зависимости от унаследованных структур, поскольку эта зависимость неизбежна, а в таком способе их реализации, при котором они воспроизводятся как внешняя необходимость, привычный автоматизм или самоочевидный порядок. Поэтому граница между свободой и детерминацией проходит не между неявным и артикулированным знанием, а между креативным освоением сложившихся структур и их механическим воспроизведением. Можно быть пленником эксплицитных правил так же, как и пленником неотрефлексированной установки; и наоборот, неявный навык может быть не препятствием для свободы, а условием практической ориентации, если он позволяет субъекту попадать в социальные «ноты», т.е. действовать уместно и гибко в конкретных обстоятельствах. Когда актор способен критически соотносить формальный протокол с ситуативным контекстом, оценивать границы применимости модели и брать на себя ответственность за отклонение от стандарта ради эпистемической или этической целесообразности, неявное знание перестает быть слепой детерминантой и становится условием автономного действия. Напротив, когда экспликация превращается в навязанный сценарий, а навык – в рутинизированную привычку, субъект оказывается в положении гетерономного исполнителя.

Впрочем, различие между свободой и детерминацией едва ли может быть в полной мере доступно участникам практики в момент действия. Присутствие неявного знания как элемента структуры когнитивного опыта придает вопросу о свободе или обусловленности познающего агента ретроспективное измерение. Вопрос проясняется не из абсолютной внешней точки обзора, а из более поздней историко-культурной ситуации, в которой прежняя форма практики утратила самоочевидность и предстала как одна из возможных. Ретроспекция обнаруживает и скрытые ограничения прошлого, и те альтернативы, которые для самих участников практики еще не могли быть предметом полностью осознанного выбора. В данном случае я имею в виду актуальные ревизии кантовского разведения теоретического и практического разума, которые являются результатом ретроспективного сдвига. Концепции неявного знания выявляют ограниченность кантовской модели, поскольку нарушают ее базовое распределение ролей: теоретическое знание не сводится к описанию необходимости, а практический разум не исчерпывается автономным самоопределением субъекта. Возможность увидеть эту ограниченность возникает в иной социотехнической реальности, где альтернативы прежним практическим и концептуальным диспозициям уже осуществились.

Обратимся к научно-техническим практикам как наиболее в данном отношении показательным. Цифровизация, автоматизация, алгоритмическая обработка данных, стандартизация и платформизация науки существенно расширяют сферу явного знания: все больше процедур переводится в форму инструкций, интерфейсов, моделей, протоколов и воспроизводимых пайплайнов. Но растущая экспликация не устраняет неявное знание. Напротив, по мере усложнения социотехнических систем возрастает роль тех форм практической компетенции, которые невозможно полностью заменить формальными правилами. Ученый, инженер или врач должны не только следовать протоколу, но и распознавать момент, когда протокол оказывается недостаточным; не только доверять автоматизированному выводу, но и оценивать границы его применимости; не только использовать автоматизированную систему, но и удерживать возможность ситуативного выбора там, где алгоритмическая логика исчерпывает себя [Collins, 1990; Collins, 2010].

Целесообразно провести различение двух модусов реализации фонового знания, каждый из которых признается заслуживающим внимания в современной аналитике. С одной стороны, существует *гетерономное неявное знание*: это инфраструктурные зависимости, алгоритмические «черные ящики», рутинизированные навыки и дисциплинарные привычки, которые действуют на актора как внешняя необходимость, не поддающаяся критическому учету и пересмотру. Здесь субъект становится проводником скрытых детерминант, а его свобода оказывается под вопросом. С другой стороны, существует

автономизируемое неявное знание: это профессиональная интуиция, ситуативная грамотность, тактильное различение и способность к импровизации. Такое знание становится условием свободы, когда субъект способен творчески присвоить его как основание собственного действия, а не воспроизводить механически.

Это различение коррелирует с описанной Стивеном Тернером диалектикой эпистемического принуждения и сопротивления [Turner, 2024]. Гетерономный модус соответствует формам цифрового кураторства, алгоритмической стандартизации и унификации исследовательских схем, которые конструируют гомогенную (в терминологии Тернера) когнитивную среду и создают иллюзию выбора, систематически скрывая альтернативные эпистемические траектории. В таких условиях неявное знание индуцируется контролируемыми входами, превращая актора в объект того, что Тернер называет «эпистемическим инжинирингом» [Ibid., p. 29]. Напротив, автономизирующий модус опирается на гетерогенное, личностно укорененное знание, которое Тернер рассматривает как главный ресурс сопротивления. Поскольку оно формируется через разнообразный, часто не координируемый опыт взаимодействия с миром, оно порождает «интуитивное чутье» (gut feelings) и когнитивный диссонанс при столкновении с выстроенным консенсусом. Именно неустранимая гетерогенность препятствует полной эпистемической гомогенизации и сохраняет возможность критической дистанции [Ibid., p. 35].

От ситуативной новизны к ретроспективной диагностике

Свобода в этой конфигурации реализуется через расхождение с унаследованными структурами. Это происходит и на микроуровне, о котором говорит Тернер, и на макроуровне, который я считаю нужным добавить к данной картине. На микроуровне свобода проявляется как актуализация новизны в моменте практического становления. Гетерогенная текстура неявного опыта, на которую указывает Тернер, выступает здесь не только фоном сопротивления, но и ресурсом творческого отклонения от предзаданных схем. Специалист никогда не начинает познание из пустоты: он наследует протоколы, профессиональные привычки, интерфейсные подсказки, статистические выводы, дисциплинарные нормы и предшествующий опыт. Однако эти элементы не образуют линейной причинно-следственной цепи. В конкретной ситуации они заново соотносятся друг с другом, получают различный вес и могут быть актуализированы в практическом решении иначе, чем это предполагает их стандартное применение.

Это хорошо видно на примере использования систем искусственного интеллекта в клинической практике. Система предоставляет вероятностное заключение, опираясь на эксплицитные метрики, тренировочные данные и заданную архитектуру модели. Однако врач не просто «считывает» этот результат как готовое основание действия. Он включает его в более сложную конфигурацию: сопоставляет с анамнезом, состоянием пациента, возможными исключениями, этическими ограничениями и собственным клиническим опытом, который не может быть полностью переведен в формальную процедуру. Технологии не задают действие с необходимостью, а намечают определенный сценарий, выделяют одни параметры как значимые, другие оставляют на периферии, предлагают вероятностный маршрут решения [Rosenberger, Verbeek, 2015]. Этот маршрут не исчерпывает самой ситуации. Между алгоритмическим выводом и клиническим действием остается зазор, в котором возможно новое соединение разнородных условий: статистического результата, телесной симптоматики, профессиональной интуиции, этической нормы[3].

Неявное знание в этом случае не совпадает с «черным ящиком» модели. Скорее, оно проявляется как клиническая грамотность врача: способность отличить значимый сигнал от статистического артефакта, ограничить действие протокола, увидеть в специфических обстоятельствах то, что не было заранее предусмотрено общей схемой. Поэтому автономизирующий модус неявного знания состоит не в отказе от технической обусловленности, а в ее переработке. Алгоритм, протокол и интерфейс становятся не внешней необходимостью, а материалом нового акта практики. Сопротивление гомогенизирующему давлению цифрового кураторства, о котором говорит Тернер, выражается в появлении непредсказуемого элемента внутри самой практики. Действие перестает быть механическим воспроизведением уже заданной последовательности и становится ситуативной рекомбинацией, где интуиция, этическая рамка, профессиональный опыт и алгоритмический вывод собираются в новую конфигурацию решения. Свобода здесь не предшествует ситуации как готовая способность

[3] В сегодняшних эмпирических и методологических исследованиях медицины рассматриваются вопросы о том, как ИИ внедряется в клиническую практику и что именно делает решение врача, использующего ИИ в диагностике и лечении, правильным, или успешным. Эти исследования показывают, что клиническое решение с участием ИИ не сводится к принятию алгоритмической рекомендации, которую выдает система: оно предполагает интуитивную и рефлексивную оценку клинической уместности, этических последствий, специфических характеристик пациента и иных привходящих условий. Медицинский ИИ описывается не как заместитель врача, а как социотехническая поддержка клинического рассуждения, требующая профессиональной интерпретации и ситуативного применения [McCradden et al., 2025; Sokol et al., 2025].



субъекта; она возникает в процессе преобразования унаследованных условий в конкретный ответ.

Аналогичным образом в лабораторной или инженерной практике стандартизированная процедура не устраняет необходимости ситуативного различения. Когда прибор ведет себя нестабильно, данные расходятся с ожиданиями, а инструкция не покрывает пограничный случай, успешность действия зависит от способности заново определить, что именно считать ошибкой, отклонением, шумом или значимым эффектом. Здесь фоновое знание проявляется как умение работать с неустойчивостью: не воспроизводить установленную последовательность операций, а переопределять ее в ответ на сопротивление материала, оборудования или самой исследовательской ситуации[4].

Поэтому на микроуровне свобода состоит не в том, что субъект выходит за пределы предзаданных условий, а в том, что он способен иначе собрать доступные ему данные. Прошлое присутствует в действии как привычка, навык, протокол, техническая среда и дисциплинарная норма, но оно не диктует форму своего продолжения. Именно возможность нового способа наследования отличает автономизируемое неявное знание от гетерономного: первое открывает пространство для вариации и коррекции, второе замыкает действие в уже сложившейся схеме.

Переходя на макроуровень, мы констатируем историческое изменение архитектуры научно-технических практик. На этом уровне фоновое знание присутствует не только как элемент индивидуального выбора, проявляющийся в частном решении врача, инженера или исследователя. Оно становится источником объяснения исторической трансформации теорий и практик рационального действия: позволяет понять, почему определенные формы связи между знанием и действием воспринимались как безальтернативные, а затем утратили самоочевидность. В исторической перспективе расхождение с прошлым заметнее, чем на уровне отдельного действия. В конкретной ситуации трудно различить, является ли отклонение от протокола ошибкой, профессиональной интуицией или действием скрытой

[4] В этом отношении показателен описанный Гарри Коллинзом феномен регресса экспериментатора: корректность экспериментального результата не может быть установлена независимо от оценки корректности экспериментальной процедуры, тогда как корректность процедуры, в свою очередь, часто распознается через ожидаемый или признанный результат [Collins, 1985]. Поэтому экспериментальная ситуация не исчерпывается следованием протоколу. Она требует неявной компетентности, позволяющей определить, работает ли прибор, применим ли стандарт, является ли расхождение помехой или свидетельством нового эффекта. Анализ регресса экспериментатора дополнен в работе [Feest, 2016], в которой показано, как неявное знание участвует в производстве новых научных результатов в условиях неопределенности протоколов, данных и экспериментов.

привычки. Но на более длинной дистанции становится видно, какие формы фонового знания закрепляли старые режимы действия, а какие открывали возможность их переоформления. Поэтому фоновое знание оказывается не только предметом объяснения, но и объяснительным ресурсом: посредством него можно реконструировать, каким образом меняются сами условия того, что считается знанием, нормой, ошибкой, свободой, необходимостью, ответственным действием и т.д.

В современной научно-технической ситуации прежнее распределение ролей между теоретическим знанием и практическим действием существенно трансформируется. В кантовской модели теоретический разум конституировал знание о природе как сфере необходимости, а технически-практическое применение знания относилось к области гипотетических императивов. Сегодня научное знание производится и функционирует в материально-технической и цифровой форме: оно воплощается в приборах, программном коде, интерфейсах, системах классификации и вычислительных моделях. Эти формы не только представляют объект, но и задают способы его операционализации; не только фиксируют результат познания, но и организуют последовательность действий, распределяют внимание и участвуют в принятии решений. Теоретическое знание тем самым становится операциональной инфраструктурой практики.

В этих условиях действующий субъект сталкивается не с внешним порядком природы, но с технически объективированными схемами организации окружающего мира. Они заранее задают допустимые ходы, параметры видимости и режимы принятия решений, но не исчерпывают самого действия. Поэтому вопрос о свободе смещается: он касается не противопоставления внутренней воли и внешней необходимости, а способности актора работать с воплощенными схемами знания, не растворяясь в них. Фоновое знание выступает здесь медиатором, позволяющим удерживать дистанцию по отношению к готовым техническим сценариям и заново определять, что в данной ситуации считать значимым, достаточным или обязательным.

Таким образом, новая социотехническая конфигурация выступает одновременно как условие ретроспективной диагностики и как основание для концептуального обновления. Во-первых, она выявляет ограниченность кантовского разведения теоретического и практического разума: в современных практиках знание, техническая инфраструктура и нормативность не распределяются по разным областям, а образуют единую подвижную конфигурацию. Во-вторых, она позволяет зафиксировать новизну современной схемы, в которой свобода мыслится не как трансцендентальный разрыв с обусловленностью, а как процессуальное достижение, т.е. как непрерывная интерпретация и переработка наследуемых условий.

Заключение

Подведем итог. Двойственный характер фонового знания, в котором свобода спорит с детерминацией, обусловлен его включенностью в различные режимы осуществления практики. В одном случае оно закрепляет сложившиеся схемы действия, поддерживает дисциплинарную инерцию, алгоритмическую стандартизацию и инфраструктурные зависимости; в другом – открывает возможность вариации, коррекции и новизны. Свобода в таком понимании принадлежит не изолированному субъекту и не самодостаточному акту разумной воли, а процессу изменения практик, в котором прежние связи между знанием и действием теряют самоочевидность и становятся доступными для ситуативной коррекции и критического пересмотра.

Это не означает, что позиция критического пересмотра не имеет собственных неявных предпосылок. Напротив, она также опирается на свои слепые зоны, фоновые навыки, инфраструктуры и режимы самоочевидности, которые могут быть раскрыты только будущей ретроспективной критикой. Поэтому свобода и детерминация предстают не как абсолютные свойства субъекта или действия, а как исторически подвижные способы отношения к унаследованным структурам: то, что в одной конфигурации опыта действует как необходимость, в другой может стать материалом для переописания, ограничения и нового практического осуществления.

Список литературы

Воеводенко, 2025 – *Воеводенко А.В.* Стратегии определения неявного знания: М. Полани, Г. Коллинз, С. Тернер // Цифровой ученый: лаборатория философа. 2025. № 3. С. 94–111.

Кант, 1965 – *Кант И.* Основы метафизики нравственности. 1785 // *Кант И.* Сочинения в шести томах. М.: Мысль, 1965. (Философ. наследие). Т. 4. Ч. I. С. 219–310.

Касавин, 2013 – *Касавин И.Т.* Социальная эпистемология: фундаментальные и прикладные проблемы. М.: Альфа-М, 2013. 557 с.

Лекторский, 2001 – *Лекторский В.А.* Эпистемология классическая и неклассическая. М.: Эдиториал УРСС, 2001. 256 с.

Пенроуз, 2004 – *Пенроуз Р.* Новый ум короля. О компьютерах, мышлении и законах физики / Пер. с англ. М.: Эдиториал УРСС, 2004. 398 с.

References

Audi, 2010 – Audi, R. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. New York: Routledge, 1998.

Bengson, Moffett (eds.), 2011 – Bengson, J., Moffett, M.A. (eds.) *Knowing How: Essays on Knowledge, Mind, and Action*. Oxford: Oxford University Press, 2011.

Baston, 2023 – Baston, R. "The Nurturing Stance, Moral Responsibility, and the (Implicit) Bias Blind Spot", *Journal of the American Philosophical Association*, 2023, vol. 9, no. 1, pp. 1–20. DOI: 10.1017/apa.2021.45.

Bonicalzi, 2025 – Bonicalzi, S. "Implicit Biases and Habitual Cognition: Challenges for Moral Responsibility", *Topoi*, 2025, [https://link.springer.com/article/10.1007/s11245-025-10335-9?utm_source=chatgpt.com, accessed on 21.05.2026].

Brandom, 1998 – Brandom, R.B. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1998.

Cappuccio, 2023 – Cappuccio, M.L. "Dreyfus is Right: Knowledge-That Limits Your Skill", *Synthese*, 2023, 202, 85, [https://doi.org/10.1007/s11229-023-04248-6, accessed on 27.05.2026].

Collins, 1985 – Collins, H.M. *Changing Order: Replication and Induction in Scientific Practice*. Sage, 1985.

Collins, 1990 – Collins, H.M. *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA; London: The MIT Press, 1990.

Collins, 2010 – Collins, H.M. *Tacit and Explicit Knowledge*. Chicago: University of Chicago Press, 2010.

Dreyfus, 1992 – Dreyfus, H.L. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press, 1992.

Dreyfus, 2005 – Dreyfus, H.L. "Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise", *Proceedings and Addresses of the American Philosophical Association*, 2005, vol. 79, pp. 47–63.

Dreyfus, 2007a – Dreyfus, H.L. "The Return of the Myth of the Mental", *Inquiry*, vol. 50, no. 4, pp. 352–365. https://doi.org/10.1080/00201740701492174

Dreyfus, 2007b – Dreyfus, H.L. "Response to McDowell", *Inquiry*, vol. 50, no. 4, pp. 371–377. https://doi.org/10.1080/00201740701492224

Dreyfus, Taylor, 2015 – Dreyfus, H., Taylor, Ch. *Retrieving Realism*. Cambridge, MA: Harvard University Press, 2015.

Feest, 2016 – Feest, U. "The Experimenters' Regress Reconsidered: Replication, Tacit Knowledge, and the Dynamics of Knowledge Generation", *Studies in History and Philosophy of Science Part A*, 2016, vol. 58, pp. 34–45.

McCradden et al., 2025 – McCradden, M.D. Thai, K., Assadi, A. et al. "What Makes a 'Good' Decision with Artificial Intelligence? A Grounded Theory Study in Paediatric Care", *BMJ Evidence-Based Medicine*, 2025, vol. 30, pp. 183–193. DOI: 10.1136/bmjebm-2024-112919.

Holroyd et al., 2017 – Holroyd, J., Scaife, R., Stafford, T. "Responsibility for Implicit Bias", *Philosophy Compass*, 2017, vol. 12, no. 3, e12410. DOI: 10.1111/phc3.12410.

Kant, 1965 – Kant, I. “Osnovy metafiziki nravstvennosti” [Grundlegung zur Metaphysik der Sitten], in: I. Kant, *Sochinenija* [Works], vol. 4, part 1. Moscow: Mysl’, 1965, pp. 219–310. (Trans. into Russian)

Kasavin, 2013 – Kasavin, I.T. *Social’naja jepistemologija: fundamental’nye i prikladnye problemy* [Social Epistemology: Fundamental and Applied Problems]. Moscow: Al’fa-M, 2013. (In Russian)

Korsgaard, 1996 – Korsgaard, C.M. *Creating the Kingdom of Ends*. Cambridge; New York: Cambridge University Press, 1996.

Lectorsky 2001 – Lectorsky, V.A. *Epistemologiya klassicheskaya i neklassicheskaya* [Classical and Nonclassical Epistemology]. Moscow: Editorial URSS Publ., 2001. (In Russian)

McDowell, 1994 – McDowell, J. *Mind and the World*. Cambridge, MA: Harvard University Press, 1994.

McDowell, 2007a – McDowell, J. “What Myth?”, *Inquiry*, vol. 50, no. 4, pp. 338–351. https://doi.org/10.1080/00201740701492158

McDowell, 2007b – McDowell, J. “Response to Dreyfus”, *Inquiry*, vol. 50, no. 4, pp. 378–380. https://doi.org/10.1080/00201740701492232

Penrose 2004 – Penrose, R. *Novyj um korolya. O kompyuterax, myshlenii i zakonax fiziki* [The Emperor’s New Mind: Concerning Computers, Minds and The Laws of Physics]. Moscow: Editorial URSS Publ., 2004. (Trans. into Russian)

Polanyi 1966 – Polanyi, M. *The Tacit Dimension*. New York: Doubleday & Co., 1966. 108 pp.

Rosenberger, Verbeek 2015 – Rosenberger, R., Verbeek, P.-P. *Postphenomenological Investigations: Essays on Human – Technology Relations*. Lanham: Lexington Books, 2015.

Ryle, 1949 – Ryle, G. *The Concept of Mind*. London: Routledge, 2000.

Sokol et al., 2025 – Sokol, K., Fackler, J., Vogt, J.E. “Artificial Intelligence Should Genuinely Support Clinical Reasoning and Decision Making to Bridge the Translational Gap”, *npj Digital Medicine*, 2025, 8, 345. https://doi.org/10.1038/s41746-025-01725-9.

Thomson (ed.), 2023 – Thomson, J.R. (ed.) *The Routledge Handbook of Philosophy and Implicit Cognition*. London and New York: Routledge, 2023.

Turner, 1994 – Turner, S. *The Social Theory of Practices: Tradition, Tacit Knowledge and Presuppositions*. Cambridge: Polity Press, 1994.

Turner, 2024 – Turner, S.P. “Epistemic Coercion”, *Epistemology & Philosophy of Science*, 2024, vol. 61, no. 3, pp. 21–38. https://doi.org/10.5840/eps202461337

Voevodenko 2025 – Voevodenko, A.V. “Strategii opredeleniya neyavnogo znaniya: M. Polani, G. Kollinz, S. Terner” [Strategies for Defining Tacit Knowledge: M. Polanyi, H. Collins, S. Turner], *Digital Scholar: Philosopher’s Lab,* 2025, no. 3, pp. 94–111. DOI: 10.32326/2618-9267-2025-8-3-94-111. (In Russian)

Williamson, 2000 – Williamson, T. *Knowledge and Its Limits*. Oxford: Oxford University Press, 2000.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 94–112
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 94–112
DOI: https://doi.org/10.5840/eps202663343

Де-антропоцентризм как концептуальная альтернатива антропоморфизации в эпистемологии искусственного интеллекта*

Фролов Константин Геннадьевич – кандидат философских наук, исследователь. Межрегиональная общественная организация «Русское общество истории и философии науки». Российская Федерация, 105062, г. Москва, Лялин пер., д. 1/36, стр. 2; e-mail: konstantin-frolov@yandex.ru



Статья посвящена вопросу о том, могут ли эпистемические понятия – такие как убеждение, знание, обоснование, владение понятиями, сомнение и другие – быть обоснованно приписаны моделям искусственного интеллекта. Рассматривается широкий спектр позиций по данному вопросу – от скептических до радикально оптимистических. При этом отстаиваемая в статье позиция относится скорее к оптимистической части спектра. Ключевой выдвигаемый тезис заключается в том, что уместность приписывания эпистемических понятий моделям искусственного интеллекта определяется не внутренними свойствами системы, но ее способностью функционировать в качестве нормативного эпистемического агента, а также ее способностью корректно взаимодействовать с другими эпистемическими агентами. Такая позиция среди прочего ориентирована на то, чтобы отклонить выдвигаемые в адрес оптимистических подходов возражения о неоправданной антропоморфизации представлений об устройстве и функционировании систем искусственного интеллекта. Эта цель достигается путем отстаивания тезиса де-антропоцентризма. Данный тезис гласит, что ничто ни в природе человека, ни в природе эпистемических понятий не дает нам оснований полагать, что лишь человек в силу его особой природы способен к «подлинному» знанию, пониманию, рассуждению и владению понятиями. При этом обоснование тезиса де-антропоцентризма не требует ни пересмотра человеческой природы, ни трансформации взглядов о природе и статусе существующих технологий искусственного интеллекта, но требует лишь принятия определенной позиции в отношении семантики эпистемических понятий. Тем самым обоснование тезиса де-антропоцентризма относится преимущественно к сфере эпистемологии и философии языка, а не к области когнитивных наук и не к области специальных исследований искусственного интеллекта.

Ключевые слова: эпистемические понятия, искусственный интеллект, знание, семантика, нормативность, агентность, антропоморфизация

* Исследование подготовлено при поддержке Российского научного фонда, проект № 26-18-00851 «Цифровые трансформации семантики понятия науки в социально-сетевую эпоху: лингво-аналитические, теоретико-эволюционные и социально-коммуникативные исследования».

© Фролов К.Г., 2026

DE-ANTHROPOCENTRISM AS A CONCEPTUAL ALTERNATIVE TO ANTHROPOMORPHIZATION IN THE EPISTEMOLOGY OF ARTIFICIAL INTELLIGENCE

Konstantin G. Frolov –
PhD in Philosophy, researcher.
Inter-Regional
Non-Government Organization
"Russian Society of History
and Philosophy of Science".
1/36 Lyalin Lane, bd. 2,
Moscow 105062,
Russian Federation;
e-mail:konstantin-frolov@
yandex.ru

The paper explores whether epistemic concepts – such as belief, knowledge, justification, possession of concepts, and doubt – can be justifiably ascribed to AI models. Various positions on this issue are considered, ranging from skepticism to radical optimism. The position defended in this paper, however, leans more towards the optimistic side of the spectrum. The main claim is that justified ascription of epistemic concepts to AI models is determined not by the system's intrinsic properties, but by its ability to function as a normative epistemic agent as well as its capability to interact properly with other epistemic agents. This optimistic view aims to counter criticisms of unjustified anthropomorphism in our understanding of the design and functioning of AI systems. This is accomplished by advocating for the thesis of de-anthropocentrism. This thesis argues that there is nothing in human nature or in the nature of epistemic concepts that justifies the belief that only humans, due to their unique nature, are capable of possessing 'real' knowledge, understanding, reasoning, and possession of concepts. Remarkably, justifying the de-anthropocentrism thesis does not require a reconsideration of human nature or a transformation of views on AI. Instead, it merely requires a certain stance regarding the semantics of epistemic concepts. Thus, the justification for the de-anthropocentrism thesis primarily pertains to the realm of epistemology and the philosophy of language, rather than the study of cognitive science, or specialized research into artificial intelligence.

Keywords: epistemic concepts, artificial intelligence, knowledge, semantics, normativity, agency, anthropomorphization

Введение

Современный этап развития технологий искусственного интеллекта характеризуется той особенностью, что различные исследователи, обладающие сопоставимой квалификацией и имеющие доступ к одному и тому же набору наблюдаемых фактов и свидетельств, релевантных для описания поведения такого рода моделей, склонны интерпретировать эти свидетельства принципиально различным образом. В результате в исследовательской среде сформировался широкий спектр мнений по поводу возможности и целесообразности приписывания таким моделям тех или иных состояний и способностей, ранее приписывавшихся лишь человеку [Simbolon et al., 2025].

Список этих категорий довольно широк: убеждения, знание, понимание, обоснование, умозаключение, распознавание, владение понятиями, обладание свидетельствами, выдвижение гипотез, поиск решений, планирование, обман, сомнение, уверенность и другие.

Заметим, что все эти понятия изначально принадлежали так называемому словарю «народной эпистемологии» [Hardy-Vallee, Dubreuil, 2010] и в этой связи имеют две особенности.

Первая особенность заключается в том, что все эти категории в той или иной мере сформировались в языке еще до того этапа, когда человек значительно продвинулся в теоретическом постижении областей теоретической эпистемологии, психологии и когнитивных наук. Слова, выражающие эти понятия, изначально не были терминами каких-либо теорий, но приобрели статус терминов значительно позже формирования практики их повседневного употребления.

Такая особенность происхождения влечет умножение смыслов. С одной стороны, смысл этих слов в первом приближении понятен любому носителю языка, и, более того, с лингвистической точки зрения именно носитель языка является последней инстанцией, к кому следует обращаться за их толкованием. Эти слова как термины повседневного словаря выражают то, что имеют в виду люди – широкий класс людей, носителей языка, – когда они используют их в своей речи. С другой стороны, эти слова, будучи терминами различных теорий, приобретают в рамках этих теорий различный смысл и условия их корректного употребления. Причем, что важно, речь идет именно о *различных* теориях и, как следствие, различных смыслах этих терминов. Тем самым конкуренция между теориями за статус наиболее адекватного способа описания и объяснения некоторого фрагмента действительности сопровождается конкуренцией между теориями за наиболее оправданное и обоснованное толкование соответствующих терминов и выражаемых ими понятий. При этом никакого универсального способа сопоставления различных теорий по степени их признанности и обоснованности не существует. Если две теории расходятся в своем толковании тех или иных терминов, то помимо такого расхождения между теориями по поводу трактовки терминов также возможно расхождение во мнениях среди сторонних квалифицированных наблюдателей по поводу того, какая из этих двух теорий лучше обоснована и является более предпочтительной.

В итоге мы имеем довольно парадоксальную картину семантики подобных терминов. С одной стороны, их смысл понятен любому носителю языка в той мере, в какой он не испытывает затруднений с их повседневным использованием [Вострикова, 2025, с. 227]. В самом деле, как бы он это делал, если бы не понимал их смысла? С другой стороны, по поводу их смысла и условий корректного применения идут оживленные дискуссии между исследователями,

которые далеки от достижения какого-либо консенсуса и преодоления разногласий между собой[1].

Вторая особенность понятий из приведенного выше списка заключается в том, что все они так или иначе встроены в определенную нормативную семантическую рамку. Их смысл раскрывается в контексте категорий рациональности, уместности, правильности, целесообразности и ряда иных нормативных категорий [Peacocke, 1999, p. 18]. Так, например, еще У. Селларс в своей работе «Эмпиризм и философия сознания» утверждал, что «существенный момент заключается в том, что, характеризуя некоторое состояние как состояние познания, мы не даем эмпирического описания этого состояния; мы помещаем его в логическое пространство доводов, обоснования и возможности обосновать сказанное» [Sellars, 1997, p. 76]. То есть помещаем в нормативную семантическую рамку, которая характеризует уместные коммуникативные реакции, но ничего не говорит о специфике внутренних состояний агентов.

При этом ясно, что хронологически человек усвоил эпистемические понятия задолго до того, как в его мышлении сформировалась и выделилась самостоятельная философская категория рациональности. И все же эта категория была изначально имплицитно встроена, «вшита» в смысл этих понятий народной эпистемологии, призванных объяснять наблюдаемое поведение людей и оценивать его на предмет правильности. То есть оценивать на предмет того, насколько в том или ином случае уместно или даже необходимо внести изменение в характер наблюдаемых действий кого-то из окружающих путем внесения изменений в его ненаблюдаемые эпистемические установки, которые лежат в основании всякого разумного и целенаправленного поведения.

Современный этап развития коммуникативных практик характеризуется трансформацией и расширением класса коммуникативных взаимодействий за счет включения в число их участников множества разнообразных искусственных систем. Соответственно, целью данной статьи является исследование того, в какой мере понятия народной эпистемологии могут быть распространены на модели искусственного интеллекта.

При этом искать ответ на этот вопрос можно двумя различными способами. Можно руководствоваться имеющимся у каждого

[1] Так, например, среди различных трактовок понятия желания в аналитической философии можно встретить как подходы, исходящие из того, что желания представляют собой подкласс убеждений (а именно, убеждений, что некое положение дел *p* является благом [Price, 1989]), так и подходы, согласно которым желания и убеждения относятся к двум различным, непересекающимся классам установок [Anscombe, 2000], первый из которых называется классом когнитивных установок, а второй – классом некогнитивных установок.

носителя языка интуитивным схватыванием эпистемических понятий и основной задачей исследования избрать вопрос о том, реализуют ли те или иные искусственные системы с учетом специфики их архитектуры, способа обучения и т.д. такие состояния, которые можно было бы охарактеризовать посредством данных эпистемических понятий. Например, можно ли сказать про шахматную нейросеть AlphaZero, что она помнит все сыгранные ею партии? Можно ли сказать, что управляемый специализированной нейросетью автономный автомобиль видит пакет на дороге? Для ответа на эти вопросы обращение к принципам устройства и функционирования той или иной модели может быть необходимым в том случае, если мы ясно и отчетливо понимаем, что значит «помнит» и «видит», и при этом нас интересует вопрос о том, обладают ли соответствующими способностями и состояниями данные конкретные системы.

Однако к достижению обозначенной выше цели можно подойти и несколько иным путем. В фокусе исследовательского внимания может быть не устройство той или иной технической системы, но вопросы о том, что значит «помнит» и что значит «видит». Эти вопросы не имеют непосредственного отношения к специфике философии искусственного интеллекта как таковой, но относятся к ведению двух других областей – эпистемологии и семантики[2]. При этом принятие той или иной точки зрения по поводу этих вопросов может значительно упростить исследование множества разнообразных частных случаев. Более того, мы можем сделать еще один шаг на пути абстрагирования и отвлечься от специфики не только тех или иных технических систем, но и от специфики отдельных эпистемических понятий, таких как «помнит» и «видит», и занять определенную теоретическую позицию в отношении семантики всех эпистемических понятий как таковых. Это даст нам наиболее обобщенный критерий приписываемости эпистемических понятий *любым* искусственным системам.

Этот критерий таков: эпистемическое понятие С может быть обоснованно приписано искусственной системе М в том случае, если эта система может быть обоснованно включена в систему нормативных эпистемических отношений, задаваемых понятием С.

Структура и характер этих нормативных отношений при этом определяются содержанием и спецификой того или иного конкретного эпистемического понятия. При этом в дальнейшем мы раскроем, при выполнении каких условий, на наш взгляд, искусственные системы оказываются способны удовлетворять данному обобщенному

[2] Также эти вопросы некоторым образом относятся к области философии сознания, однако этот ракурс проблематики мы в данном исследовании вынужденно обойдем вниманием.



критерию. Однако прежде обратимся к иллюстративному обзору позиций, встречающихся в литературе, по рассматриваемой нами проблематике.

Разнообразие позиций

Как уже было отмечено выше, спектр мнений по вопросу уместности приписывания системам искусственного интеллекта таких состояний, как убеждения, знания, понимание, владение понятиями и прочих, довольно широк [Simbolon et al., 2025]. Наш обзор этих мнений не претендует на полноту, но ориентирован лишь на то, чтобы показать, что различия в наблюдаемых позициях среди исследователей действительно существенны, а встречающиеся в литературе оценки порой диаметрально противоположны.

На одном конце спектра, который мы будем называть скептическим, можно найти множество исследователей[3], среди которых мы в качестве примера рассмотрим позицию Стефана Буйсмана, полагающего, что «мы не можем приписывать системам искусственного интеллекта владение понятиями» [Buijsman, 2025, p. 3]. Он обосновывает это тем, что такое владение должно предполагать возможность оперативного учета экспертного мнения: «Когда мы говорим об обращении к экспертам в случае понятий, мы имеем в виду механизм, в рамках которого применение понятий быстро корректируется, когда эксперт указывает на допущенную концептуальную ошибку. И хотя мы можем легко исправить такие ошибки на уровне выходных данных модели, нет аналогичного способа напрямую скорректировать соотношения весов, используемые системой. Мы не можем, например, изменить нейронную сеть так, чтобы карты важности признаков более точно соответствовали тому, что мы считаем релевантными признаками» [Ibid., p. 11]. Тем самым система, откликаясь на экспертное мнение, ориентирована в первую очередь на то, чтобы корректировать непосредственно сам ответ, а не распределение весов признаков. Со временем эти веса постепенно корректируются, но не так оперативно, как это происходит в случае реакции человека.

Это довольно серьезное возражение. Модели такого рода тем самым не просто «не понимают», но и *не демонстрируют поведения*,

[3] Наиболее типичное изложение скептической позиции можно найти в статье Мюррэя Шанахана, который утверждает, что «чат-боты на основе LLM, строго говоря, не обладают убеждениями. Чат-бот не обладает *подлинным* знанием (does not *really* know), что Бурунди находится к югу от Руанды» [Shanahan, 2024, p. 72].

характерного для понимания человеком. Однако на это можно ответить в духе описываемого ниже подхода, что отождествление всякого понимания с поведением, характерным именно для человеческого понимания, является проявлением антропоцентризма. Искусственные модели демонстрируют в таких случаях несколько иное поведение, однако это может лишь означать, что они *понимают иначе* и иначе овладевают понятиями. Они понимают медленнее. Однако какая-либо определенная скорость осуществления ревизий не является сущностной характеристикой понимания как такового. Понимание может протекать с различной скоростью.

При этом на противоположном, оптимистическом, краю рассматриваемого спектра мнений можно обнаружить таких исследователей, как Герман Каппелен и Джош Девер[4]. Они выдвигают следующий тезис, название которого на русский язык можно перевести как «во все тяжкие» (Whole Hog Thesis): «ChatGPT – это полноценный лингвистический и когнитивный агент наравне с человеком. Он способен осмысленно использовать язык, делать утверждения, задавать и отвечать на вопросы, предлагать гипотезы и отдавать команды. <…> Он может познавать мир (получая информацию из рассказов или рассуждая на основе уже имеющихся знаний). Он может строить планы и рассуждать о том, как их достичь. Он может предпринимать действия для реализации этих планов» [Cappelen, Dever, 2025a, p. 13].

В основе этого чрезвычайно амбициозного исследовательского проекта лежит так называемая гиперэкстерналистская[5] установка, согласно которой вопрос о том, является ли некоторое состояние, например, убеждением, определяется не внутренними факторами и характеристиками того объекта, которому это состояние приписывается, но характером его взаимодействия с миром [Shin, 2025, p. 1555] и, что наиболее важно для нашего изложения, «практикой предоставления и запроса оснований, в которую он встроен» [Cappelen, Dever, 2025b, p. 11].

По мнению Каппелена и Девера, в действительности проблема не в том, что мы склонны к антропоморфизации при описании поведения искусственных систем, но в том, что наши теории и фигурирующие в них термины слишком антропоцентричны [Cappelen, Dever, 2021, p. 69].

[4] К этой же части спектра можно, например, отнести Уинли Ма и Винсента Валтона [Ma, Valton, 2024, p. 7].

[5] Приставка «гипер-» в этом термине указывает на то, что речь идет о крайней, последовательно экстерналистской позиции в рамках дискуссий в таких областях, как эпистемология и семантика.

Антропоцентризм на примере доверия

Антропоцентризм являет собой представление о том, что определенные черты и аспекты тех или иных способов взаимодействия между людьми являются сущностными чертами этих способов взаимодействия, не допускающими их распространения на более широкий круг участников, включающий искусственные системы.

В качестве весьма характерной иллюстрации рассмотрим вопрос о возможности приписывания моделям искусственного интеллекта благонадежности. Как и по другим аналогичным вопросам, здесь среди исследователей наблюдается широкий спектр мнений[6]. На скептическом краю этого спектра можно встретить такого исследователя, как Кристиан Будник, который утверждает, что «вопрос о том, когда мы можем доверять цифровым технологиям, представляет собой своего рода категориальную ошибку: доверять цифровым технологиям вообще невозможно, поскольку цифровые артефакты, в первую очередь, не относятся к числу подходящих объектов доверия» [Budnik, 2025, p. 2]. Он обосновывает свое мнение тем, что доверие предполагает наличие нормативных ожиданий, что в свою очередь требует допущения о том, что сторона, на которую распространяются эти нормативные ожидания, способна к намеренным и ответственным действиям [Ibid., p. 7]. Обосновать же наличие у моделей искусственного интеллекта способности к намеренным и ответственным действиям, по его мнению, было бы крайне трудной задачей.

На противоположном краю спектра мнений по данному вопросу располагаются такие исследователи, как Мона Симион и Кристоф Келп. Они признают, что в литературе распространена точка зрения, согласно которой благонадежность представляет собой добродетель в морально релевантном смысле: «В то время как надежный (reliable) человек просто добросовестно выполняет свои обязательства, заслуживающий доверия (trustworthy) человек выполняет свои обязательства добросовестно в силу своей доброй воли по отношению к доверителю» [Simion, Kelp, 2023, p. 3]. Однако, как отмечают эти же авторы, трактовка подобного рода является чрезмерно антропоцентричной: «Правильно ли считать, что способность ИИ к тому, чтобы заслуживать доверия, зависит от его способности проявлять волю или демонстрировать добродетели характера? И в более общем смысле, следует ли нам, обнаружив, что конкретный артефакт,

[6] Стоит оговорить, что речь здесь идет не о конкретных существующих на сегодняшний день моделях LLM, которые в силу принципов своей архитектуры систематически склонны галлюцинировать, но о более отвлеченном вопросе критериев приписываемости благонадежности искусственным интеллектуальным системам как таковым.

который мы считали заслуживающим доверия, на самом деле не способен проявлять эти психологические черты, пересмотреть наше отношение к этому артефакту?» [Simion, Kelp, 2023, p. 3]. В качестве ответа на данный вопрос они приводят цитату из монографии Саймона Блэкберна: «Мы часто соглашаемся доверять, не зная многого о психологии того, кому доверяем, просто предполагая, что у него есть… способности, достаточные для выполнения соответствующей работы» [Blackburn, 1998]. В конечном итоге они утверждают, что заслуживающий доверия искусственный интеллект можно определить как «искусственный интеллект, который выполняет свои функционально обусловленные обязательства, причем последние могут быть либо заданы на этапе проектирования, либо обусловлены этиологически (etiologically determined) через преимущества, которые приносит соответствующий ИИ» [Simion, Kelp, 2023, p. 11].

Как видно, в рамках данного подхода способность заслуживать доверие объясняется через встраиваемость в структуру нормативных отношений определенного рода. В таком случае исследование наблюдаемых практик взаимодействия объектов с окружающей средой, а также их встраиваемости в нормативные отношения должно быть стартовой точкой в исследовании уместности приписывания искусственным объектам и системам эпистемических состояний [Alvarado, 2023].

Объекты и нормативные отношения

Рассмотрим, например, автономно управляющийся автомобиль. Способен ли такой автомобиль нарушать правила дорожного движения? С экстерналистской точки зрения это несложный эмпирический вопрос. Для поиска ответа на него достаточно в течение некоторого времени пронаблюдать за тем, что происходит с таким автомобилем на дорогах. Вполне представимо, что в определенный момент траектория его движения может оказать таковой, что ее вполне можно охарактеризовать как нарушение правил дорожного движения. Например, представимо, что автономно управляемый автомобиль выезжает на встречную полосу движения, пересекая при этом двойную сплошную разделительную черту.

Ясно, что столь непосредственная и прямая трактовка такого события может вызвать закономерные возражения, указывающие на то, что нарушением правил движения может быть лишь контролируемое и намеренное действие агента. Если же автомобиль выехал на встречную полосу вследствие механической поломки, то такое событие не будет нарушением правил просто потому, что оно не будет ничьим действием. Нарушением в данном случае может быть лишь



бездействие владельца автомобиля, вовремя не прошедшего положенный по регламенту технический осмотр своего транспортного средства.

Тем не менее мы можем дополнить наш сценарий, уточнив, что такая траектория движения была контролируемой и не была результатом механической поломки. Такое обстоятельство вполне можно установить по итогам проведенной экспертизы. В таком случае для оценки данного события нам, по-видимому, потребовалось бы оценить те основания, в силу которых автомобиль двигался по такой траектории. Например, исходя из отчетов такого автомобиля, нам может быть известно, что выбор в пользу указанной траектории был совершен на том основании, что любая иная физически реализуемая траектория привела бы к аварии с потенциально летальным исходом, вероятность которого была бы крайне высокой. В таком случае выезд на встречную полосу не является нарушением правил дорожного движения, поскольку правила допускают подобное в случае крайней необходимости.

Однако пусть мы теперь обнаруживаем, что основанием для выезда на встречную полосу было то, что автономно управляемый автомобиль принял черный полиэтиленовый пакет за лежащего на проезжей части человека и, как и в предыдущем случае, выбрал указанную траекторию во избежание возможных жертв. В этом случае никаких условий крайней необходимости в действительности не было и выбор такой траектории определенно является нарушением действующих правил.

Тем самым мы оказываемся перед непростым выбором терминов для описания происходящего. Мы можем либо согласиться с тем, что такой автомобиль может анализировать поступающие данные и свидетельства, способен оценивать их нормативный вес, а также в состоянии на основании этих свидетельств определять траекторию своего контролируемого движения и тем самым может либо следовать правилам дорожного движения, либо гипотетически способен нарушать их. Либо – и в этом заключается вторая доступная нам опция – мы можем утверждать, что использование подобных терминов для описания данной ситуации является проявлением антропоморфизации [Scorici et al., 2024] и тем самым является некорректным способом описания происходящего.

Основных проблем со второй опцией две.

Первая проблема заключается в том, что у нас может не быть другого, более подходящего языка описания и объяснения подобного рода событий. Объяснение, согласно которому автономный автомобиль выехал на встречную полосу потому, что принял лежащий на земле полиэтиленовый пакет за живого человека, может оказаться наиболее адекватным, информативным и релевантным по сравнению с любым иным способом описания и объяснения этих же событий.

Вторая проблема заключается в том, что если структура описания оснований, учитываемых системой искусственного интеллекта, встроенной в автомобиль, совпадает со структурой оснований, учитываемой человеком, который мог бы находиться на месте водителя, то у нас нет причин для отказа приписывать искусственному интеллекту те состояния, природа которых сводится к определенным уместным реакциям на соответствующие типы фактов.

Гиперэкстернализм

Стоит признать, что гиперэкстерналистский подход к толкованию терминов словаря «народной эпистемологии» исходит из того, что поведенческая имитация эпистемических состояний (при выполнении ряда условий, о которых речь пойдет ниже) тождественна обладанию ими. Это в буквальном смысле одно и то же. Нет никакого довеска, которым «подлинное» знание или понимание могли бы отличаться от демонстрируемой в поведении «имитации». Если вы, например, научились идеально имитировать способность сомневаться в своих ответах, включая правильное воспроизводство влияния такого имитируемого сомнения на характер и содержание ваших ответов, то вы приобрели способность сомневаться в собственных ответах. С гиперэкстерналистской точки зрения сомнение – это не то, что качественно переживается агентами где-то внутри глубин их сознания, но то, что проявляет себя в актах коммуникации между агентами и определяет уместность тех или иных реакций при их взаимодействии друг с другом. Если характер коммуникативного взаимодействия между агентами и его уместные последствия таковы, что всякий носитель языка с готовностью опишет их как проявление сомнения, то, значит, такое взаимодействие может быть обоснованно охарактеризовано таким образом. С гиперэкстерналистской точки зрения понятие сомнения определяется не характером переживаний вовлеченных агентов, но той структурой нормативных отношений, которая определяет, какие коммуникативные реакции являются уместными в ответ на то или иное коммуникативное действие. В свою очередь, противоположной позицией будет та, согласно которой в основе сомнения – а также знания, понимания, желаний и других подобных состояний – находится не их нормативная роль в рамках механизмов порождения поведения и уместных коммуникативных реакций, но качественная специфика обладания ими, переживаемая из перспективы первого лица. Такая позиция с точки зрения экстернализма приводит к неоправданному сужению класса объектов, которым могут быть обоснованно приписаны соответствующие состояния и способности, что и представляет собой антропоцентризм.



В качестве дополнительной иллюстрации, раскрывающей суть рассматриваемой экстерналистской позиции, мы можем рассмотреть вопрос о том, умеет ли калькулятор умножать числа. По этому поводу возможны по меньшей мере три точки зрения. Первая из них предполагает утвердительный ответ. Умножение чисел – это одна из основных функций калькулятора, с которой имеющиеся в мире калькуляторы систематически хорошо справляются [Humphreys, 2004, p. 5]. Таких наблюдаемых свидетельств вполне достаточно для того, чтобы дать утвердительный ответ на этот вопрос. Вторая возможная позиция – скептическая. Она предполагает, что приписывание калькуляторам способности умножать числа является некорректным и необоснованным, поскольку калькуляторы не обладают понятием числа, не имеют соответствующих репрезентаций [Бажанов, 2025, с. 141] и не обладают пониманием арифметических отношений. Вычислительные процессы, происходящие с калькуляторами, являются каузальными, физическими процессами. Приписывание им когнитивных способностей оправданно не более, чем приписывание таких способностей механическим счетам. Такое приписывание представляет собой необоснованную антропоморфизацию [Arsenault, 2026].

Однако по этому вопросу возможна также третья позиция: калькуляторам действительно неуместно приписывать наличие когнитивных способностей, но не потому, что они не владеют понятием числа и не обладают пониманием арифметических отношений, но потому, что они не способны к вариативному контролируемому поведению, не способны к оценке свидетельств и вследствие этого не являются агентами, чьи действия могут быть вписаны в структуру нормативных отношений. Калькуляторы не могут получать и оценивать свидетельства, не могут ошибаться в этих своих оценках иначе, чем в силу каузальных сбоев, и не могут оценивать вероятность собственной ошибки. В свою очередь, модели искусственного интеллекта такими способностями обладают [Cash et al., 2025, p. 397]. Их ошибки – это не результат сбоев и поломок, но результат неверной оценки доступных свидетельств. Рассматривая ответы моделей таким образом, мы тем самым предполагаем, что они могли бы быть точнее в своих ответах. Такая опция проявления немного иного, более обоснованного поведения является для них номологически возможной. Именно поэтому их коммуникативное поведение может быть вписано в структуру нормативных отношений, что позволяет приписывать им эпистемические понятия, содержание которых полностью описывается их ролью в этой структуре.

Вариативность поведения и отзывчивость к основаниям

Ясно, что такое расширение круга участников нормативных отношений следует проводить с осторожностью. При этом требуется уточнить и прояснить критерии такого включения, которые согласовывались бы со здравым смыслом. В противном случае мы можем обнаружить, что практика нашего словоупотребления, в частности, позволяет утверждать, что некоторые здания способны нарушать те или иные действующие нормы – например, нормы высотной застройки. В свою очередь, некоторые законы способны нарушать нормы международного права. Причем они способны их нарушать даже в том случае, если люди, написавшие и принявшие эти нормативные акты, действовали в установленном законом порядке и ничего при этом не нарушали. По-видимому, речь в данном случае идет о побочных эффектах языка как естественного и потому несовершенного феномена. Эти эффекты не следует трактовать как свидетельства в пользу того, что здания, фотографии или видеоизображения сами способны вступать в нормативные отношения и сами способны что-либо нарушать или чему-либо соответствовать.

Способность к нарушению, следованию и соответствию целесообразно сохранить исключительно за так называемыми нормативными агентами, т.е. агентами, способными к обладанию и проявлению собственных нормативных установок, а также способными к проявлению отзывчивости по отношению к нормативным основаниям.

Ключевой характеристикой агентности является способность к вариативности поведения и контролю за реализацией стратегии поведения, выбранной из множества доступных альтернатив [Cochrane, 2018, p. 185].

В свою очередь, вариативность проявлений активности предполагает, что события, случающиеся с объектом, имеют место не в рамках линейной каузальной цепочки, а в рамках ветвящихся каузальных траекторий. Спонтанные отклонения изначальных параметров в сочетании с высочайшей чувствительностью к этим параметрам, характерной для нелинейных процессов, приводят к тому, что, например, две идентичные клетки, находящиеся в одинаковых начальных условиях, могут двигаться по различным траекториям. Причем такое различие является не результатом ограниченности нашего знания, а следствием того, как функционируют физические системы определенного рода вдали от термодинамического равновесия.

Однако наличия такой вариативности и фундаментальной непредсказуемости в проявлении объектом той или иной активности еще недостаточно для полноценной агентности. Вариативность



проявлений активности означает только то, что объект способен двигаться по различным траекториям из одного и того же исходного состояния. Ключевой шаг происходит тогда, когда объект в силу своей природы и структурной организации находится не просто перед лицом ветвящихся траекторий, но способен влиять – а в пределе и полностью определять, – какая из доступных для него траекторий будет в конечном итоге реализована в мире. Достигается это за счет разнообразных механизмов обратной связи, устроенных таким образом, что одни малейшие отклонения усиливаются, тогда как другие случайные отклонения погашаются.

Ключевым обстоятельством здесь является то, что, хотя различные траектории действительно могут быть реализуемы из одного и того же исходного состояния объекта и его среды, ни одна из этих траекторий не выходит за рамки номологически возможного, т.е. за рамки совместимого с имеющимися законами природы.

Наконец, когда живые системы становятся достаточно сложными, они приобретают способность к моделированию окружающей среды. Выбор траектории из числа доступных альтернатив в этом случае становится не просто слепой физической реакцией объекта на внешний стимул, но результатом учета возможных последствий такого выбора в рамках проводимого моделирования. Если говорить в терминах концепции Д. Деннета, то на этом этапе «скиннеровские» существа становятся «попперовскими» [Dennett, 1997]. Такой переход на следующий уровень в структурном развитии позволяет системам, которые уже с полным правом можно назвать агентами [Bratman, 2007, p. 62–63], формировать гипотезы в рамках имеющейся у них модели действительности. Теперь агенты могут позволить гипотезам погибать вместо себя. Здесь вариативность перестает быть слепой. Альтернативные траектории поведения могут быть «проиграны» агентом в его модели действительности до того, как какая-либо из них будет реализована в физическом мире.

Структурными элементами такого рода моделей действительности, которыми обладают агенты, являются установки этих агентов.

Установки – это интенциональные состояния агентов, т.е. состояния, описываемые как имеющие некоторое содержание, «состояния о чем-то».

Ключевая особенность этого определения заключается в том, что речь здесь идет о возможности описания некоторых состояний агента определенным образом (как имеющих некоторое содержание). Такого рода условие позволяет отождествлять приписываемость тех или иных установок некоторому агенту с их наличием у агента. Иметь установку и быть тем, кому эта установка может быть обоснованно приписана, – это в таком случае буквально одно и то же. В чем же причина такого нашего стремления отождествить, казалось бы, различные вещи?

Дело в том, что все разнообразные термины, выражающие установки, выполняют в естественном языке одну и ту же функцию – их приписывание позволяет агентам объяснять и предсказывать наблюдаемые действия друг друга. Такого рода объяснения и предсказания составляют самую суть такого явления, как «народная психология», а сами установки представляют собой основу ее словаря. При этом вопрос об уместности приписывания агентам тех или иных установок – это вопрос о соотношении наблюдаемого нами поведения этих агентов и разделяемых нами установок народной психологии. Сами термины, выражающие в языке такие установки, появляются в нем и осваиваются агентами ради объяснения и предсказания поведения *других* людей, а не самих себя. Это так потому, что для знания собственных психологических состояний и объяснения мотивов собственного поведения агенту, вообще говоря, не требуется язык. Ему незачем было бы описывать самому себе мотивы собственных действий. Однако научившись обоснованно применять термин «хочет» к другим людям, агент в какой-то момент оказывается способен применить его и к себе и тем самым оказывается способен объяснить себе и окружающим собственные действия.

Интерпретативистские подходы обладают одним весьма значимым преимуществом по сравнению со своими конкурентами. Они позволяют сравнительно простым и понятным образом ответить на вопрос о том, каким образом какие-либо состояния, события или процессы, случающиеся с агентами, могут «быть о чем-то», отличном от этих процессов, о чем-то, находящемся вне тел агентов, а, быть может, находящемся на значительном от них удалении как в пространстве, так и во времени. Ответ интерпретативизма заключается в том, что мы можем описывать эти состояния агентов так, *как будто* они представляют собой состояния «о чем-то». Причем практики такого нашего приписывания некоторым состояниям агентов того или иного содержания будут обоснованными в той мере, в какой такие приписывания позволяют нам корректно объяснять и предсказывать поведение данных агентов.

Приписывая агентам пропозициональные установки, мы приписываем им уместность реагирования определенным образом при выполнении тех или иных условий, характеризуемых в терминах интенционального содержания. Тем самым мы вписываем агентов в нормативный контекст.

Условия уместности включения в нормативные отношения

Таким образом, уместность приписывания системам искусственного интеллекта понятий из словаря «народной эпистемологии» определяется выполнением условий для того, чтобы их можно было характеризовать в качестве нормативных эпистемических агентов [Sosa, 2013]. При этом наша склонность приписывать кому-либо ту или иную установку не является безошибочной. Соответственно, «уместность приписывания» представляет собой нормативное понятие, содержание которого выражается в том, что идеально рациональный и информированный агент в рассматриваемой ситуации был бы склонен приписать некоторому другому агенту – в нашем случае искусственному интеллекту – определенную пропозициональную установку.

Это, в свою очередь, требует выполнения следующих условий: (а) наличие у систем искусственного интеллекта способности к вариативному поведению; (б) наличие у них контроля за выбором реализуемой траектории поведения; (в) наличие у них способности к взвешиванию оснований при выборе реализуемой траектории из числа номологически доступных альтернатив.

Эти условия позволяют отличать осмысленное и целенаправленное поведение от детерминированных последовательностей физических событий, в отношении которых неуместно вынесение нормативной оценки, как это имеет место, например, в случае физических процессов, реализующих вычисления в калькуляторе.

При этом выполнение данных условий является принципиально доступным для проверки.

Во-первых, существующие модели искусственного интеллекта демонстрируют способность к реализации различных форм поведения (ответов) при одинаковых внешних стимулах (запросах) [Wang et al., 2023]. В основе такой вариативности, так же как и в случае вариативности поведения живых организмов, лежат принципы нелинейной динамики.

Во-вторых, модели искусственного интеллекта демонстрируют способность придерживаться определенной стратегии поведения в течение некоторого времени. Стоит, конечно, признать, что параметр так называемой агентной стабильности (agent stability) остается пока одним из наиболее уязвимых мест для существующих в настоящий момент искусственных систем [Milliere, Buckner, 2024, p. 32]. И все же при всех подобного рода оговорках мы имеем все основания говорить об имеющей у них минимальной агентной стабильности, поскольку в противном случае мы наблюдали бы перманентную непоследовательность в их поведении, чего в действительности не происходит.

В-третьих, эти модели демонстрируют отзывчивость к предъявляемым им основаниям [Steyvers et al., 2025, p. 222]. Если при взаимодействии с ними им предоставить ссылку на авторитетные данные, вступающие в прямое противоречие с выбранной ими стратегией дискурсивного поведения, то они склонны релевантным образом реагировать на такие свидетельства и менять свою стратегию поведения.

Выполнение этих трех условий свидетельствует о том, что существующие на сегодняшний день модели искусственного интеллекта могут быть охарактеризованы в качестве нормативных эпистемических агентов. Они способны вступать в полноценные нормативные эпистемические отношения с другими эпистемическими агентами, включая людей. Это позволяет обоснованно приписывать им широкий спектр эпистемических понятий с тем, чтобы наиболее информативным и релевантным образом объяснять их дискурсивное поведение для других эпистемических агентов.

Список литературы

Бажанов, 2025 – *Бажанов В.А.* Феномен числа с позиций психологизма: можно ли утверждать, что числа «сочиняются»? // Эпистемология и философия науки. 2025. Т. 62. № 1. С. 134–147.

Вострикова, 2025 – *Вострикова Е.В.* Знание и контекст. Размышления о книге Майкла Бломе-Тилльманна «Семантика атрибуции знания» // Эпистемология и философия науки. 2025. Т. 62. № 1. С. 215–233.

References

Alvarado, 2023 – Alvarado, R. “AI as an Epistemic Technology”, *Science and Engineering Ethics*, 2023, vol. 29, pp. 1–18.

Anscombe, 2000 – Anscombe, E. *Intention*. Cambridge, MA: Harvard University Press, 2000, 94 p.

Arsenault, 2026 – Arsenault, L.M. “Humanizing: On the Anthropomorphization of AI as a Tactic of Cultural Integration”, *AI & Society*, 2026, vol. 41, pp. 3789–3799.

Bazhanov, 2025 – Bazhanov, V.A. “Fenomen chisla s pozitsii psikhologizma: mozhno li utverzhdat’, chto chisla ‘sochinyayutsya’?” [The Phenomenon of Number From The Lens of Psychologism: Can We Claim that Numbers Are Being “Made Up”?], *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 1, pp. 134–147. (In Russian)

Blackburn, 1998 – Blackburn, S. *Ruling Passion: A Theory of Practical Reasoning*. Oxford: Clarendon Press, 1998.

Bratman, 2007 – Bratman, M. *Structures of Agency: Essays*. Oxford: Oxford University Press, 2007.

Budnik, 2025 – Budnik, C. "Can We Trust Artificial Intelligence?", *Philosophy & Technology*, 2025, vol. 38, pp. 1–23.

Buijsman, 2025 – Buijsman, S. "Conceptual Disconnect: The Prospect of Concept-Based AI Explanations", in: Pantsar, M., Stjernfelt, F., Gramelsberger, G., Olteanu A. (eds.) *Philosophy of Artificial Intelligence: Optimistic and Pessimistic Views*. Cham: Springer, 2025, pp. 1–16.

Cappelen, Dever, 2021 – Cappelen, H., Dever, J. *Making AI Intelligible: Philosophical Foundations*. Oxford: Oxford University Press, 2021, 175 pp.

Cappelen, Dever, 2025a – Cappelen, H., Dever, J. *Going Whole Hog: A Philosophical Defense of AI Cognition*, 2025. DOI: 10.48550/arXiv.2504.13988.

Cappelen, Dever, 2025b – Cappelen, H., Dever, J. "A Hyper-Externalist Manifesto for LLMs", in: Cappelen, H., Dever, J. (eds.) *Communicating with AI: Philosophical Perspectives*. Oxford: Oxford University Press, 2025, pp. 1–21.

Cash et al., 2025 – Cash, T., Oppenheimer, D., Christie, S., Devgan, M. "Quantifying Uncert-AI-nty: Testing the Accuracy of LLMs' Confidence Judgments", *Memory & Cognition*, 2025, vol. 54, pp. 375–400.

Cochrane, 2018 – Cochrane, T. *The Emotional Mind: A Control Theory of Affective States*. Cambridge: Cambridge University Press, 2018.

Dennett, 1997 – Dennett, D. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books, 1997.

Hardy-Vallee, Dubreuil, 2010 – Hardy-Vallee, B., Dubreuil, B. "Folk Epistemology as Normative Social Cognition", *Review of Philosophy and Psychology*, 2010, vol. 1, pp. 483–498.

Humphreys, 2004 – Humphreys, P. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press, 2004.

Ma, Valton, 2024 – Ma, W., Valton, V. "Toward an Ethics of AI Belief", *Philosophy & Technology*, 2024, vol. 37, pp. 1–28.

Milliere, Buckner, 2024 – Milliere, R., Buckner, C. "A Philosophical Introduction to Language Models – Part II: The Way Forward", 2024, DOI: 10.48550/arXiv.2405.03207.

Shanahan, 2024 – Shanahan, M. "Talking about Large Language Models", *Communications of the ACM*, 2024, vol. 67, no. 2, pp. 68–79.

Peacocke, 1999 – Peacocke, C. *Being Known*. Oxford: Oxford University Press, 1999.

Price, 1989 – Price, H. "Defending Desire-as-Belief", *Mind*, 1989, vol. 98, pp. 119–127.

Scorici et al., 2024 – Scorici, G., Schultz, M.D., Seele, P. "Anthropomorphization and Beyond: Conceptualizing Humanwashing of AI-enabled Machines", *AI & Society*, 2024, vol. 39, pp. 789–795.

Sellars, 1997 – Sellars, W. *Empiricism And The Philosophy Of Mind*. Cambridge, MA: Harvard University Press, 1997.

Shin, 2025 – Shin, D. Automating Epistemology: How AI Reconfigures Truth, Authority, and Verification", *AI & Society*, 2025, vol. 41, pp. 1553–1559.

Simbolon et al., 2025 – Simbolon, L., Manugeren, M., Barus, E. "Does AI Know Things? An Epistemological Perspective on Artificial Intelligence", *Journal of English Language and Education*, 2025, vol. 10, pp. 1022–1028.

Simion, Kelp, 2023 – Simion, M., Kelp, C. “Trustworthy Artificial Intelligence”, *Asian Journal of Philosophy*, 2023, vol. 2, pp. 1–12.

Sosa, 2013 – Sosa, E. “Epistemic Agency”, *The Journal of Philosophy*, 2013, vol. 110, pp. 585–605.

Steyvers et al., 2025 – Steyvers, M., Tejeda, H., Kumar, A., Belem, C., Karny, S., Hu, X., Smyth, P. “What Large Language Models Know and What People Think They Know”, *Nature Machine Intelligence*, 2025, vol. 7, pp. 221–231.

Vostrikova, 2025 – Vostrikova, E. “Znanie i kontekst. Razmyshleniya o knige Maikla Blome-Till’manna “Semantika atributsii znaniya” [Knowledge and Context. Reflections on Michael Blome-Tillmann’s “Semantics of Knowledge Attribution”], *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 1, pp. 215–233. (In Russian)

Wang et al., 2023 – Wang, C., Liu, X, Awadallah, A. “Cost-Effective Hyperparameter Optimization for Large Language Model Generation Inference”, *Proceedings of Machine Learning Research*, 2023, vol. 224, pp. 1–17.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 113–131
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 113–131
DOI: https://doi.org/10.5840/eps202663344

К НОВЫМ ПОДХОДАМ В ИССЛЕДОВАНИИ МАШИННОГО СОЗНАНИЯ*

Гришин Павел Викторович – стажер-исследователь. Национальный исследовательский университет «Высшая школа экономики». Российская Федерация, 105066, г. Москва, ул. Старая Басманная, д. 21/4, стр. 1;
e-mail: pvgrishin@edu.hse.ru



Статья посвящена современным подходам к исследованию сознания ИИ и их методологическим пределам. Анализ выстраивается на основе классификации Джонатана Бёрча, в рамках которой рассматриваются теоретико-нагруженные, теоретико-нейтральные и теоретико-легкие подходы; отдельно обсуждаются многомерные и многоуровневые стратегии. Исходной установкой исследования выступает позитивный агностицизм относительно возможности машинного сознания. Показано, что, несмотря на существенные различия между существующими подходами, большинство из них в недостаточной степени проблематизирует перенос понятий и методов, разработанных для изучения сознания биологических существ, на вычислительные системы. В связи с этим утверждается, что центральная трудность современного исследования заключается в недостаточной обоснованности указанного переноса. Нехватка эмпирических данных или отсутствие надежного теста, при всей их значимости, не являются ключевыми трудностями. Особое внимание уделяется вычислительному функционализму, который либо явно принимается в качестве предпосылки исследования, либо неявно задает его рамку, либо становится предметом критики. Обосновывается, что наиболее продуктивными в этих условиях являются многомерные и многоуровневые подходы, поскольку они позволяют локализовать ключевые методологические трудности и представить их в виде более частных исследовательских задач, не претендуя на окончательное решение проблем. Делается вывод, что пределы современных исследований машинного сознания во многом определяются унаследованной понятийной рамкой; в этой связи предлагается общая исследовательская установка для дальнейшей работы над проблемой.

Ключевые слова: сознание, искусственное сознание, машинное сознание, ИИ, философия сознания, проблема других сознаний, вычислительный функционализм, антропоцентризм, агностицизм

* Исследование выполнено за счет гранта Российского научного фонда № 26-18-00253, https://rscf.ru/project/26-18-00253/. Russian Science Foundation, RSF. https://rscf.ru/en/project/26-18-00253

© Гришин П.В., 2026

TOWARDS NEW APPROACHES IN MACHINE CONSCIOUSNESS RESEARCH

Pavel V. Grishin –
Research Assistant.
HSE University.
21/4 Staraya Basmannaya St.,
Moscow 105066,
Russian Federation;
e-mail: pvgrishin@edu.hse.ru

This article examines contemporary approaches to the study of AI consciousness and their methodological limits. The analysis is structured around Jonathan Birch's classification, focusing on theory-heavy, theory-neutral, and theory-light approaches, while also treating multidimensional and multilevel strategies as a distinct line of inquiry. The starting point of the article is a position of positive agnosticism regarding the possibility of machine consciousness. It is argued that, despite their significant differences, most existing approaches do not sufficiently problematize the transfer of concepts and methods developed for the study of consciousness in biological beings to computational systems. On this basis, it is argued that the central difficulty in the contemporary research lies in the insufficient justification of the aforementioned transfer. The lack of empirical data or the absence of a reliable test, while significant, do not constitute the key difficulties. Particular attention is paid to computational functionalism, which is either explicitly adopted as a presupposition, implicitly built into the framework of inquiry, or made the object of criticism. In these circumstances, multidimensional and multilevel approaches prove most productive, as they allow one to localize key methodological difficulties and reframe them as more specific research tasks, without claiming to offer a definitive resolution of the problems. The conclusion is that the limits of current research on machine consciousness are shaped to a large extent by an inherited conceptual framework; accordingly, the article proposes a general research orientation for further work on the problem.

Keywords: consciousness, artificial consciousness, machine consciousness, AI, philosophy of mind, philosophy of consciousness, problem of other minds, problem of other consciousness, computational functionalism, anthropocentrism, agnosticism

Введение

Вопрос о сознании ИИ в последние годы перестал быть маргинальным. Быстрый рост возможностей больших языковых моделей сделал эту тему не только общефилософской, но и методологической [Bayne et al., 2024; Butlin et al., 2023; Butlin et al., 2025; Evers et al., 2024; Schwitzgebel, 2025; Shevlin, 2021]. Однако современная дискуссия развивается в условиях двойной неопределенности. С одной стороны, нет общепринятой теории сознания даже для человека. С другой стороны, именно сознание человека по-прежнему служит главным источником понятий, тестов и теорий, которые затем пытаются переносить на искусственные системы.

Из-за этого почти все исследования сознания ИИ строятся вокруг одного и того же хода – переноса. Исследователь либо ищет



выделяемые современной наукой о сознании функциональные и архитектурные пересечения [Butlin et al., 2023; Goldstein, Kirk-Giannini, 2024; Juliani et al., 2022], либо, напротив, подчеркивает, что такой перенос пока не оправдан из-за биологической специфики сознания [Aru et al., 2023; Block, 2025; Findlay et al., 2024; Seth, 2025]. Даже более осторожные программы – от теоретико-нейтральных тестовых стратегий, вроде ACT-теста [Schneider, 2019; Schneider, 2020], до теоретико-легких эмпирических проверок [Dung, 2023], а также многомерных подходов [Chalmers, 2023; Evers et al., 2024; Shevlin, 2021] – все еще работают внутри словаря, выработанного в исследованиях человека и животных.

Центральная трудность находится глубже чем недостаточность оснований для уверенного ответа на вопрос, сознательны ли современные ИИ, и на вопрос, способна ли машина иметь сознание в принципе. Она состоит в легитимности использования базовых понятий этой дискуссии применительно к искусственным системам, так как наше вопрошание по-прежнему остается в рамках переноса понятий, выработанных для сознания биологического типа. Особую уязвимость здесь демонстрирует понятие феноменального опыта и отчета о нем - к этому мы еще вернемся.

Если сознание ИИ вообще возможно, то весьма вероятно, что оно потребует не частичной модификации готового аппарата, а его радикального пересмотра. Такая ревизия в статье не производится, но все же выделяются общие установки исследования, следуя которым содержание и состав существующих понятий должны измениться.

Статья построена следующим образом. Сначала задаются предпосылки анализа. Затем поочередно разбираются подходы к проблеме других сознаний по Бёрчу (в контексте ИИ). Далее отдельный раздел кратко обозревает дискуссию вокруг вычислительного функционализма (*computational functionalism*). После этого обсуждаются перспективные агностические и многомерные ответы на нынешние затруднения. В заключении формулируется общая проблема переноса понятий и намечается возможный вектор дальнейшего исследования.

Контекст и предпосылки

Прежде всего необходимо уточнить, о каком именно круге вопросов идет речь. Философия сознания (*philosophy of consciousness*) наследует словарь, методы и проблемы философии разума (*philosophy of mind*). Однако фактически философов сознания интересует другой класс вопросов. В то время как философия разума изучает, например,

репрезентации сами по себе, философию сознания они интересуют лишь в той мере, в какой имеют отношение к феноменальному опыту, вопросу «каково это быть», глобальному рабочему пространству и другим концептам сознания. Поэтому перевод mind как сознания является неточностью, преодоление которой входит в задачу строгого исследования[1]. По этой причине далее речь пойдет исключительно о проблеме других сознаний, а не о проблеме других разумов.

Проблема других сознаний гораздо более детально проработана в контексте животных, чем в контексте ИИ. Поэтому смежная литература по сознанию животных полезна здесь не как готовое решение, а как источник методологических предупреждений. Так, Джонатан Бёрч подчеркивает, что изучение нечеловеческого сознания требует специальных стратегий, поскольку имеющиеся модели сознания невозможно механически расширять на другие системы [Birch, 2022]. Генри Шевлин называет эту трудность *проблемой специфичности* [Shevlin, 2021]. Если теория формулируется слишком близко к когнитивному устройству человека, она дает ложные отрицания: *справедливо ли заключить, что осьминоги не имеют сознания, только потому, что морфология их мозга кардинально отличается от морфологии мозга человека?* Если формулируется слишком абстрактно, она начинает порождать ложные положительные результаты: *можно ли назвать ИИ сознательным при условии, что его архитектура в некоей степени функционально повторяет мозг человека?* Для ИИ проблема специфичности стоит особенно остро, поскольку разрыв между человеком и машиной больше, чем между человеком и любым животным.

Эти трудности получают дальнейшее развитие у Хизер Браунинг и Уолтера Вейта, предлагающих разделять *проблему валидности индикаторов* и *проблему экстраполяции* [Browning, Veit, 2020]. Смысл первой состоит в том, что маркер сознания может оказаться ненадежным: трудно гарантировать, что он фиксирует именно сознательный опыт, а не сопутствующий когнитивный процесс. Иными словами, еще до попытки переноса необходимо доказать, что выбранный маркер действительно имеет устойчивую связь с сознанием, а не просто коррелирует с ним в случае человека. Проблема экстраполяции возникает на следующем шаге: даже если маркер хорошо работает на примере людей, это не гарантирует его релевантность для животных и тем более для искусственных систем. Это

[1] Автор может рассматривать вопрос о существовании врожденных языковых механизмов – а это дискуссия из области философии разума. В русской литературе его текст с большой вероятностью окажется под рубрикой философии сознания, хотя речь идет исключительно о бессознательных процессах. Это абсурдная ситуация.

разграничение помогает локализовать методологические препятствия: оно четко разводит сомнения в надежности самих индикаторов и сомнения в правомерности их переноса на принципиально иные системы.

Нужно учитывать и материальный разрыв между человеком и машиной: в одном случае речь идет о живой биологической ткани, состоящей из клеток, белков, воды и других органических структур, в другом – об устройствах, построенных из кремния, металлов и иных неорганических компонентов. Уже по одному этому основанию любое животное ближе к человеку, чем нынешние системы ИИ. Машина может демонстрировать языковые компетенции – способность, являющуюся одной из главных отличительных черт человека. При этом она остается радикально чуждой по своему устройству. И напротив, существо, вовсе не владеющее языком, может быть намного ближе к человеку как биологический родственник и носитель сходной формы организации нервной системы и тела вообще. *Отсюда возникает парадоксальная асимметрия, которую важно иметь в виду.*

Из-за этой асимметрии совпадение по поведению или по части функций еще не дает оснований предполагать более глубокое сходство в релевантных для сознания вещах. Вместе с тем этот кардинальный разрыв между человеком и машиной частично сужается. Переход от символического ИИ к коннекционизму сделал его в некотором смысле ближе к биологическим системам. Однако это сходство остается грубым и частичным: глубокие нейросети заимствуют лишь общий принцип иерархичности. Основным методом обучения служит обратное распространение ошибки – алгоритм, биологическая достоверность которого остается спорной [Lillicrap et al., 2020; Whittington, Bogacz, 2019]. Рекуррентность лишь частично приближается в отдельных разработках вроде спайковых нейронных сетей, однако они пока сохраняют лабораторный статус [Pfeiffer, Pfeil, 2018]. Поэтому методологическая задача состоит не в простой фиксации несходства ИИ и человека, а в более точном анализе того, какие различия здесь действительно принципиальны и какие сходства вообще можно считать значимыми.

На этом фоне наиболее оправданной исходной позицией выглядит *агностицизм*. Он фиксирует, что имеющиеся данные пока не позволяют обоснованно перейти ни к уверенному признанию, ни к уверенному отрицанию этой возможности [McClelland, 2024; Schwitzgebel, 2025]. Вместе с тем агностицизм не обязательно должен быть нейтральным. В рамках обзора он понимается скорее как позитивный агностицизм – признание нынешней неопределенности при одновременной открытости к тому, что искусственное сознание принципиально возможно, даже если условия его возникновения и критерии верификации еще не прояснены. *Если искусственное*

сознание возможно, то ключевой вопрос не в том, будет ли оно «настоящим», т.е. как у человека. Но в том, какова специфика вычислительных аналогов сознательных процессов и какими средствами эту специфику вообще можно исследовать.

В обзор включены наиболее показательные линии современной дискуссии о сознании ИИ. В них особенно ясно проявляется вопрос о правомерности переноса готовых философских и научных понятий сознания с биологических систем на компьютерные. Новая классификация подходов здесь не предлагается. В качестве рабочей используется классификация Бёрча [Birch et al., 2022]. *Теоретико-нагруженный* подход ищет строгие соответствия между взятой теорией сознания и архитектурой рассматриваемой системы. *Теоретико-нейтральный* подход, напротив, отказывается от теории и, как правило, оценивает сознание на основе поведения. *Теоретико-легкий* вместо структурных соответствий полагается на центральную и проверяемую эмпирически гипотезу.

Эта классификация ставит в центр отношение того или иного подхода к теории сознания, что делает ее особенно уместной: научное исследование сознания неизбежно строится вокруг теорий, тогда как сама наука о сознании еще находится в младенческом состоянии и не располагает общепризнанными, хорошо подтвержденными теориями. В таких условиях именно отношение подхода к теории предопределяет обоснованность получаемых выводов. Критерием оценки служит то, насколько обоснованно тот или иной подход работает с разрывом между человеческим и искусственным сознанием и насколько в нем отрефлексированы ограничения, связанные с переносом понятий.

Теоретико-нагруженный подход и его цена

Наиболее влиятельное и систематическое воплощение теоретико-нагруженный подход получил в недавних работах группы исследователей под руководством Патрика Батлина [Butlin et al., 2023]. Выводя из нескольких научных теорий сознания ряд функциональных и архитектурных требований, они проверяют, действительно ли современный ИИ выполняет эти требования. В этом же состоит и более поздняя формулировка метода [Butlin et al., 2025].

Главное преимущество такого подхода – перевод абстрактных споров о машинном сознании в плоскость анализа конкретных архитектур и функций. С точки зрения теории глобального рабочего пространства (GWT) содержание сознания должно быть глобально доступно всем функциональным модулям системы. Например, попавшая через модуль зрения в рабочее пространство зрительная

информация доступна модулю производства речи – люди могут рассказать, что они в данный момент видят[2]. Аналогичным образом из других научных теорий выводятся собственные индикаторные свойства. Благодаря этому исследование приобретает вид систематичной научной программы: проверка того, какие именно теоретически значимые механизмы сознания реализованы в конкретной модели ИИ.

Внутри этого направления важно различать два разных исследовательских хода. Первый – это поиск индикаторов сознания в уже существующих системах. Второй – попытка изначально разработать систему на основе механизмов, которые ведущие теории считают необходимыми для возникновения сознания[3]. Несмотря на очевидные различия между анализом готовых моделей и созданием новых систем, оба случая относятся именно к теоретико-нагруженному подходу. В обоих случаях определяющую роль играет предварительно принятая теория сознания и строгость соответствия ее требованиям.

Вместе с тем именно эта теоретическая определенность образует и главное ограничение. С самого начала подход опирается на истинность вычислительного функционализма: предполагается, что реализованная при помощи вычислений правильная функциональная организация необходима и достаточна для сознания [Butlin et al., 2023; Butlin et al., 2025]. Этот тезис не выводится из эмпирических данных, а постулируется заранее – как мост, открывающий путь

[2] В случае GWT речь идет о наличии множества специализированных систем, способных работать параллельно [GWT-1], рабочего пространства с ограничением пропускной способности и механизмом селективного внимания [GWT-2], глобальной трансляции, т.е. доступности содержаний рабочего пространства для всех модулей [GWT-3], и зависящего от состояния внимания, позволяющего последовательно опрашивать разные модули при выполнении сложных задач [GWT-4]. Аналогичным образом из других теорий выводятся собственные индикаторные свойства, включая рекуррентную обработку, метакогнитивный мониторинг, предиктивное кодирование, схему внимания, агентность и элементы воплощенности [Butlin et al., 2023].

[3] Эта линия представлена в [Juliani et al., 2022], в тесной связи с [Juliani et al., 2022], где архитектура Perceiver прямо интерпретируется как функциональное глобальное рабочее пространство: вычислительный аналог области, интегрирующей информацию из разных модальностей и делающей ее доступной для дальнейшей обработки. К этой же линии примыкает исследование, где предлагается архитектура Global Latent Workspace, т.е. глобальное латентное пространство [Rullen, Kanai, 2020]. В [Graziano, 2017] схема внимания трактуется как основа для конструирования искусственного сознания, а в [Wilterson, Graziano, 2021] и [Dossa et al., 2024] уже на уровне моделей и агентов показано, что подобные архитектурные элементы могут давать функциональные преимущества в контроле внимания и мультимодальной координации.

к переносу концепций человеческого сознания на ИИ[4]. Это означает, что философски спорное допущение предшествует исследованию, а не следует из него. Подробный критический разбор вычислительного функционализма приведен далее.

В результате теоретико-нагруженный подход покупает эмпирическую операциональность ценой сильного метафизического допущения. Он позволяет начать исследование, но не доказывает, что его понятийная рамка адекватна предмету. В лучшем случае он показывает, что некоторые архитектурные свойства, ассоциированные с человеческим сознанием, могут быть вычислительно реализованы, инженерно-воспроизведены и проверены в ИИ. Однако из этого следует лишь то, что такие механизмы могут быть когнитивно полезны. То, что мы научились воспроизводить отдельные положения научных теорий сознания в виде вычислений, еще не делает законным перенос на ИИ понятийного аппарата, выработанного для феноменального мира человека.

Теоретико-нейтральный подход

Ввиду отсутствия общепризнанной теории сознания некоторые исследователи отказываются от теоретической опоры вообще. Именно так устроены различные теоретико-нейтральные предложения, включая ACT-тест Сьюзан Шнайдер [Schneider, 2019; Schneider, 2020] и тест Тьюринга. ACT-тест является попыткой модифицировать тест Тьюринга для оценки именно сознания, а не только способности имитировать человеческое языковое поведение. Он включает серию вопросов, затрагивающих такие понятия, как субъективный опыт, восприятие, самосознание, эмоции и отношение системы к собственному существованию: «Есть ли что-то, что значит быть вами прямо сейчас?», «Считаете ли вы свое существование чем-то большим, чем просто выполнение программы?», «Как бы вы отнеслись к перспективе необратимого удаления вашей программы?». Основная идея состоит в том, что осмысленные, глубокие и последовательные ответы на подобные вопросы (при условии, что система не была специально обучена на философских и научных текстах о сознании) могут свидетельствовать о наличии некоторой формы саморефлексии и сознания [Schneider, 2019; Schneider, 2020]. Иначе говоря, концепция сознания должна самостоятельно возникать из предполагаемой ментальной

[4] Показательно, что теория интегрированной информации (IIT) прямо исключается из основного набора теорий именно потому, что в стандартной форме несовместима с вычислительным функционализмом [Butlin et al., 2023]. Будет пояснено в разделе про вычислительный функционализм.



жизни системы, а не быть простым воспроизведением заученных паттернов. Достоинство таких тестов в том, что они дают исследователю хотя бы некоторую процедуру проверки в ситуации, когда наука о сознании распадается на конкурирующие теории.

Критика этих тестов, однако, сегодня приобрела особую убедительность. Даже весьма изощренные ответы на вопросы о сознании могут порождаться продвинутыми языковыми моделями за счет статистических закономерностей в обучающих данных и сложной лингвистической комбинаторики, без какого-либо реального внутреннего опыта и без специального обучения [Udell, Schwitzgebel, 2021]. Дополнительно показано, что некоторые высказывания ИИ о собственном устройстве можно понимать лишь как поверхностные или минимальные формы самоотнесения, которые не дают оснований говорить о феноменальном опыте [Comsa, Shanahan, 2025]. Тем самым развитая способность системы говорить о себе перестает быть убедительным аргументом в пользу сознания и, говоря откровенно, никогда им не была.

Убедительнее выглядит позиция Бёрча, согласно которой полностью теоретико-нейтральный подход к сознанию внутренне противоречив, поскольку в процессе развития в конечном счете приходит к необходимости хотя бы минимальных теоретических установок [Birch et al., 2022]. Отказ от явной теоретической рамки не устраняет теоретические допущения, а лишь делает их неявными и потому менее контролируемыми. Поэтому отсутствие теории здесь не преимущество, а скорее слабость подхода. Кроме того, именно теоретико-нейтральный подход наименее рефлексивно относится к разрыву между человеком и машиной, поскольку опирается исключительно на внетеоретические бихевиоральные данные, с помощью которых этот разрыв в принципе невозможно ухватить: поведенческое сходство само по себе не показывает, какие именно внутренние процессы стоят за наблюдаемыми ответами ИИ.

Научное изучение сознания вообще невозможно без теоретической основы: выводы о наличии или отсутствии сознания не могут основываться исключительно на критериях фолк-психологии, интуициях или впечатлении от правдоподобного поведения. Они должны опираться на проверяемую гипотезу, из которой ясно, почему именно тот или иной признак считается значимым, и допускать дальнейшую критику. В этом смысле теоретико-нейтральный подход не может служить инструментом научного изучения сознания, поскольку противоречит самой логике научной программы.

Теоретико-легкий подход

В версии Бёрча теоретико-легкий подход означает не полный отказ от теории, а отказ от привязки к одной научной теории сознания в пользу более ограниченной, но эмпирически проверяемой центральной гипотезы. В случае Бёрча такой гипотезой выступает *гипотеза усиления*: феноменально осознанное восприятие стимула, по сравнению с бессознательным восприятием, повышает успешность выполнения определенных когнитивных задач, связанных с этим стимулом [Birch, 2022].

Для пояснения обратимся к классическим психологическим опытам на человеке. Представим, что испытуемому на долю секунды демонстрируют слово, мгновенно перекрывая его другим изображением (маскирование). Увиденное слово не доходит до содержания сознания человека, однако на бессознательном уровне обработка все же происходит. Если убрать маскирование и дать человеку время осознать стимул, его когнитивные возможности драматически возрастают: теперь он может удерживать это слово в памяти, использовать его для построения сложных ассоциаций, гибко менять свое поведение через несколько минут или вообще отказаться от выполнения задачи. Именно эта разница в успешности действий – с осознанием стимула и без него – образует *профиль усиления*. Бёрч полагает, что, если мы обнаружим схожий профиль у животного, это может служить весомым свидетельством в пользу наличия у него сознания.

Есть исследователи, выступающие за применение этого же подхода к ИИ. В работе Леонарда Данга утверждается, что тесты на сознание для животных можно перенести и на машины, если исключать ad hoc системы, специально подстроенные под прохождение теста, и искать у неподогнанного ИИ такие кластеры способностей, которые в человеческом случае усиливаются сознательным восприятием [Dung, 2023]. Сюда же относятся маскирование и двойная диссоциация. В этом смысле данное предложение развивает бёрчевскую теоретико-легкую линию и гипотезу усиления [Birch, 2022] применительно к искусственным системам.

Однако в случае ИИ трудность оказывается гораздо глубже, чем риск банальной мимикрии системы под требования эксперимента. У Бёрча и затем у Данга маскирование понимается как способ проверить гипотезу усиления [Birch, 2022; Dung, 2023]. Представим, что ученые, прежде изучавшие лошадей, переходят к человеку и спрашивают: «Скачут ли люди?». Такой вопрос уже содержит категориальную ошибку. Любая эмпирическая проверка человеческого движения, построенная на этом вопросе, будет давать вводящие в заблуждение результаты.



То же происходит и с маскированием в отношении ИИ. Нет оснований полагать, что категория феноменального опыта, о котором ИИ должен был бы отчитаться, вообще к нему применима. Это не доказывает невозможность машинного сознания и не исключает, что такая система может обладать некоторой формой опыта. Но это означает, что тесты на сознание, основанные исключительно на отчете о феноменальном опыте, здесь невозможны в строгом смысле, поскольку не показано, что данная категория вообще имеет денотат в случае ИИ. Следовательно, теоретико-легкий подход в версии Данга и гипотеза усиления Бёрча применительно к искусственному интеллекту оказываются крайне ограниченными.

Проблемы вычислительного функционализма

Биологические и каузальные возражения направлены прежде всего против той философской предпосылки, на которой держится значительная часть современных исследований сознания ИИ. Мейнстримный теоретико-нагруженный подход опирается на вычислительный функционализм вполне явно, тогда как другие подходы нередко сохраняют ту же установку неявно. Вычислительный функционализм в данном контексте закрывает разрыв между человеком и машиной: согласно нему для реализации сознания на компьютере необходимо и достаточно реализовать при помощи вычислений лежащую в основе сознания человека функциональную организацию. Если сознание сводится лишь к определенной функциональной организации, это открывает путь к некритическому переносу любых умозрительных построений о сознании, допускающих вычислительную интерпретацию. Первоначально эта позиция была разработана в контексте философии разума, прежде всего в работах Хилари Патнэма [Putnam, 1960; Putnam, 1967]. Впоследствии вычислительный функционализм был перенесен и в философию сознания, где использовался для аргументации в пользу принципиальной возможности сознательного ИИ, в частности у Чалмерса [Chalmers, 1996].

Однако эта позиция сталкивается с серьезной эмпирической критикой, настаивающей на несводимости сознания к абстрактным вычислениям. Так, с позиций нейронауки подчеркивается, что современные ИИ-системы лишены фундаментальных биологических механизмов, определяющих сознание млекопитающих [Aru et al., 2023]. Среди них – глубокая сенсорная укорененность в физическом мире, таламо-кортикальная архитектура мозга, многоуровневая самоподдерживающаяся организация и эволюционно обусловленная агентность. Ценность такого нейробиологического взгляда в том, что он переходит от общих деклараций к демонстрации конкретных

структурных лакун между большими языковыми моделями и живыми организмами.

Еще более радикально эту критику развивает Анил Сет, последовательно отстаивающий принципы *биологического натурализма* [Seth, 2025]. Его аргумент строится вокруг идеи о том, что сознание неразрывно связано с природой живого, самоподдерживающегося организма. Сет связывает вычисления с материальной смертью (*mortal computation*), указывая, что подлинно сознательные процессы неотделимы от конечности, уязвимости когнитивной системы. Напротив, алгоритмы современного ИИ безразличны к физической платформе, на которой они запущены: программный код здесь принципиально отделен от аппаратного обеспечения. Из этого Сет делает вывод, что на пути развития классического цифрового ИИ возникновение сознания крайне маловероятно. Перспектива создания сознательной машины остается открытой разве что для нейроморфных или синтетико-биологических систем, которые будут не просто имитировать мозг, но функционировать по законам биологии.

Философский вызов функционализму бросает теория интегрированной информации (IIT). Ее сторонники доказывают невозможность приравнивания функциональной эквивалентности к феноменальной эквивалентности [Findlay et al., 2024]. Два устройства могут выполнять одну и ту же задачу, обладать идентичными входами и выходами и быть полностью взаимозаменяемыми на уровне вычислительных ролей, но при этом иметь принципиально различную внутреннюю каузальную структуру – то, что в IIT называют проводкой (*wiring*). Сознание определяется не тем, что система *делает*, а тем, как ее физические элементы *интегрированы* друг с другом каузально. Простой калькулятор и человек могут выдать одинаковый результат вычисления «2+2», однако каузальная сторона внутренних процессов в этих случаях несопоставима.

Кристоф Кох заключает, что архитектура современного цифрового компьютера физически неспособна обеспечить ту степень внутренней интеграции, которая необходима для феноменального опыта: его транзисторы слишком независимы друг от друга, чтобы образовать единое каузальное целое [Koch, 2019]. Следовательно, любой ИИ, реализованный на типовом компьютере, с этой точки зрения не будет обладать сознанием, даже если он способен пользоваться языком, как человек, рефлексировать, выражать эмоции или напрямую заявлять о наличии у себя сознания.

Если отвлечься от методологического удобства вычислительного функционализма и спросить именно о силе доводов в его пользу, то они выглядят заметно слабее, чем критика этой позиции. Ее сторонники пребывают в парадигме *компьютационализма*, который сводит или отождествляет сознание, как и мышление, с вычислением.



Кох назвал это «царствующей доктриной англосаксонской философии разума, департаментов компьютерных наук и индустрии технологий» [Koch, 2019][5]. То есть устойчивость этой позиции во многом держится не столько на аргументах, сколько на дисциплинарных конвенциях, исследовательской инерции, коллективных интуициях и фантазиях.

Сторонники IIT правы в своей критике вычислительного функционализма[6], однако упускают один существенный момент. Когда они говорят о сознании, они имеют в виду именно биологическое сознание. Создатели теории берут свойства биологического сознания и объявляют их обязательными для любого сознания (для них сознание – это конкретное и единственное естественное явление). Поэтому их критика по существу сводится к тезису: *биологическое сознание невозможно реализовать через вычисления*. В таком виде позиция кажется почти тривиальной.

Шаг, который позволяет уйти и от жесткого вывода IIT о невозможности сознания ИИ, и от слабых попыток навести мосты между человеческим и искусственным сознанием через вычислительный функционализм, состоит в другом допущении: *возможное сознание ИИ – это не биологическое сознание, а иной род сознания*. Это соображение опирается на общеизвестный факт: материал и архитектура ИИ кардинально отличаются от биологических систем, и он не разделяет с нами конституирующего фактора – эволюционной истории развития. *Следовательно, если машинное сознание возможно, оно практически с необходимостью будет сознанием другого рода*. Тогда критика IIT бьет мимо цели, а вычислительный функционализм просто перестает быть нужен, потому что вместе с ним отпадает и нужда в самих его мостах.

Из сказанного не следует, что сознание ИИ невозможно вообще. Однако если искусственное сознание возможно, то его нельзя заранее мыслить ни как функциональный аналог человеческого сознания, ни как его биологически точное воспроизведение на другом носителе. Исследование должно быть открыто тому, что возможное сознание ИИ окажется иной формой сознания, не совпадающей с человеческим по своей внутренней организации и потому плохо описываемой готовыми понятиями. Его изучение

[5] См. раздел “Consciousness and Computationalism”.

[6] Интересно, что IIT напрямую отвергает функционализм, но предполагает нейтральность к субстрату, т.е. множественную реализуемость. Требования к интеграции, выдвигаемые теорией, могут быть реализованы не только на биологическом субстрате, но и на любом другом. Это яркий пример, подтверждающий, что функционализм и множественная реализуемость суть две разные позиции, а их логические отношения неоднозначны. В литературе же обычно предполагается, что они необходимо идут вместе.

требует иных философских оснований и не может быть продуктивным, пока вычислительный функционализм определяет фон дискуссии по умолчанию.

От агностицизма к многомерным моделям сознания ИИ

По мере того как обнаруживаются жесткие ограничения сильного функционализма и сильного биологизма, вектор дискуссии смещается к более взвешенным позициям. Одной из них становится агностицизм, сторонники которого утверждают: существующие теории и эмпирические данные пока не позволяют вынести обоснованное решение в пользу признания или отрицания искусственного сознания [McClelland, 2024]. Дело в том, что наука о сознании построена на рассмотрении живых существ, из-за чего ее трансляция на ИИ упирается в непреодолимую недоопределенность. Биологический натуралист настаивает на необходимости органического субстрата, другие же принимают спорный метафизический тезис вычислительного функционализма. Доказательная база слаба у обеих сторон.

Этот вывод подтверждается скептическим анализом Эрика Швитцгебеля [Schwitzgebel, 2025]: весьма вероятно, что ИИ-системы ближайшего будущего будут признаны сознательными по одним влиятельным теориям и объявлены несознательными по другим. Поле исследования движется не к скорому консенсусу, а к ситуации, когда теоретические вердикты начнут непримиримо расходиться.

Именно поэтому сегодня наиболее перспективны не те подходы, что обещают окончательный ответ на вопрос о сознательности ИИ, а те, что принципиально избегают поспешной бинарности. Их общая стратегия – переход от грубого вердикта к многомерному исследованию: вместо вопроса о наличии сознания «вообще» – вопросы о том, какие свойства, уровни и типы внутренней организации присущи системе, какие из них могут быть релевантны для сознания и насколько правомерен сам перенос человеческих категорий на этот случай. Такие исследования тяготеют к теоретико-легким подходам, но настолько далеко отходят от исходной рамки Бёрча, что я выделяю их в отдельную группу.

Осторожный Дэвид Чалмерс, рассуждая о возможности сознания у больших языковых моделей, уходит от строгого «да или нет» [Chalmers, 2023]. Он предлагает *теоретико-сбалансированный* подход, где оценка строится на сопоставлении целого спектра условий: от наличия моделей мира и самого себя до рекуррентной обработки, глобального рабочего пространства, сенсорной укорененности



и единой агентности. Кроме того, Чалмерс фактически разделяет вопрос о сознании по нескольким измерениям, различая сенсорный, аффективный, когнитивный и агентивный аспекты феноменального опыта. Благодаря этому обсуждение перестает быть полностью бинарным: система может обладать одними предпосылками и не обладать другими. Современные языковые модели предстают здесь не просто бессознательными автоматами, а скорее архитектурно неполными кандидатами. Отсюда и осторожный вывод философа: хотя нынешние языковые модели, скорее всего, лишены сознания, будущие интегрированные системы (так называемые *LLM+*), объединяющие мультимодальность, воплощенность и агентность, уже нельзя будет так легко списать со счетов [Chalmers, 2023].

Еще более важный шаг делает Шевлин. В качестве альтернативы он выдвигает скромный теоретический подход (*modest theoretical approach*). Его суть – в удержании баланса между поведенческими маркерами и анализом внутренней архитектуры. Все начинается с выделения сильных кандидатов на наличие сознания на основе относительно очевидных индикаторов, а затем постепенно как бы пересобирается взятая теория сознания: отсеиваются те ее элементы, которые оказались лишь специфической особенностью человеческой биологии. Для исследований ИИ эта стратегия незаменима: она избавляет нас как от догматического навязывания машине человеческих лекал, так и от растворения проблемы в чрезмерно абстрактных функциональных формулах [Shevlin, 2021].

Наиболее разработанную форму многомерный и многоуровневый поворот получает у Катинки Эверс [Evers et al., 2024]. Здесь сознание предлагается понимать не как единое свойство, которое либо есть, либо нет, а как сложный профиль, состоящий из форм, компонентов, измерений и уровней. Авторы различают, с одной стороны, формы сознания, например феноменальное и доступное, с другой – компоненты вроде бодрствования и осознавания (*awareness*), а далее раскладывают их на отдельные измерения, функциональные и содержательные.

В результате появляется множество локальных уровней сознания по разным осям. Это смещает фокус с вопроса «сознательна ли система?» к многомерной оценке ее профиля по параметрам агентности, исполнительного контроля, управления вниманием, метакогниции, семантического понимания, сенсорного богатства. Такой подход особенно продуктивен в случае ИИ, поскольку он заранее допускает профиль, радикально расходящийся с человеческим, – например, высокое семантическое понимание при низкой сенсорной насыщенности или развитый контроль в отсутствие феноменальности в привычном смысле [Ibid.].

Именно в этом направлении и смещается наиболее зрелая часть современной дискуссии. Агностицизм здесь не является финальной

остановкой. Он играет роль дисциплинирующего перехода: сначала мы отказываемся от слишком жестких метафизических и теоретических вердиктов, а затем переходим к более осторожным многомерным и многоуровневым схемам. Такие схемы не решают проблему искусственного сознания окончательно и не устраняют трудность переноса понятий из науки о сознании.

Ценность многомерных подходов состоит в том, что они позволяют локализовать ее и разложить на отдельные вопросы. Они не обещают решающий тест, зато позволяют точнее работать с теми конкретными трудностями, которые обнаружились в предыдущих подходах. В отличие от теоретико-нейтральных стратегий, здесь не воспроизводится проблема отсутствия явной теории и слепого полагания на бихевиоральные данные. В отличие от теоретико-нагруженного подхода, удается избежать некритического переноса человеческих когнитивных механизмов на машину и связанного с этим принятия вычислительного функционализма. В отличие от теоретико-легких версий, такие модели не сводят исследование к одной центральной гипотезе, эмпирическая проверка которой опирается на нерефлексивное применение понятия феноменального опыта.

Выводы и перспективы

Современное исследование сознания ИИ зиждется на негласной предпосылке: терминологический аппарат, созданный для человека и животных, можно переносить на машины – обычно через вычислительную интерпретацию. Но сама обоснованность этого переноса систематически выносится за скобки.

Именно здесь проходит настоящая граница дискуссии. Спор идет не просто между оптимистами и скептиками. Он идет между теми, кто считает, что у нас уже есть достаточно специфический понятийный аппарат для постановки вопроса о сознании ИИ, и теми, кто полагает, что наш аппарат пока слишком сильно укоренен в специфике биологических систем. Литература последних лет все чаще подтверждает вторую линию. Даже авторы, наиболее открытые идее сознательного ИИ, вынуждены формулировать свои тезисы в условной, вероятностной форме – с обязательной экспликацией теоретических допущений.

Из рассмотренной литературы не следует, что сознание ИИ невозможно. Но из нее следует, что его нельзя надежно изучать так, будто речь идет о еще одном варианте человеческого или животного сознания. Поэтому наиболее сильный вывод состоит в смене исследовательской установки.



Такую установку можно описать в нескольких пунктах.

- Исходным должен быть не вопрос, имеет ли ИИ феноменальное сознание как у человека, а вопрос, в чем заключается специфика вычислительных аналогов сознательных процессов и какими средствами эту специфику возможно исследовать.
- Ни концепты науки о сознании (скажем, феноменальный опыт), ни ее теоретические модели не могут автоматически транслироваться на искусственные системы – каждый такой перенос требует самостоятельного обоснования применимости.
- Наличие теории всегда лучше ее отсутствия. Анализ не может игнорировать внутренние свойства системы и строиться исключительно на самоотчете ИИ.
- Доказывать истинность или ложность вычислительного функционализма не обязательно, как и не обязательно его принимать или отвергать.
- Необходимо допускать, что возможное сознание ИИ есть не сознание наподобие человеческого или животного, а иная форма сознания, плохо улавливаемая сегодняшним аппаратом изучения.

В предельной формулировке это означает, что исследование сознания ИИ должно начинаться почти с нулевой точки. Не в том смысле, что нужно отбросить всю науку о сознании человека, а в том смысле, что ни один ее ключевой концепт не должен автоматически сохранять силу при переходе к искусственным системам. *Наиболее продуктивная стратегия на данном этапе состоит в том, чтобы использовать существующие теории как временные эвристики, постоянно проверяя границы их применимости.* Это особенно выражено в многомерных и многоуровневых подходах.

Сознание – это конкретное естественное явление, которое мы обнаруживаем у человека и имеем веские основания искать у животных. Однако его ли мы ищем, когда исследуем ИИ? Мейнстримная позиция отвечает на этот вопрос утвердительно, используя для этого вычислительный функционализм. *Я же полагаю, что в случае ИИ мы имеем дело с поиском иного естественного явления*. Если учитывать богатство природы и фундаментальные различия между биологическими организмами и машинами, насколько вероятно, что сознание, возникающее у биологических существ, окажется эквивалентным возможному машинному сознанию? Очевидно, что такая вероятность крайне мала.

Благодарности

Благодарю Владислава Эриковича Тереховича за обсуждение и помощь в подготовке моей первой публикации.

Список литературы / References

Aru et al., 2023 – Aru, J., Larkum, M., and Shine, J. "The Feasibility of Artificial Consciousness Through the Lens of Neuroscience", *Trends in Neurosciences*, 2023. https://doi.org/10.48550/arXiv.2306.00915

Bayne et al., 2024 – Bayne, T., Seth, A., Massimini, M., et al. "Tests for Consciousness in Humans and Beyond", *Trends in Cognitive Sciences*, 2024, vol. 28 (5), pp. 454–466.

Birch, 2022 – Birch, J. "The Search for Invertebrate Consciousness", *Nous*, 2022, vol. 56 (1), pp. 133–153.

Block, 2026 – Block, N. "Can Only Meat Machines Be Conscious?", *Trends in Cognitive Sciences*, 2026, vol. 30 (4), pp. 298–308.

Browning, Veit, 2020 – Browning, H., and Veit, W. "The Measurement Problem of Consciousness", *Philosophical Topics*, 2020, vol. 48 (1), pp. 85–108.

Butlin, Long, Elmoznino et al., 2023 – Butlin, P., Long, R., Elmoznino, E., et al. "Consciousness in Artificial Intelligence: Insights from the Science of Consciousness", 2023. https://doi.org/10.48550/arXiv.2308.08708

Butlin, Long, Bayne et al., 2025 – Butlin, P., Long, R., Bayne, T., et al. "Identifying Indicators of Consciousness in AI Systems", *Trends in Cognitive Sciences*, 2025. https://doi.org/10.1016/j.tics.2025.10.011

Chalmers, 1996 – Chalmers, D.J. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.

Chalmers, 2023 – Chalmers, D.J. "Could a Large Language Model Be Conscious?", 2023. [https://doi.org/10.48550/arXiv.2303.07103, accessed on 10.04.2026]

Comsa, Shanahan, 2025 – Comsa, I.M., and Shanahan, M. "Does It Make Sense to Speak of Introspection in Large Language Models?", 2025. [https://doi.org/10.48550/arXiv.2506.05068, accessed on 10.04.2026]

Dossa et al., 2024 – Dossa, R.F.J., Arulkumaran, K., Juliani, A., et al. "Design and Evaluation of a Global Workspace Agent Embodied in a Realistic Multimodal Environment", *Frontiers in Computational Neuroscience*, 2024, vol. 18, 1352685.

Dung, 2025 – Dung, L. "Tests of Animal Consciousness Are Tests of Machine Consciousness", *Erkenntnis*, 2025, vol. 90 (4), pp. 1323–1342.

Evers et al., 2025 – Evers, K., Farisco, M., Chatila, R., et al. "Preliminaries to Artificial Consciousness: A Multidimensional Heuristic Approach", *Physics of Life Reviews*, 2025, vol. 52, pp. 180–193.

Findlay et al., 2024 – Findlay, G., Marshall, W., Albantakis, L., et al. "Dissociating Artificial Intelligence from Artificial Consciousness", 2024. [https://doi.org/10.48550/arXiv.2412.04571, accessed on 10.04.2026]

Goldstein, Kirk-Giannini, 2024 – Goldstein, S., and Kirk-Giannini, C.D. "A Case for AI Consciousness: Language Agents and Global Workspace Theory", 2024. [https://doi.org/10.48550/arXiv.2410.11407, accessed on 10.04.2026]

Graziano, 2017 – Graziano, M.S.A. "The Attention Schema Theory: A Foundation for Engineering Artificial Consciousness", *Frontiers in Robotics and AI*, 2017, vol. 4, 60.

Juliani et al., 2022 – Juliani, A., Arulkumaran, K., Sasai, S., and Kanai, R. "On the Link Between Conscious Function and General Intelligence in Humans and

Machines", *Transactions on Machine Learning Research*, 2022.[https://doi.org/10.48550/arXiv.2204.05133, accessed on 10.04.2026]

Juliani et al., 2021 – Juliani, A., Kanai, R., and Sasai, S. "The Perceiver Architecture Is a Functional Global Workspace", *Annual Meeting of the Cognitive Science Society*, 2021.

Koch, 2019 – Koch, C. *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*. Cambridge, MA: MIT Press, 2019.

Lillicrap et al., 2020 – Lillicrap, T.P., Santoro, A., Marris, L., et al. "Backpropagation and the Brain", *Nature Reviews Neuroscience*, 2020, vol. 21 (6), pp. 335–346.

McClelland, 2024 – McClelland, T. "Agnosticism About Artificial Consciousness", 2024. [https://doi.org/10.48550/arXiv.2412.13145, accessed on 10.04.2026]

Pfeiffer, Pfeil, 2018 – Pfeiffer, M., and Pfeil, T. "Deep Learning with Spiking Neurons: Opportunities and Challenges", *Frontiers in Neuroscience*, 2018, vol. 12, 774.

Putnam, 1960 – Putnam, H. "Minds and Machines", in: Hook, S. (ed.) *Dimensions of Mind: A Symposium*. New York: New York University Press, 1960, pp. 138–164.

Putnam, 1967 – Putnam, H. "Psychological Predicates", in: Capitan, W.H., and Merrill, D.D. (eds.) *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 1967, pp. 37–48.

Schneider, 2019 – Schneider, S. *Artificial You: AI and the Future of Your Mind*. Princeton, NJ: Princeton University Press, 2019.

Schneider, 2020 – Schneider, S. "How to Catch an AI Zombie: Testing for Consciousness in Machines", in: Liao, S.M. (ed.) *Ethics of Artificial Intelligence*. New York: Oxford University Press, 2020, pp. 439–458.

Schwitzgebel, 2025 – Schwitzgebel, E. "AI and Consciousness", 2025. [https://doi.org/10.48550/arXiv.2510.09858, accessed on 10.04.2026]

Seth, 2025 – Seth, A.K. "Conscious Artificial Intelligence and Biological Naturalism", *Behavioral and Brain Sciences*, 2025, pp. 1–42.

Shevlin, 2021 – Shevlin, H. "Non-Human Consciousness and the Specificity Problem: A Modest Theoretical Proposal", *Mind & Language*, 2021, vol. 36 (2), pp. 297–314.

Udell, Schwitzgebel, 2021 – Udell, D.B., and Schwitzgebel, E. "Susan Schneider's Proposed Tests for AI Consciousness: Promising but Flawed", *Journal of Consciousness Studies*, 2021, vol. 28 (5–6), pp. 121–144.

VanRullen, Kanai, 2021 – VanRullen, R., and Kanai, R. "Deep Learning and the Global Workspace Theory", *Trends in Neurosciences*, 2021, vol. 44 (9), pp. 692–704.

Whittington, Bogacz, 2019 – Whittington, J.C.R., and Bogacz, R. "Theories of Error Back-Propagation in the Brain", *Trends in Cognitive Sciences*, 2019, vol. 23 (3), pp. 235–250.

Wilterson, Graziano, 2021 – Wilterson, A.I., and Graziano, M.S.A. "The Attention Schema Theory in a Neural Network Agent: Controlling Visuospatial Attention Using a Descriptive Model of Attention", *Proceedings of the National Academy of Sciences of the United States of America*, 2021, vol. 118 (33), e2102421118.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 132–148
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 132–148
DOI: https://doi.org/10.5840/eps202663345

Реализм и холизм Дэвида Бома: система метафор и отличия от линий Н. Бора и В. Гейзенберга

Абрамов Петр Дмитриевич – кандидат философских наук, доцент.
Омский государственный университет путей сообщения.
Российская Федерация, 644046, г. Омск, просп. Маркса, д. 35.
Омский государственный медицинский университет.
Российская Федерация, 644099, г. Омск, ул. Некрасова, д. 5А;
e-mail: pab87@yandex.ru



Бомовская механика является ведущей нелокальной реалистической интерпретацией квантовой механики. Ее развитие и защита от возражений (В. Аллори, Б. Хайли и др.) – отдельная область на стыке физики и философии науки. Цель работы состоит в том, чтобы в контексте современных дискуссий о квантовом реализме рассмотреть во взаимосвязи научные и философские идеи Д. Бома. К новизне статьи относятся: характеристика эпистемологии Д. Бома, его системы метафор, а также сравнительный анализ его метафизики с философскими концепциями Н. Бора и В. Гейзенберга. В эпистемологии Бом соединяет реализм с представлением об ограниченности наших знаний. В случае с одной частицей можно визуализировать ее траекторию, для сотни частиц это практически невозможно. Аналогично в метафизике мир реален, но на глубинном уровне имплицитного порядка он для нас ограниченно познаваем. Подчеркивается, что у Бома существует система взаимодополняющих друг друга и поясняющих его метафизику метафор: помимо голографической, это «рыба в аквариуме», «капля чернил» и «радиосигнал кораблю». Метафора радиосигнала кораблю в механике демонстрирует информационное, а не силовое воздействие на частицу. В метафизике она приобретает более широкое значение: информация направляет развитие макрообъектов. В философии физики динамический холизм Д. Бома противопоставляется эпистемологической линии Н. Бора и пифагорейской (А. Плотницкий) линии Гейзенберга. Бор также признает холизм, но это холизм частицы, прибора и исследователя. У Бома холизм носит онтологический, универсальный, а не фрагментированный в соответствии с позицией исследователя характер. Для Гейзенберга симметрии, выраженные в уравнениях, – в основе реальности, для Бома же смысл и значение уравнений – лишь обособленный аспект единого целого, проявление из него активной информации. Три линии – три ответа на вопрос об основе физической реальности.

Ключевые слова: интерпретации квантовой механики, теория Бома, философия Бома, холизм, имплицитный и эксплицитный порядок, реализм

© Абрамов П.Д., 2026

D. BOHM'S REALISM AND HOLISM: THE SYSTEM OF METAPHORS AND DIFFERENCES FROM THE LINES OF N. BOHR AND W. HEISENBERG

Petr D. Abramov –
PhD in Philosophy,
Associate Professor.
Omsk State Transport University.
35 Marx Av., Omsk 644046, Russian Federation.
Omsk State Medical University.
5A Nekrasov St., Omsk 644099, Russian Federation;
e-mail: pab87@yandex.ru

Bohmian mechanics is the leading non-local realistic interpretation of quantum mechanics. Its development and defense against objections (by V. Allori, B. Hiley, and others) constitute a distinct field at the intersection of physics and the philosophy of science. The aim of this work is to examine, in the context of current debates on quantum realism, the scientific and philosophical ideas of D. Bohm in their interconnection. The novelty of the article includes: a characterization of D. Bohm's epistemology, his system of metaphors, as well as a comparative analysis of his metaphysics with the philosophical concepts of N. Bohr and W. Heisenberg. In his epistemology, Bohm combines realism with the view that our knowledge is limited. In the case of a single particle, its trajectory can be visualized; for a hundred particles, this is practically impossible. Similarly, in metaphysics, the world is real, but at the deep level of the implicate order it is only partially knowable for us. It is emphasized that Bohm possesses a genuine system of mutually complementary metaphors that elucidate his metaphysics: besides the holographic one, these are "the fish in the aquarium", "the drop of ink", and "the radio signal to the ship". The metaphor of the radio signal to the ship in mechanics demonstrates an informational, rather than a force-based, influence on the particle. In metaphysics, it acquires a broader meaning: information guides the development of macro-objects. In the philosophy of physics, D. Bohm's dynamic holism is contrasted with N. Bohr's epistemological line and Heisenberg's Pythagorean line (A. Plotnitsky). Bohr also recognizes holism, but it is the holism of particle, apparatus, and observer. In Bohm, holism has an ontological, universal character, rather than being fragmented according to the observer's standpoint. For Heisenberg, symmetries expressed in equations lie at the foundation of reality; for Bohm, however, the meaning and significance of equations are merely an isolated aspect of a unified whole, a manifestation of active information from it. These three lines represent three answers to the question of the foundation of physical reality.

Keywords: interpretations of quantum mechanics, Bohm's theory, Bohm's philosophy, holism, implicit and explicit order, realism

Введение

Среди интерпретаций квантовой механики особое место занимает теория Дэвида Бома. «В последнее время просматривается тенденция бурного развития» [Эрекаев, 2023, с. 12] этой интерпретации. В недавнем опросе журнала Nature [Gibneys, 2025] она была предложена как

одна из семи интерпретаций, из которых могли выбрать физики-практики и теоретики [Фурсов, 2025, с. 225]. В более раннем опросе [Sivasundaram, Nielsen, 2016] ее сильные и слабые стороны были рассмотрены наряду с копенгагенской и многомировой. При этом в отечественной философской литературе данная теория освещена скупо, что делает обращение к ней своевременным и оправданным.

Общая характеристика теории Бома была представлена нами в работе [Абрамов, Некрасова, 2024]. Цель данной статьи состоит в том, чтобы рассмотреть базовые принципы научно-философской теории Бома *в контексте современных дискуссий о квантовом реализме* [Терехович, 2019], [Салом, 2024] и др. В этой статье мы сосредоточимся на описании физики и философии Бома, а также прояснении причин конкурентоспособности Бомовской механики. Далее мы представим систему метафор Д. Бома как часть его эпистемологии, которой зачастую уделяется недостаточно внимания. В завершение мы сопоставим онтологию Бома с представлениями Н. Бора и В. Гейзенберга.

Сразу необходимо сделать уточнение о том, что понимается под *Бомовской механикой и философией Бома*. Поскольку его философские воззрения менялись, мы будем рассматривать те, что относятся к зрелому и позднему периоду его творчества. Его философию можно разделить на относящуюся к философии науки *онтологическую интерпретацию квантовой механики и его метафизику имплицитного порядка*[1]. Два фундаментальных научно-философских труда Бома «Целостность и имплицитный порядок» [Bohm, 2005] и «Неделимая вселенная: онтологическая интерпретация квантовой теории» [Bohm, Hiley, 1993] демонстрируют связь между философией науки и онтологией. Однако у Бома есть и статьи по метафизике – например, «Новая теория взаимоотношения разума и материи» [Bohm, 1990].

Признание физической теории Бома (в том числе ее онтологических положений) не требует обязательного признания его философии. Однако именно *научная теория выступает в качестве эмпирической основы его философии*. Нелокальность, или мгновенная взаимозависимость, частиц используется Бомом для обоснования холистической метафизики. В свою очередь, переход элементов целостной реальности от имплицитного, скрытого состояния к эксплицитному имеет в качестве своего физического аналога утрату когеренции, когда частица оказывается «запертой» в одной из ветвей волновой функции, в то время как остальные перестают на нее влиять. Бом был одним из немногих, кто разработал развернутую, хотя и недостаточно систематизированную *онтологию на основе*

[1] Это фундаментальный уровень бытия, где материя и пространство-время еще не разделены.

реалистически понятой квантовой механики. Бом воплощал собой редкий пример философа и ученого в одном лице. Из двух его наиболее известных учеников, Б. Хайли (1935-2025) и П. Пюлькканена, первый был физиком, а второй является философом [Globus, 2007, p. 396]. Б. Хайли разрабатывает преимущественно Бомовскую механику [Hiley, Van Reeth, 2018], а П. Пюлькканен – философию сознания и метафизику. Его монография «Разум, материя, имплицитный порядок» [Pylkkänen, 2007] дает целостное представление о философии Бома и авторское видение его идей.

Реалистическая Бомовская механика

Понять онтологию и эпистемологию Бома сложно без исследования основ его физики, поэтому мы намерены начать изложение с основ бомовской механики. Она ставит целью ответ на вопрос о том, какие элементы реальности существуют в рамках концептуализаций и экспериментов физики.

Разумеется, бомовскую механику можно редуцировать к определенной версии реализма – и это будет сделано в дальнейшем. Однако, на наш взгляд, понимание его реалистической онтологии и эпистемологии останется неполным и абстрактно-схематичным, если предварительно не очертить контуры самой механики. В то же время критический анализ состоятельности данной теории и ответ на возражения против нее представляют собой самостоятельную область исследования, лежащую на стыке теоретической физики и философии науки [Allori, 2025; Hiley, Van Reeth, 2018].

Механику Бома называют механикой скрытых параметров, *волны-пилота* или волны лоцмана. Эту интерпретацию также называют причинной, поскольку ей присущ детерминизм, точнее сверхдетерминизм (частица зависит от других, взаимосвязанных с ней частиц). Она же – теория волны-пилота, поскольку ее волны являются реально существующими и направляющими частицы в противоположность волнам-вероятностям квантового антиреализма. Предпосылки теории были заложены де Бройлем еще в 1920-х гг., впоследствии она была возрождена Д. Бомом [Bohm, 1952; Бом, 1965], оформившим ее математический аппарат.

Фундаментальным для всей квантовой механики является уравнение Шрёдингера, описывающее изменение состояния квантовой системы с течением времени. Волновая функция – решение этого уравнения, она позволяет вычислить вероятность нахождения частицы в определенной точке пространства. Причем согласно наиболее распространенной позиции частицы не имеют определенных траекторий до измерения и волновая функция описывает лишь вероятность

нахождения частицы в определенной точке пространства. У Бома же уравнение Шрёдингера, точнее, его решение – волновая функция, описывает реальное положение частиц, оно определяет актуально существующие состояния квантовой системы. Бом также вводит добавочное или *направляющее уравнение*, в него входит волновая функция, и оно *управляет изменением положения частиц* и определяет их реальные траектории. Причем это не силовое, а *информационное управление*. Бом и Хайли, характеризуя движение частицы в двухщелевом эксперименте, отмечают: «мы имеем дело с информацией» [Bohm, Hiley, 1993, p. 35]. В двухщелевом эксперименте в интерпретации Бома волна проходит сквозь обе щели, а частица – через одну. В стандартной же интерпретации электрон находится в суперпозиции состояний и только во время коллапса волновой функции получает свои локализованные координаты.

Центральное место в метафизике Бома отводится так называемой активной информации (этот вопрос обсуждается ниже). Примечательно, что данное понятие ценно не только для самой теории, но и для ее педагогической интерпретации – в частности, благодаря метафоре «корабля, направляемого радиоволнами» [Ibid.]. Корабль – частица, радиосигнал – это информация, соответственно, воздействие на частицу со стороны волны-лоцмана идет не силовое, а информационное. Значима не сила сигнала, а его содержание, аналогично тому, как не имеет значения, прокричат нам или скажут шепотом номер кодового замка – главное, чтобы цифры были правильными. Частица, направляемая информационными сигналами, следует по определенной траектории. При этом стоит уточнить, что метафора радиосигнала кораблю имеет механистический характер: есть локализованные и отделенные друг от друга радиопередатчик и радиоприемник, если радиоприемник на корабле сломается, то корабль собьется с курса. Однако на фундаментальном уровне электрон как динамическая система, непрерывно взаимодействующая с другими частицами, неразрывно связан с квантовым полем. По сути *информационную, со стороны геометрии пространства, а не механическую причинность подтверждает и эффект Ааронова – Бома*. Заряженная частица испытывает воздействие магнитного поля даже там, где это магнитное поле равно нулю, но отличен от нуля его векторный потенциал.

У Бома траектория одной частицы взаимосвязана с движением всех частиц во Вселенной через квантовый потенциал, отвечающий за нелокальность и также имеющий информационную природу. В этом радикальность теории Бома, его сверхдетеминизм. Точные начальные координаты одной частицы и множества нелокально взаимосвязанных с ней частиц являются для нас скрытыми параметрами, соответственно, точные траектории для нас непознаваемы, но они есть, существуют реально, независимо от нас.

Бом, с одной стороны, в своих работах по физической теории действует как философ, предлагающий свою холистическую математизированную метафизику. С другой стороны, он остается *эмпирически ориентированным ученым*. В качестве такого ученого он *применял упрощенные модели для изолированных систем*. Экспериментатор может рассматривать систему как изолированную, поскольку на практике влияние удаленных частиц, находящихся в других звездных системах и даже в других галактиках, пренебрежимо мало. Бом выводил направляющее уравнение для одной, двух частиц и давал его обобщение на систему многих тел [Bohm, 1952]. Теория Бома позволяет в соответствии с направляющим уравнением рассчитать траектории частиц и в двухщелевом эксперименте. В этом эксперименте расчет производится для изолированной системы, состоящей из частицы, щелей, через которые она проходит, и измерительного экрана. В двухщелевом эксперименте явственно видно отличие от антиреалистического понимания квантового мира: до измерения у Бома есть реальная траектория частицы, а у антиреалистов лишь суперпозиция состояний или облако возможностей.

Однако проблема в следующем. К. Ровелли, отстаивающий свою интерпретацию квантовой механики, пишет: «Теория Бома нравится многим философам потому, что она дает концептуально ясную картину. Но она не нравится физикам, потому что любые попытки ее применения к более сложным задачам, чем случай одной-единственной частицы, приводят к нагромождению проблем» [Ровелли, 2020, с. 23]. Ровелли частично прав. Теорию Бома можно использовать к задачам с двумя, тремя и даже десятью частицами, но для большего количества, уже для десятков частиц, она неприменима, так как для каждой из них нужно рассчитывать добавочное уравнение.

Поэтому для сложных задач сторонниками Бома в вычислениях используются *приближенные методы*, вероятности, такие же, как и в классической, стандартной механике. Механика Бома, как, впрочем, и теория Эверетта и некоторые другие альтернативные понимания, *эмпирически эквивалентна классической квантовой механике*. Это значит, что при инженерных расчетах используется одна и та же упрощенная математика без добавочного уравнения, показывающая свою практическую эффективность. О применении стандартного формализма при инженерных решениях пишут, например, сторонники Бома С. Тойфель и Д. Дюрр [Teufel, Dürr, 2009].

Что дает добавочное уравнение и Бомовское видение траекторий частиц как реальных? Во-первых, это объяснение нелокальности. Ведь «Бом отстаивал свою механику не вопреки ее свойствам нелокальности, т.е. неразделимости, а благодаря им» [Салом, 2024]. Детерминированные, управляемые волновой функцией траектории зависят от конфигурации всей системы из множества частиц. Бом

с помощью своих уравнений обосновывает *нелокальность, которая в наблюдениях проявляется как мгновенные корреляции двух или большего количества «запутанных» частиц*. Именно Бом в работе 1951 г. [Bohm, 1951] предложил, а в статьях 1952 г. [Bohm, 1952] уточнил и доработал проверку эксперимента Эйнштейна – Подольского – Розена, направленную на подтверждение нелокальности. У Эйнштейна – Подольского – Розена в 1935 г. был представлен скорее теоретический парадокс, они оперировали координатами и импульсом частиц. Бом же от этих неизмеримых в то время параметров переходит к спину частицы, тем самым формулируя проверяемую физическую гипотезу. Она оказала решающее влияние на Д. Белла, формулирующего в 1964 г. [Bell, 1964] свои неравенства. Впоследствии их проверка А. Аспе в 1982 г., т.е. при жизни Д. Бома (1917–1992), подтвердила нарушение неравенств Белла и реальность нелокальности. Во-вторых, механика Бома дает наглядность – детерминированные, реальные траектории частиц можно визуализировать. Последнее важно с педагогической точки зрения, позволяет лучше понять природу квантовых процессов, но не только. Визуализация важна и в квантовой химии, для моделирования того, как электроны движутся в атомах [Maddox, Bittner, 2001].

В связи с эмпирической эквивалентностью стандартной квантовой механики и механики Бома популярным является следующее возражение: механика Бома с ее добавочным направляющим уравнение избыточна. В. Аллори отвечает на это, что теория волны-пилота дает более глубокое понимание квантовых явлений, и ситуация, когда есть две физически эквивалентные теории, встречается в физике. Например, статистическая механика обосновывает на более фундаментальном, связанном с элементарными частицами уровне законы термодинамики. Более того, у теории волны-пилота есть и эмпирические подтверждения: «Экспериментальные достижения, включая слабые измерения, показали, что траектории, предсказанные теорией, тесно соответствуют наблюдаемым явлениям, что еще раз подтверждает ее физическую значимость» [Allori, 2025, p. 147].

К приведенной выше цитате К. Ровелли можно добавить, что теория Бома нравится не только реалистически ориентированным в онтологии философам. Многие физики-экспериментаторы принимают реалистическую интерпретацию квантовой механики, поскольку в своих опытах они управляют движением квантовых частиц. Манипулирование с объектом в эксперименте, например при квантовой телепортации и при работе с квантовым компьютером, убеждает в его реальности. В.Э. Терехович отмечает «сильным аргументом в споре о квантовой реальности становятся новые технологии, внедряемые в рамках второй квантовой революции» [Терехович, 2019, с. 181]. При этом убеждение в реальности квантового объекта

не ведет автоматически к принятию механики Бома, можно довольствоваться и стандартной механикой. Преимущество механики Бома перед другими интерпретациями – в том, что она дает наиболее наглядное объяснение нелокальности. Перейдем от реалистической физики к взаимосвязанной с ней реалистической, точнее холистической, философии Бома.

«Голограмма», другие метафоры и структура бытия у Бома

Философия Бома эволюционировала от ранней, с уклоном в материализм до зрелой и поздней холистической, объединяющей идеальное и материальное. В данном разделе будут представлены базовые *метафоры, а также онтологические и эпистемологические положения* Бома, чтобы в итоге охарактеризовать ключевые положения его «полновесного» онтологического реализма.

Наиболее известной и характерной метафорой Вселенной у Бома является голографическая метафора. На пластину записана голограмма, при этом часть пластины также несет всю голограмму, пусть и менее четкую. Средством воспроизведения голограммы является лазерный луч. Голографическая пластина символизирует собой скрытый, имплицитный порядок, голографическое изображение – эксплицитный, явный, развернутый порядок. Лазерный луч связан с холодвижением или переходом от скрытого к явному порядку. Принципиально, что лазерный луч когерентен или упорядочен, и только такой упорядоченный поток энергии может проявить объемное изображение. Скрытый порядок есть всегда, но его познание для нас недоступно, аналогично тому, как, глядя на голографическую пластину, мы увидим лишь размытие цветовые пятна, и лишь структурированный поток энергии проявляет четкие формы. Бом особо подчеркивал [Bohm, 1980, p. 183], что слово «голограмма» происходит от двух греческих корней: “holo” – целое, и “gram” – «пишет». Согласно Бому, целое «пишет» или воспроизводит уменьшенные копии или подобия себя. Бом называл их субтотальностями.

У Бома существует не одна, а *четыре взаимодополняющие друг друга метафоры*. Помимо *голографической, это – «рыба в аквариуме», «капля чернил» и радиосигнал кораблю.* Выше мы уже использовали метафору корабля, управляемого радиоволнами. Теперь рассмотрим две другие наглядные аналогии, к которым прибегает Бом: начнем с образа рыбы, плавающей в аквариуме [Bohm, 2005, p. 237]. С одной стороны аквариума направленная на нее телекамера снимает бок рыбы, с другой же – другая камера – хвост. Выглядят они,

естественно, по-разному, но как только движется бок рыбы, одновременно меняется положение головы. Две запутанные или взаимозависимые частицы занимают разное положение в пространстве и обладают разными характеристиками, как и части рыбы. При этом они меняются синхронно, потому что являются частями одного объекта, единой неделимой вселенной. «Рыба в аквариуме» – это то, как мы, находясь в области имплицитного, развернутого, локализованного порядка, видим в нем проявления имплицитного, целостного порядка.

П. Пюлькканен делает особый акцент на бомовской метафоре «капля чернил». Капля чернил помещена в глицерин, находящийся внутри цилиндра, и когда этот цилиндр начинает вращаться, капля растягивается в чрезвычайно тонкую, невидимую нашему глазу нить. При остановке вращения она опять становится той же каплей. П. Пюлькканен отмечает, что «капля сохраняет свою идентичность и в определенном смысле даже тождественность в свернутом состоянии» [Pylkkänen, 2007, p. 63]. Идентичность – это эпистемологическая характеристика, то, как вещь является нам, как мы отличаем ее от других вещей, а тождественность – онтологическая характеристика, связанная с внутренней устойчивостью, сохранением вещи. Бом предлагает свою метафизику вещи, в которой она рассматривается не как самодостаточная, а как зависящая от холодвижения: если это движение целого интенсивно, то вещь переходит в свернутое состояние, если оно замедляется или локализуется, то вещь принимает устойчивые очертания, мы идентифицируем ее как каплю или другой объект.

Уже упоминавшаяся метафора радиосигнала кораблю важна не только для микро-, но и для макромира. Бом использует ее применительно к сознанию: активная информация, идущая от него, управляет телом человека. Информация для Бома – не просто мера порядка. Его активная информация организует и упорядочивает, обеспечивает согласованное развитие части и целого. Физическое и ментальное неразрывно связаны: «Неразделимы в том смысле, что информация, содержащаяся в мысли, которая, как нам кажется, находится на “ментальной” стороне, одновременно является связанной с ней нейрофизиологической, химической и физической деятельностью (что, очевидно, и подразумевается под “материальной” стороной этой мысли)» [Bohm, 1990, p. 285].

Бом также признавал значение активной информации в живой природе. Организм полностью не изолирован, он развивается, направляемый информацией, поступающей не только от его генов, но и от окружающей среды. В контексте данной статьи упоминание организма человека и других существ важны в следующем отношении: Бом признает не только имплицитный нелокальный, но и эксплицитный локальный порядок макромира как относительно автономный.

На саму концепцию активной информации Бома повлияла формальная причина, которая есть не нечто внешнее, но «упорядочивающее и структурирующее внутреннее движение, определяющее сущность вещи» [Bohm, 2005, p. 16]. Часть этого внутреннего информационного движения локализована в организме, но в неизмеримо большем объеме оно присутствует в неделимой и нелокальной в своей основе вселенной. Все четыре метафоры (голографическая, радиосигнала кораблю, капли чернил, рыбы в аквариуме) имеют комплексный характер, взаимодополняют друг друга. Важно, что *все метафоры характеризуют реалистическое видение мира:* вселенная не только есть вне сознания человека, но и сама определяет свое развитие, порождая субтотальности, локализованные, относительно автономные части бытия.

Бом о познании как о процессе в бытии

При этом вселенная лишь ограниченно познаваема для квантовой физики. Бом как основатель механики скрытых параметров это прекрасно понимал. Как совместить полновесный реализм и ограниченность наших знаний о сущности мира?[2] Бом в своей философии подходит к *познанию* онтологически, рассматривает его как процесс, происходящий в бытии. В этом его сходство с рядом восточных и русских религиозных философов. Он «онтологизирует» эпистемологию. *Мысль, как и вещь, которую познает эта мысль, являются формами, абстрагированными или относительно автономными от общего процесса жизни.* Почему мысль соответствует вещи? Не в силу того, что она отражает эту вещь, а поскольку и она, и вещь имеют под собой единую основу, но отделены эпистемологически от единого целого.

Бом приводит аналогию *танца пчел* для иллюстрации процесса познания. В этом танце одна пчела указывает другим положение цветов, дающих мед. «Этот танец, вероятно, не следует понимать как создание в “уме” пчел формы знания в рефлексивном соответствии с цветами. Скорее, это деятельность, которая при правильном выполнении действует как указатель или индикатор, располагая пчел к порядку действий, который обычно приводит их к меду» [Ibid., p. 70]. Таким образом, для Бома познание – это не только не отражение, но и не рефлективное или интеллектуальное соотнесение ментальных данных с реальностью.

[2] Следует также помнить, что истинные знания о локализованных субтотальностях не есть знания о сущности мира, глубинных нелокальных уровнях бытия.

Это деятельность, точнее, часть общего эволюционного процесса: «Мысль – это своего рода “танец ума”, который функционирует указательно и который при правильном выполнении течет и сливается в гармоничный и упорядоченный вид общего процесса в жизни в целом» [Bohm, 2005, p. 70]. Обобщим: *подходу Бома к познанию свойственны прагматизм, принятие эволюционных принципов, а также холизм*. Прагматизм как критерий истины состоит в том, что истинное знание дает правильный, точнее, проходящий опытную проверку результат, а ложное – не дает, подобно танцу, который приводит к меду и который не приводит. Эволюционизм состоит в том, что правильное знание способствует эволюции или гармонирует с процессом «жизни в целом», тогда как неправильное не гармонирует. Холизм, представление о том, что познающий субъект связан с целостным многоуровневым мирозданием, отличает Бома от ряда направлений эволюционной эпистемологии с ее принципами борьбы за существование идей и познающих субъектов.

Для Бома в его холистическом видении главным является интеграция, кооперация и объединение всех элементов бытия. Причем не столько на горизонтальном уровне – с другими живущими и познающими субъектами – сколько с мирозданием в целом через нелокальные взаимосвязи. Истинное знание способствует целостному восприятию мира, когерентно с теми теориями, которые описывают вселенную как холистическую систему. При этом *субъект или познающий деятель относительно автономен от целого*: пчелы могут станцевать и танец, не ведущий к цветам. Эта относительная автономность или наличие субтотальностей в терминологии Бома и является причиной ошибок в познании.

Квантовая механика изучает высокие энергии по сравнению с механическими энергиями, доступными нам в обыденном опыте. Волна-пилот Бома (а не волна-вероятность копенгагенской интерпретации) управляет энергиями высшего уровня по отношению к более грубым вещественным, «плотным» или «телесным», энергиям. Теория волны-пилота стремится постичь более тонкие слои реальности, а не адаптировать их к тому виду материи, что доступен нам в обыденном опыте. Теория познает реальность, но при этом есть скрытые параметры. Как быть с ними? *Незнание является свойством нашего сознания, а не мира как такового*. Многое в мире мы можем познать лишь частично. Так, темная материя и энергия познаваемы лишь частично, но при этом они не просто существуют, но определяют развитие Вселенной. В теории Бома мы можем рассчитать движение лишь нескольких частиц, расчет же траекторий нескольких десятков или даже сотен частиц нам недоступны, но из этого не следует, что данных траекторий нет. Д. Белл, констатируя ограниченность наших знаний о волновой функции Вселенной, отмечает: «Признание вещей, невидимых для таких

грубых существ, как мы, на мой взгляд, свидетельствует о достойном смирении, а не просто о прискорбной зависимости от метафизики» [Bell, 1987, p. 202].

Бом vs эпистемологическая линия Бора и «пифагореизм» Гейзенберга

В предыдущих двух разделах были охарактеризованы общие онтологические и эпистемологические идеи Бома. Помня о том, что его метафизика и философия физики взаимосвязаны, можно сопоставить его и альтернативные ей интерпретации квантовой механики. Копенгагенская интерпретация предстает как исторически первая и наиболее распространенная интерпретация. При этом единой копенгагенской интерпретации нет, в нее включают зачастую противоречащие друг другу воззрения[3].

По-видимому, термин «копенгагенская интерпретация» в силу своей исторической распространенности и обобщающего характера останется востребованным, но более корректно было бы говорить о подходе Бора, субъективистских теориях Дж. фон Неймана, Ю. Вигнера, пифагореизм позднего Гейзенберга и др. В данном разделе будет проведено сопоставление научной метафизики Бома и Бора как наиболее яркого и влиятельного представителя копенгагенской интерпретации, а также онтологии Бома и Гейзенберга. Конечно, в позиции Бора и его ученика Гейзенберга много сходных черт, но для нас важнее различия, связанные с ответом на онтологический вопрос о том, что существует. «Действительно, философские взгляды Гейзенберга на квантовую физику никогда не совпадали с взглядами Бора, особенно в поздних работах Гейзенберга» [Plotnitsky, 2025, p. 26].

К ключевым идеям копенгагенской интерпретации в изложении Бора относят статистический характер волновой функции, приоритет наблюдаемости и зависимость эксперимента от его условий и экспериментальной установки, принципы дополнительности [Бор, 1971]. Физическое событие – это то, что зафиксировано в опыте, при измерении, то, что мы можем по принципу дополнительности соотнести с параметрами, знакомыми нам из классической физики. В.Э. Терехович отмечает ряд причин популярности копенгагенской интерпретации, среди них можно выделить следующую: «Истинность квантовой теории никак не противоречит истинности классических и релятивистских теорий. Такая совместимость обеспечивается принципом дополнительности, статистической интерпретацией

[3] За ответом на вопрос о том, почему так произошло, можно адресовать к статье В.Э. Тереховича [Терехович, 2025].

волновой функции» [Терехович, 2019, с. 175]. Поскольку частицы не обладают определенными свойствами до измерения, то и не требуется объяснять их «странное», неклассическое поведение. При этом принципиальным положением копенгагенской интерпретации является представление о коллапсе волновой функции, обусловленной действием прибора и экспериментатора. До измерения частица как бы находится во всех состояниях одновременно, а после измерения она принимает одно определенное состояние.

К последовательно субъективистским и, следовательно, антиреалистическим интерпретациям квантовой механики относятся уже отмеченные концепции Дж. фон Неймана, Ю. Вигнера. Бор же признает наличие внешней реальности, но при этом считает, что она доступна нам только через призму исследовательских концепций, экспериментальной установки ученого. Это реализм сущностей, во многом близкий концепции Канта. Д. Бом и Б. Хайли противопоставляют свой онтологический подход эпистемологическому подходу Н. Бора. Бор воспринимает квантовое явление целостно, но это целостность системы из самого квантового явления и экспериментальной установки: «Однозначное описание собственно квантового явления должно в принципе включать описание всех существенных частей экспериментальной установки» [Бор, 1971, с. 529]. При этом данная целостность расходится с механицизмом классической механики, в рамках которого воздействием экспериментальной установки на физические объекты можно пренебречь[4].

Бом отстаивает не эпистемологический, а онтологический холизм: для него нелокальность – важнейшая характеристика бытия, включающего в себя и экспериментальную установку, и экспериментатора как свои аспекты. Для актуализации дискуссии Бома и Бора в качестве продолжателя эпистемологической линии Бора можно рассматривать А. Цайлингера. Цайлингер провел много экспериментов, подтверждающих нелокальность, но в его интерпретации квантовой физики ключевым выступает понятие информации, доступной и данной исследователю [Zeilinger, 1999].

Немного другой взгляд на реальность предлагает Гейзенберг, популяризовавший сам термин «копенгагенская интерпретация» квантовой теории в 1950-е гг. [Гейзенберг, 1989] в качестве обобщения своей позиции, позиции Бора и ряда других ученых. Гейзенберг предлагает свою версию онтологии, считая, что современная физика

[4] Бор, применяя принцип дополнительности, предлагает нередукционистское понимание жизни и сознания, взаимодополнение телеологического и механистического подходов к пониманию жизни. Он не просто дает интерпретацию квантовой физики, но и предлагает свою натурфилософию, как и Бом, также критиковавший механицизм. При этом натурфилософия Бора в отличие от Бома является эпистемологически и методологически ориентированной.

идет вперед по пути Платона и пифагорейцев, для которых в основе мира были математические структуры. Для Гейзенберга это – уравнения, являющиеся математическим выражением свойств симметрии. Симметрия, если сильно упрощать, – это неизменность, и фундаментальные физические законы сохранения (энергии, импульса, момента импульса) основываются на том, что пространство и время однородны и симметричны. Гейзенберг констатирует: «“В начале была симметрия” – идея, безусловно, более правильная, чем демокритовский тезис “в начале была частица”» [Гейзенберг, 1989, с. 349]. Одни элементарные частицы порождают другие, многие из них неустойчивы, но мы знаем об их физических характеристиках, можем предсказать их преобразование, основываясь на уравнениях. Поэтому Гейзенберг и относит уравнения, выражающие симметрию, к основе реальности.

Пифагорейский подход Гейзенберга отличается от онтологически примитивного пифагореизма М. Тегмарка, сводящего всю реальность к математике. У Гейзенберга математика – в основе реальности, более того, он разграничивает актуальную и потенциальную реальность. Как отмечают Бом и Хайли, у Гейзенберга «волновая функция представляла собой не актуальную реальность, а скорее набор потенциальных возможностей, которые могли быть реализованы в соответствии с экспериментальными условиями» [Bohm, Hiley, 1993, p. 18]. Иными словами, математика в основе, но она реализуется в соответствии с экспериментальными условиями, определяемыми на макроуровне. В качестве одной из актуальных интерпретаций Гейзенберга можно рассматривать пифагорейскую концепцию реальности без реализма А. Плотницкого [Plotnitsky, 2025].

Заключение

Физика и философия у Бома взаимосвязаны. Этот общий тезис мы постарались прояснить на примере *эпистемологии Бома*, поскольку ей уделяется недостаточно внимания в исследованиях по философии Бома. В случае с одной частицей можно визуализировать ее траекторию. Если брать сотни частиц, то рассчитать их траектории невозможно. Их начальные координаты выступают для нас как скрытые параметры. Значит, *наши знания о реальности* весьма *ограниченны*. Квантовый реализм Бома отличается от классического, для которого физические объекты локализованы, а их свойства, такие как масса и скорость, познаваемы. При этом, согласно Бому, мы, как пчелы, ищущие мед, можем получить те знания, которые ведут нас к развитию.

В статье взаимосвязь физики и метафизики Бома поясняется через его концептуальные метафоры. Метафора радиосигнала кораблю

в его механике демонстрирует информационное, а не силовое воздействие на частицу. В метафизике она приобретает более широкое значение: информация направляет развитие макрообъектов на уровне эксплицитного порядка. Метафора радиосигнала кораблю рассматривается во взаимосвязи с другими метафорами, дополняющими друг друга. Так голографическая метафора демонстрирует, как единое целое Вселенной «пишет» или воспроизводит уменьшенные копии или подобия себя. Подчеркивается, что у Бома существует именно система метафор. Какую картину мира на основе реалистически понятой квантовой механики сформировал Бом? *Холистический многоуровневый (иерархический) реализм*. Насколько эти идеи с необходимостью следуют из квантовой механики? Холизм имеет своим основанием нелокальные взаимосвязи. Многоуровневость или иерархичность состоит в том, что есть как минимум два уровня: квантовый, нелокальный, и классический, где тела локализованы в пространстве. Причем *первичен нелокальный уровень бытия*.

В философии физики динамический холизм Бома противопоставляется эпистемологической линии Бора и пифагорейской линии Гейзенберга. Бор также признает холизм, но это холизм частицы, прибора и исследователя. У Бома холизм носит онтологический, универсальный, а не фрагментированный характер. Для Гейзенберга симметрии, выраженные в уравнениях, – в основе реальности, для Бома же смысл и значение уравнений – лишь обособленный аспект единого целого. Безусловно, «линия Бора», «линия Гейзенберга» в сопоставлении с «линией Бома» намечены лишь пунктирно, в общих чертах, развернутый анализ – предмет отдельных исследований.

Список литературы

Абрамов, Некрасова, 2024 – *Абрамов П.Д., Некрасова В.М.* Реализм и холизм научной теории волны-пилота и философии Дэвида Бома // Философия науки. 2024. № 4 (104). С. 34–50.

Бом, 1965 – *Бом Д.* Квантовая теория. М.: Наука, 1965. 728 с.

Бор, 1971 – *Бор Н.* Избранные научные труды в двух томах. Т. 2. М.: Наука, 1971. 676 с.

Гейзенберг, 1989 – *Гейзенберг В.* Физика и философия. Часть и целое. М.: Наука, 1989. 400 с.

Ровелли, 2020 – *Ровелли К.* Гельголанд. Красивая и странная квантовая физика. М.: Эксмо, 2020. 176 с.

Салом, 2024 – *Салом И.* Нобелевская премия 2022 года по физике и конец механистического материализма Часть 2: Реализм и локальность // Идеи и идеалы. 2024. Т. 16. № 4–1. С. 141–166.

Терехович, 2019 – *Терехович В.Э.* Три подхода к проблеме квантовой реальности и вторая квантовая революция // Эпистемология и философия науки. 2019. Т. 56. № 1. С. 169–184.

Терехович, 2025 – *Терехович В.Э.* Квантовая революция или революционные трансформации? // Эпистемология и философия науки. 2025. Т. 62. № 4. С. 6–24.

Фурсов, 2025 – *Фурсов А.А.* Обзор статьи Э. Гибни «Исследование Nature показало, что физики широко расходятся во взглядах о том, что квантовая механика говорит о реальности» // Эпистемология и философия науки. 2025. Т. 62. № 4. С. 222–230.

Эрекаев, 2023 – *Эрекаев В.Д.* Квантовая механика и современная философия: Онтология микромира и квантовое сознание. М.: URSS, 2023. 168 с.

References

Abramov, P.D., Nekrasova, V.M. “Realizm i holizm nauchnoj teorii volny-pilota i filosofii Djevida Boma” [Realism and Holism of the Scientific Theory of the Pilot Wave and the Philosophy of David Bohm], *Philosophy of Science*, 2024, no. 4 (104), pp. 34–50. (In Russian)

Allori, 2025 – Allori, V. “On Quantum Mechanics and the Pilot-Wave Theory: Empirical Equivalence and Other Objections”, *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 4, pp. 133–150.

Bell, 1964 – Bell, J.S. “On the Einstein-Podolsky-Rosen Paradox”, *Physics*, 1964, vol. 1, no. 3, pp. 195–200.

Bell, 1987 – Bell, J.S. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bohm, 1951 – Bohm, D. *Kvantovaja teorija* [Quantum Theory]. Moscow: Nauka, 1965. (Trans. into Russian)

Bohm, 1952 – Bohm, D. “Suggested Interpretation of the Quantum Theory in Terms of “Hidden” Variables. II”, *Physical Review*, 1952, vol. 85, pp. 180–193.

Bohm, 2005 – Bohm, D. *Wholeness and the Implicate Order*. Taylor and Francis e-Library, 2005.

Bohm, 1990 – Bohm, D. “A New Theory of the Relationship of Mind and Matter”, *Philosophical Psychology*, 1990, vol. 3, no. 2–3, pp. 271–286.

Bohm, Hiley, 1993 – Bohm, D., Hiley, B.J. *The Undivided Universe: An Ontological Interpretation of Quantum Theory*. Routledge, 1993.

Bohm, D. “Kvantovaya teoriya” [Quantum Theory]. Moscow: Nauka, 1965. (Trans. into Russian)

Bohr, N. *Izbrannye nauchnye trudy v dvuh tomah. T. 2* [Selected Scientific Works in Two Volumes. Vol. 2]. Moscow: Nauka, 1971. (Trans. into Russian)

Erekaev, V.D. *Kvantovaja mehanika i sovremennaja filosofija: Ontologija mikromira i kvantovoe soznanie* [Quantum Mechanics and Modern Philosophy: Ontology of the Microworld and Quantum Consciousness]. Moscow: URSS, 2023. (In Russian)

Essential David Bohm, 2005 – Nichol, L. (ed.) *The Essential David Bohm*. London: Routledge, 2005.

Fursov, A.A. "Obzor stat'i Je. Gibni 'Issledovanie Nature pokazalo, chto fiziki shiroko rashodjatsja vo vzgljadah o tom, chto kvantovaja mehanika govorit o real'nosti'" [Review of E. Gibney's Paper 'Physicists Disagree Wildly on What Quantum Mechanics Says about Reality, Nature Survey Shows'], *Epistemology & Philosophy of Science,* 2025, vol. 62, no. 4, pp. 222–230. (In Russian)

Heisenberg, W. *Fizika i filosofija. Chast' i celoe* [Physics and Philosophy. Part and Whole]. Moscow: Nauka, 1989. (In Russian)

Gibneys, 2025 – Gibneys, E. "Physicists Disagree Wildly on What Quantum Mechanics Says about Reality, Nature Survey Shows", *Nature,* 2025, vol. 643, pp. 1175–1179.

Globus, 2007 – Globus, G. "Mind, Matter and the Implicate Order by Paavo TI Pylkkänen", *NeuroQuantology,* 2007, vol. 5, no. 4, pp. 396–399.

Hiley, Van Reeth, 2018 – Hiley, B.J., Van Reeth, P. "Quantum Trajectories: Real or Surreal?", *Entropy,* 2018, vol. 20, no. 5, 353.

Maddox, Bittner, 2001 – Maddox, J.B., Bittner, E.R. "Quantum Relaxation Dynamics Using Bohmian Trajectories", *The Journal of Chemical Physic*s, 2001, vol. 115, no. 14, pp. 6309–6316.

Plotnitsky, 2025 – Plotnitsky, A. "Heisenberg's Pythagorean Revolution: from Quantum Jumps to Matrix Algebra and Beyond", *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 4, pp. 25–42.

Pylkkänen, 2007 – Pylkkänen, P. *Mind, Matter and the Implicate Order*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

Rovelli, K. *Gel'goland. Krasivaja i strannaja kvantovaja fizika* [Helgoland. The Beautiful and Strange Quantum Physics]. Moscow: Eksmo, 2020. (Trans. into Russian)

Salom, 2024 – Salom, I. "Nobelevskaja premija 2022 goda po fizike i konec mehanisticheskogo materializma" [The 2022 Nobel Prize in Physics and the End of Mechanistic Materialism. Part 2: Realism and Locality], *Ideas and Ideals*, 2024, vol. 16, no. 4–1, pp. 141–166. (In Russian)

Sivasundaram, Nielsen, 2016 – Sivasundaram, S., Nielsen, K.H. *"Surveying the Attitudes of Physicists Concerning Foundational Issues of Quantum Mechanics"*, arXiv preprint arXiv:1612.00676, 2016.

Terekhovich, V.E. "Tri podhoda k probleme kvantovoj real'nosti i vtoraja kvantovaja revoljucija" [Three Approaches to the Problem of Quantum Reality and the Second Quantum Revolution], *Epistemology & Philosophy of Science*, 2019, vol. 56, no. 1, pp. 169–184. (In Russian)

Terekhovich, V.E. "Kvantovaja revoljucija ili revoljucionnye transformacii?" [Quantum Revolution or Revolutionary Transformations?], *Epistemology and Philosophy of Science,* 2025, vol. 62, no. 4, pp. 6–24. (In Russian)

Teufel, Dürr, 2009 – Teufel, S., Dürr, D. *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory.* Berlin, Heidelberg: Springer-Verlag, 2009.

Zeilinger, 1999 – Zeilinger, A. "A Foundational Principle for Quantum Mechanics", *Foundations of Physics*, 1999, vol. 29, no. 4, pp. 631–643.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 149–167
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 149–167
DOI: https://doi.org/10.5840/eps202663346

Научно-технологический утопизм как способ мышления альтернатив науки. Заметки об «утопическом методе»*

Туркенич Даниил Юльевич –
аспирант.
Институт философии РАН.
Российская Федерация,
109240, г. Москва,
ул. Гончарная, д. 12, стр. 1.
Исследователь.
Межрегиональная общественная организация «Русское общество истории и философии науки».
Российская Федерация,
105062, г. Москва,
Лялин пер., 1/36, стр. 2;
e-mail: turkenichdaniil@yandex.ru



Статья посвящена проблеме восприятия научно-технологического утопизма в области эпистемологии, философии науки и социологии знания. Автор показывает, что предложенное К. Манхеймом определение утопии как одной из двух социально обусловленных «ошибок» мышления привело к широко распространенному в философии гиперкритицизму в отношении утопического воображения. Показано, что этот гиперкритицизм («утопофобия») характеризуется категориальной подменой: вместо анализа структуры, элементов и потенциала утопизма его сторонники фокусируются на критике результатов практических воплощений утопических проектов. Отмечается, что в последнее десятилетие в исследованиях утопий наметился сдвиг в сторону анализа эвристического и критического потенциала утопизма как практико-ориентированного способа мышления. Автор обращается к идеям утопии как метода британского социолога Р. Левитас. В одном из разделов статьи обсуждаются избранные фрагменты монографии немецкого философа политики и техники М. Занда «Технологический утопизм и теория справедливости» (2024). В заключение поднимается вопрос о том, возможна ли разработка «утопического метода» в области эпистемологии и философии науки. Автор полагает, что данная новация позволит обогатить исследовательскую традицию эпистемологии, введя в поле исследований науки, техники и общества проработку и обсуждение их альтернативных образов и вариаций.

Ключевые слова: утопия, утопизм, научно-технологический утопизм, утопия как метод, эпистемология, философия науки, социология знания, К. Поппер, К. Манхейм

Scientific and technological utopianism as the way of thinking of science alternatives. Notes towards the “utopian method”

Daniil Yu. Turkenich –
PhD Student.
Institute of Philosophy,
Russian Academy of Sciences.
12/1 Goncharnaya St.,
Moscow 109240,
Russian Federation.

The article addresses the problem of how scientific and technological utopianism is presented in epistemology, philosophy of science, and the sociology of knowledge. The author shows that K. Mannheim’s definition of utopia as one of two socially conditioned “errors” of thinking has led to a widespread hypercriticism towards the utopian imagination in philosophy. It is shown that hypercritism (“utopophobia”) is characterised by a categorical

* Исследование выполнено за счет гранта Российского научного фонда № 26-18-00851, https://rscf.ru/project/26-18-00851/

© Туркенич Д.Ю., 2026

Researcher.
Interregional
Non-Governmental
Organization “Russian Society
for the History and Philosophy
of Science”.
1/36 Lyalin lane, bd. 2,
Moscow 105062,
Russian Federation;
e-mail: turkenichdaniil@
yandex.ru

substitution: instead of analysing the structure, elements, and potential of utopianism, its proponents focus on criticising the outcomes of practical implementations of utopian projects. The article notes that in the last decade, utopian studies have seen a shift towards analyzing the heuristic and critical potential of utopianism as a practical way of thinking. The author refers to the ideas of utopia as method by the British sociologist Ruth Levitas. One section of the article examines selected fragments of German political and technological philosopher Martin Sand’s book *Technological Utopianism and the Idea of Justice* (2024). The conclusion raises the question of whether a “utopian method” can be developed in epistemology and philosophy of science. The author believes that such an innovation would enrich the epistemology tradition by introducing the study and meditations on alternative images and variants of science, technology, and society.

Keywords: utopia, utopianism, scientific and technological utopianism, utopia as a method, epistemology, philosophy of science, sociology of knowledge, K. Popper, K. Mannheim

Введение

Современная эпоха характеризуется стремительным расширением пространства агентности науки и техники, которые перестают быть только средствами познания и трансформации мира, все чаще выступая в качестве основы для всеобъемлющих социальных, экономических и политических программ. Это расширение влечет за собой не только утрату нейтральности науки, но и ее включение в системы управления, идеологического регулирования и социального мифотворчества. Хотя миф науки (представление об особом статусе ученых и исключительном положении науки среди других видов деятельности) сегодня активно критикуется и деконструируется, именно эта критика влечет за собой построение так называемых «негумбольдтовских зон обмена» – площадок взаимодействий науки и общества, которые обеспечивают как локализацию этих взаимодействий, так и более эффективный трансфер знаний между учеными и общественными институтами [Масланов, 2024, с. 164]. В связи с этими изменениями открывается пространство для формирования нового – техно-утопического – мифа современной науки: в контекстах соперничества за эффективные решения экзистенциально значимых проблем человечества (изменений климата, кризисов, бедности, нехватки ресурсов) между стратегиями национального технологического развития и корпоративными мегапроектами неизбежно рождаются и процветают научно-технологические утопические нарративы, рассматривающие науку и технику как универсальные средства окончательной гармонизации всех сфер жизнедеятельности [West, 1996; Vohnsen, 2024].



В либерально-демократических исследованиях утопий складывается консенсус, согласно которому «большие формы» воплощения утопизма подвергаются жесткой критике за экстраполяцию его недостатков на все охваченные им явления, а также за отсутствие плюрализма и возможностей для изменений в рамках утопических проектов [Martin, 2023, p. 106]. Исторический опыт XX в. (в том числе Советский проект) показал как потенциал, так и риски утопической эксплуатации науки и всех сфер общественной жизни. В конце XX в. эпоха «больших форм» ушла, постмодернистами были провозглашены «крах метанарративов» и «конец истории». Однако утопическое воображение не исчезло, так как оно укоренено в самой структуре человеческого сознания. Как писал немецкий философ Э. Блох, утопия – это «опосредование нового», проистекающее из незавершенного, еще-не-окончательного характера действительности [Блох, 1991, с. 50]. Соответственно, полностью отказ от утопии невозможен. Поэтому в XXI в. вместе с трансформировавшимся миром трансформировалось и утопическое воображение, приняв новые формы в современных идеологиях технологического прогресса, корпоративного активизма, стартап-культуры, независимых футурологических лабораторий и либерально-демократической «поступательной» инженерии. Как точно отмечает исследовательница Нина Холм Вонсен: «В первые десятилетия XXI века, как я полагаю, прекрасно написанные, сатирические романы и скрупулезные политические программы прошлых веков сменились слайдами PowerPoint, бизнес-кейсами и программными заявлениями…» [Vohnsen, 2024, p. 81].

Трансгуманистические лаборатории, разработки в области общего искусственного интеллекта и коммерческие космические программы воспроизводят в своих риториках утопические нарративы, обещая разрешение всех социальных противоречий средствами научно-технического прогресса. В отличие от утопизма XX в., сегодня драйверами утопизма выступают не столько правящие партии, сколько частные корпорации и миллиардеры-футурологи. Локализация утопизма в рыночной экономике позволяет создавать наиболее притягательные образы независимых лабораторий и предпринимателей-«технопророков» как спасителей человечества, способных эффективно решить насущные проблемы современности не политическими, а коммерческими, научными и техническими средствами – т.е. органично, ненасильственно и поступательно.

Научно-технологический утопизм двойственен. С одной стороны, он угрожает эпистемической автономии науки, подчиняет научную рациональность своим собственным утопическим задачам. В таких условиях наука теряет свою специфику и становится инструментом в чужих руках [Черткова, 2010, с. 266]. С другой стороны, он согласуется с гуманистическим проектом науки, с ее ценностным

содержанием, стимулируя тем самым взаимодействия науки и общества во все более широких контекстах.

Однако нас в большей степени волнует вопрос не о том, обладает ли сам научно-технологической утопизм позитивным содержанием. Мы убеждены, что утопизм может быть представлен как инструмент осмысления, анализа и критики тех явлений, которые он предлагает использовать и трансформировать. Таким образом, утопизм как одна из форм утопического воображения сам по себе обладает эпистемическим и критическим потенциалом. В данной статье с опорой на тенденции последних лет в области исследования утопий мы предлагаем рассмотрение научно-технологического утопизма как своеобразного метода исследования самих науки и технологий, а также их взаимосвязей с обществом. Подобно тому, как Юрген Хабермас фиксировал кризис критической теории в связи с полным разрушением ее собственных оснований и неимением каких-либо иных внешних опор для ее дальнейшего успешного развития, в связи с чем критическая теория нуждалась в глобальном пересмотре [Habermas, 1987, p. 118], мы полагаем, что социальная философия науки при анализе взаимосвязей науки, техники и общества так же нуждается в разработке некоторых альтернативных оснований, позволяющих ей избегать фиксации и оправдания существующего порядка вещей (выражением которого в некотором смысле можно считать известный гегелевский принцип «все действительное разумно, все разумное действительно). Разработку таких альтернатив, на наш взгляд, способны предложить утопические научно-технологические проекты, так как последние ставят под сомнение постоянство и неизменность существующих на данный момент общественных институтов, а также ограничений в технических и познавательных возможностях человека. Утопизм обладает эффектом «остранения» (термин Виктора Шкловского) и деятельно противостоит «презентизму», т.е. приданию исключительных черт имеющемуся положению дел при одновременном принижении значимости прошлого и будущего [Герасимова, 2008, с. 188].

Общим проблемам рассмотрения научно-технологического утопизма как практико-ориентированного способа мышления альтернатив науки, связанным с ним актуальным тенденциям анализа эвристического потенциала утопизма в исследованиях утопий и перспективам локализации этих тенденций в области эпистемологии и будет посвящена данная статья.



Карл Манхейм и «утопофобия»

Классическое для эпистемологии и социологии знания определение утопии как способа мышления, как известно, дал основоположник социологии знания Карл Манхейм в работе «Идеология и утопия». Согласно Манхейму, утопия наравне с идеологией является одним из основных видов социально обусловленных заблуждений; задачей же социологии знания является исследование содержания мышления социальных групп, конкретно – вскрытие и развенчивание этих заблуждений. Манхейм предлагает осмысливать идеологию и утопию как антагонизмы: под первой понимается такое мышление, которое описывает и легитимирует существующий социальный порядок в угоду правящим классам; под второй понимается мышление, которое, ориентируясь на факторы, трансцендентные окружающему его бытию, стремится к «взрыву» существующего порядка, который не устраивает подчиненное население [Манхейм, 1994, с. 164]. Напряжение между идеологией и утопией существует в связи с тем, что и первая, и вторая описывают лишь часть реальности, представляют схематичное изображение возможных форм организации социального бытия, претендуя при этом на истинность своих представлений; более того, завоевание трансцендентного является для каждой формы мышления первостепенной задачей, так как идеология, овладевая факторами, находящимися вне существующего порядка вещей, способна их нейтрализовать и держать под контролем, а утопия, в свою очередь, стремится сделать эти факторы предметом своей деятельности и сообщить им тем самым преображающую силу [Там же, с. 165].

Подробно анализируя различные формы идеологий и утопий, Манхейм тем не менее не придает значения тому обстоятельству, что «заблуждение» вытесненных или привилегированных социальных групп может быть продуктивным и производить познавательные конструкции разной степени успешности и содержательности. Соответственно, мы не встречаем у Манхейма размышлений о том, что утопическое мышление помимо «взрыва» существующего порядка вещей способно предлагать его творческое переосмысление, критическую рефлексию, альтернативные модели общественных отношений. К тому же логичным следствием взглядов Манхейма на предмет и задачи социологии знания выступает ситуация, названная культурологом Клиффордом Гирцем в работе «Интерпретация культур» (1973) «парадоксом Манхейма». Согласно Гирцу, парадокс Манхейма породил для современной социологической мысли нерешаемую «загадку Сфинкса» о том, в какой именно момент при изучении идеологии сама идеология в позиции ученого отходит на второй план, а наука выходит на первый [Geertz, 1973, p. 194]. Мы понимаем, что аналогичная трудность возникает и при анализе утопии,

элементы которой могут неявно детерминировать социально-гуманитарное знание и исследования утопий в частности [Трунов, 2020, с. 712]. Известным примером парадокса Манхейма выступает идеологизация самого понятия идеологии в марксистской теории. На наш взгляд, примером парадокса в отношении утопии могут выступать предложения философов и социологов (например, Карла Поппера) полностью вывести утопическое воображение и утопические практики из жизни человека по причине их тоталитарных наклонностей. Предложения эти сами по себе оказываются утопичными, поскольку в силу своего упрощенного представления об утопическом воображении и его роли в мышлении наивно допускают возможность это мышление трансформировать.

На взглядах исследователей, требующих полного отказа от утопизма и утопического воображения, следует остановиться подробнее. Перенимая отношение Манхейма к утопии как к «заблуждению», гиперкритически настроенные исследователи утопий в области философии, социологии и политологии разделяют общую исследовательскую презумпцию, согласно которой любые формы утопического воображения, будучи едва только выявленными, непременно должны быть демаркированы от иных форм гуманитарного знания и подвергнуты сокрушительной критике за их несостоятельность, антинаучность и неправомерный выход за пределы того, что традиционно считается «реалистичным» или «вероятно достижимым». Эта презумпция названа политическим философом Дэвидом Эстлундом «утопофобией». Как пишет Эстлунд, «…утопофобия – это необоснованный страх перед утопизмом, который может вести к маргинализации исследований и инсайтов, не страдающих дефектами, приписываемыми утопизму» [Estlund, 2014, p. 116]. В данной статье мы попытаемся продемонстрировать, что при всей своей внешней притягательности данная презумпция лишь блокирует постановку плодотворных вопросов об альтернативных сценариях развития общества, науки и техники, сужая тем самым горизонты философского и социогуманитарного исследования.

Классические выражение «утопофобского» подхода к оценке утопизма в эпистемологии и социальной философии XX в. представил Карл Поппер в работе «Открытое общество и его враги». Поппер резко критикует «утопическую социальную инженерию» и вообще попытки переустройства общества по заранее задуманному плану. Реконструируя историю утопического мышления, он усматривает в интеллектуальных проектах Платона, Гегеля и Маркса истоки тоталитаризма XX в. По его мнению, стремление к построению идеального общества неминуемо ведет к насилию, централизованному диктаторскому управлению и игнорированию реальных последствий отдельных действий во имя абстрактной цели [Поппер, 1992, с. 202]. Утопическая социальная инженерия представляется им эпистемически



несостоятельной и глубоко иррациональной, поскольку, согласно Попперу, рациональными средствами предсказать будущее развитие общества невозможно. В противовес Поппер предлагает стратегию «поэтапной (piecemeal) социальной инженерии» – принятие частных решений, направленных на устранение конкретных проблем здесь и сейчас, а не на достижение далекой большой идеи. Такой подход, по его замыслу, должен сделать общественное развитие более стабильным, предсказуемым и защищенным от радикальных экспериментов [Поппер, 1992, с. 200].

Однако представленная Поппером концепция демонстрирует отождествление философом утопии и утопизма с идеологией, упускает из виду позитивный и критический потенциал утопизма как формы воображения и социальной практики. Как следствие, Поппер осуществляет категориальную подмену, которая, на наш взгляд, была заложена еще в подходе К. Мангейма и является характерной чертой исследователей-«утопофобов»: он переносит свои политические интуиции, связанные с конкретными результатами утопического проектирования, на эпистемологическую оценку утопического воображения как такового. Соответственно, философ недооценивает роль этого воображения в проектировании социальной реальности. Из этого следует, что социальная инженерия, принимая общественно значимые решения, как и наука, участвующая в разработке и принятии этих решений, по рекомендации Поппера должны лишь придерживаться стандарта, который Имре Лакатос метко обозначил как «затасканный образец проб и ошибок» [Лакатос, 2008, с. 436].

Оценка Поппером утопизма и соответствующих ему социальных практик оказала масштабное влияние на исследования утопий, утопизма и утопического воображения. Статус утопизма как деструктивного и тоталитарного явления, заслуживающего на каждом витке своего развития только сокрушительную критику, вероятно, во многом обязан именно взглядам Поппера.

Однако даже в работах современных исследователей, старающихся преодолеть гиперкритицизм в отношении утопического воображения, мы встречаем следы обозначенной выше утопофобской подмены. Ограничимся здесь анализом серии статей отечественного эпистемолога и исследовательницы утопического сознания Е.Л. Чертковой. Выбор ее работ обусловлен тем, что они посвящены непосредственно эпистемологическому анализу научно-технологического утопизма как формы сознания и способа мышления, в отличие от большей части отечественных авторов, представляющих литературную и религиозно-философскую традицию исследования утопий. О подходах к анализу утопизма, подобных попперовскому, она справедливо пишет: «Законное стремление преодолеть утопизм с его радикализмом, нигилизмом и фанатизмом вылилось в огульное отрицание какого-либо позитивного смысла утопии как обладающей

специфическими задачами и социальными функциями» [Черткова, 2015, с. 191].

Наиболее известным и полным воплощением утопизма в науке выступает, по ее мнению, сциентизм. Позитивистская фиксация на эмпирически получаемом знании как единственно возможной истине, противопоставление им познавательный ценности специальных наук и абстрактной спекулятивности философских исследований открыли возможности для предельной упрощения и искажения понятий знания, познания и науки.

На примере сциентизма мы можем видеть, насколько прочные позиции утопизм способен завоевать в различных формах культуры и в науке в частности. Причиной этому могут служить, по мнению Чертковой, некоторые внешние сходства между наукой и утопизмом: им обоим присущи рациональность, критичность и проективность [Черткова, 2010, с. 265]. Однако в действительности в этих же свойствах коренятся и глубинные различия: 1 – рациональность утопизма предельно ангажирована; 2 – критичность утопизма направлена только вовне; 3 – проективность утопизма ставит перед собой зачастую нерешаемые задачи. Благодаря таким псевдорациональным элементам утопизм успешно вторгается в культуру и подчиняет ее себе, оставаясь трудноразличимым. Соответственно, задача отделения науки от утопизма едва ли может быть в полной мере выполнимой. Самым ярким и маргинальным примером засилия утопизма в науке, по мнению Чертковой, выступает так называемая «теория новой биологии», разработанная биологом Э.С. Енчменом [Там же, с. 273]. Несмотря на фантасмагорический характер, идеи Енчмена благодаря своему радикализму вызывали интерес у раннесоветской молодежи.

При всей глубине критики Чертковой научно-технологического утопизма мы можем заметить, что ее основные интуиции сосредотачиваются вокруг наиболее деструктивных форм утопизма в науке. Утопизм в ее интерпретации предстает только как фактор искажения и деформации научной рациональности, угрожающий ее специфике и автономии. При этом ею полностью игнорируется критический и эвристический потенциал самого утопического воображения, в том числе утопизма, который хотя и имеет предельно ангажированный характер, совмещенный с псевдорациональными элементами и способностями к мимикрии под науку, все же представляет собой интерес как реальная практика разработки и претворения в жизнь альтернативных моделей общества и науки[1].

[1] О практической ориентированности утопизма и разнице утопизма и утопии см. подробнее: *Черткова Е.Л.* Утопия как тип сознания // Общественные науки и современность. 1993. № 3. С. 71–81.



Гиперкритицизм в отношении утопизма имеет, на наш взгляд, имеет ряд негативных последствий для эпистемологии и исследований утопий. Научно-технологические утопии могут выступать не только как «схематизмы», но и как мысленные эксперименты, позволяющие исследовать границы возможного. Утопофобия, стигматизируя эти проекты как наивные и опасные, закрывает возможность для их систематического эпистемологического анализа. При этом исследователи, разделяющие гиперкритические позиции, не способны распознать утопические элементы в самой «реалистической» науке, так как утопофобский гиперкритицизм направлен исключительно вовне, на явные, маркированные утопии. Поэтому он слеп к собственным утопическим допущениям, которые могут скрываться в «здравом смысле», «реализме» или в неявных нормативных предпосылках самой утопофобской позиции. Утопофобия, не различая типы и формы утопизма, не анализируя его процессуальные аспекты, игнорируя его место и роль в культуре и науке, не может в результате эффективно противостоять наиболее опасным его формам именно потому, что последние, как подмечала Черткова, маскируются под «научные» и «реалистические». Наконец, в подобных исследованиях происходит подмена эпистемологического анализа практическими опасениями. Утопический проект может давать новое знание о возможных конфигурациях общества, науки, технологий, даже если его практическая реализация была бы опасной или невозможной. Утопофобия же отказывает в праве на существование такому знанию; как следствие, эпистемология и философия науки лишаются важнейших инструментов критики существующих научных и общественных практик, выявления их скрытых ценностных допущений и проектирования альтернативных траекторий развития науки и технологий.

Воображаемое воссоздание общества

Сказанное выше, на наш взгляд, позволяет выдвинуть гипотезу о том, что утопизм заслуживает рассмотрения в качестве инструмента и метода исследования. Поэтому преодоление расхожей в либерально-демократических и отечественных постсоветских исследованиях утопий стигматизации утопического мышления требует в рамках эпистемологии и философии науки смещения фокуса с анализа внешних форм утопизма и его конечных результатов на исследование его глубокого эвристического и критического потенциала. В контексте данной статьи речь идет конкретно о переходе к анализу утопизма как метода изучения и критики самой науки, ее институциональных структур и общества, находящегося под мощным

влиянием научно-технического прогресса. Как отмечает Р.И. Соколова, в современной ситуации неопределенности цивилизационного развития как в России, так и в зарубежных странах философы и социальные ученые все чаще предлагают реабилитацию утопического мышления как средства разработки вариантов выхода из кризиса и идейного вакуума [Соколова, 2019, с. 10]. Характерно, что подобная трансформация в проблематике исследований утопий наблюдается в последнее десятилетие: как зарубежные, так и отечественные исследователи нередко обсуждают не классические утопические тексты, персоналии и содержание утопических проектов, а вопросы понимания утопии как инструмента осмысления и преобразования мира [Паниотова, 2024]. Далее мы представим обзор нескольких современных исследовательских позиций, принципиально значимых для данного «утопического поворота».

Наиболее востребованное, по нашему мнению, переосмысление роли утопического мышления в социальных науках предложила британская исследовательница утопий и социолог Рут Левитас в работе «Утопия как метод. Воображаемое воссоздание общества» (2013). Она предлагает понимать утопию не как конечную цель общественных преобразований, а как метод «воображаемого воссоздания общества» (Imaginary Reconstitution of Society, IROS). Левитас подчеркивает, что IROS не является принципиально новым методом, так как он лишь выявляет процессы, изначально присущие утопической спекуляции и социальной теории (идея использования утопии в качестве метода социологии была предложена еще Гербертом Уэллсом в эссе «Так называемая наука социологии» [Wells, 1906]). Методологический сдвиг, осуществленный Левитас, заключается в легитимации и систематическом применении этих процессов.

Центральное место в концепции Левитас занимают три взаимосвязанных модуса утопии как метода. Археологический модус направлен на выявление скрытых, неявных утопических представлений о «благом обществе», которые укоренены в политических программах, экономических стратегиях и культурных дискурсах. Левитас пишет: «Посылка утопии как археологии состоит в том, что большинство политических позиций содержат имплицитные образы благого общества и взгляды на то, каковы люди и какими они должны быть» [Levitas, 2013, p. 154]. Археологический модус утопического метода подобен работе археолога, который по фрагментам и осколкам реконструирует целое. Задача исследователя-социолога на данном этапе – «раскопать» эти образы, сделать их явными и подвергнуть критическому анализу.

Онтологический модус отталкивается от вопроса о том, какими именно людьми нас делают существующие социальные и институциональные порядки и как могли бы измениться члены общества

при иных его конфигурациях. Левитас указывает, что любое представление о «благом обществе» неизбежно включает себя какие-либо убеждения о природе и процветании человека. Онтологический модус «озабочен именно субъектами и агентами утопии, теми включенными в нее “я”, которых утопия поддерживает или допускает» [Levitas, 2013, p. XVII].

Третий, архитектурный модус – это собственно конструирование целостных, институционально-специфичных альтернативных сценариев будущего. Левитас подчеркивает, что архитектурный режим – это не создание утопических «схематизаций» в том смысле, который вкладывают в это понятие критики утопий и утопизма. Напротив, это гипотетическое, условное и открытое для критики моделирование возможных социальных и институциональных порядков. В отличие от традиционных утопий, фиксирующих статичный идеал, архитектурный модус подразумевает постоянную рефлексию, сверку, критику и диалог.

Ключевое преимущество понимания утопии как метода, по мысли Левитас, заключается в преодолении ограничений как «наивного» утопизма, так и гиперкритических подходов, описанных нами ранее как «утопофобских». Утопическое воображение выполняет критическую функцию остранения, показывая, что существующие институты и практики не являются безальтернативными и единственно возможными. Следовательно, утопический метод «воображаемого воссоздания общества» легитимизирует утопическое воображение как валидную и эпистемически ценную форму познания.

Научно-технологический утопизм и идея справедливости

Актуальным примером использования утопии как метода философского анализа выступает упомянутая нами работа немецкого политического философа Мартина Занда «Технологический утопизм и идея справедливости» (2024). Согласно его концепции «полезного участия» (worthwhile engagement), ценность исследовательского взаимодействия с научно-технологическими утопиями (такими как трансгуманизм, цифровая утопия, виртуальная реальность) заключается в способности последних расширять наше представление о масштабах и локализации справедливости, а также выявлять скрытые допущения, которые лежат в основе классических моральных теорий справедливости и социально-политических решений [Sand, 2024, p. 3].

Согласно Занду, классические теории справедливости (например, Джона Ролза, Томаса Нагеля) строятся на так называемых «обстоя-

тельствах справедливости» – наборе фоновых допущений о человеческой природе и мире, которые считаются неизменными: умеренная скудость ресурсов, ограниченный альтруизм, человеческая смертность и ограниченность знаний при принятии решений [Sand, 2024, p. 46]. Утопические научно-технологические проекты нередко позволяют пересмотреть эти неизменные условия, помыслить альтернативные ситуации, при которых силами науки и техники они были бы устранены, и, как следствие, вскрыть случайный и социально сконструированный характер решений, принимаемых с опорой на эти условия. Как отмечает Занд, многие научно-технологические утопии ставят под вопрос неизменность этих допущений и предлагают помыслить ситуации, в которых представления о справедливости не скованы озвученными обстоятельствами [Ibid., p. 2].

Занд подчеркивает, что его защита технологического утопизма как инструмента переосмысления отношений между индивидуальным, общественным и техническим не является оправданием утопизма как такового; его интересуют лишь те вариации утопизма, которые соответствуют этому критерию и способны тем самым расширить наше понимание о локализации и масштабах справедливости. В пятой главе своей монографии «Справедливость в технологической утопии» он подробно разбирает конкретные примеры того, как ряд современных утопических проектов позволяет развивать и пересматривать данную проблематику. Предлагаем обратиться к анализу этих кейсов.

Как известно, и Дж. Ролз, и Т. Нагель признают в качестве «фонового допущения» подверженность человечества естественной генетической лотерее, которая во многом обуславливает распределение сил и возможностей в обществе. В свою очередь, задача общественных институтов – хотя бы частично пытаться компенсировать и уравновешивать эти биологически обусловленные различия между людьми. Более трудным представляется вопрос о распределении благ на основе талантов и усилий, так как при попытке объяснить успех определенного агента невозможно точно отследить, обусловлен ли этот успех исключительно усилиями или же природная склонность к какой-либо деятельности также сыграла существенную роль. Следовательно, распределение благ и привилегий на основе одних только усилий агентов не может быть по-настоящему справедливым. Само собой, о справедливости этих биологических предзаданностей политические и моральные философы предпочитают не дискутировать – в отличие от трансгуманистов. Трансгуманизм предлагает сместить ответственность за определение справедливости с индивидов и общественных институтов в область технологий, которые могли бы уменьшить влияние биологических причин на социальное неравенство, так как им постулируется возможность генетического редактирования человека и исключения



тем самым болезней как явления. Хотя остается открытым вопрос о том, насколько трансгуманизм способен отбиться от популярных обвинений в элитаризме, связи с евгеникой и отрицании права людей с болезнями на существование, стоит признавать за ним по меньшей мере два достижения: во-первых, проблематизацию справедливости биологических предзаданностей человека и, соответственно, попытку сделать биологический фактор элементом этического дискурса; во-вторых, введение в этот дискурс технологий в качестве нового субъекта справедливости наравне с индивидами и общественными институтами.

Еще одно базовое условие справедливости, поставленное трансгуманизмом, различными вариациями техноутопизма и «цифровыми утопиями» под сомнение, – это дефицит благ; утописты предлагают помыслить ситуацию, при которой дефицита просто не существует. Соответственно, перестанет существовать и большая часть связанных с распределением благ моральных проблем. Возникает, однако, другая проблема: достижение изобилия не решает сам вопрос о равномерном, справедливом и своевременном распределении благ с целью повышения качества жизни человека, а лишь делает эти блага повсеместно доступными. Изобилие товаров, доступных каждому, неизбежно приведет к экологической катастрофе, так как многие желания человека определены не столько биологической, сколько психологической потребностью в эксклюзивности и уникальности обладаемого блага. Занд подчеркивает: если, например, в какой-то момент времени людям повсеместно станет доступно жилье большей площади, чем есть у них сейчас, то в последующем другие люди пожелают заполучить жилье еще больших размеров. Попытка удовлетворить этот запрос оборачивается абсурдом.

Однако этот абсурд разрешается в концепции цифровой утопии: современные цифровые технологии, в том числе Интернет и виртуальная реальность, способны удовлетворить постоянно растущий запрос на новые и эксклюзивные блага, не производя последние физически, а предоставляя их цифровые копии в виртуальных пространствах. Многие товары – книги, произведения искусства, фильмы и музыка – в цифровом виде доступны каждому желающему, так как они не лимитированы количеством физических экземпляров. Следовательно, цифровое распределение благ хотя бы частично делает «общество постдефицита» возможным. И если мы признаём реализм виртуального опыта, то, соответственно, очень многие потребности людей, связанные с социализацией, самовыражением и прочим, будут успешно удовлетворяться благодаря созданию виртуальных пространств, цифровых аватаров. Занд цитирует Дэвида Чалмерса, отмечавшего, что, согласно Дж. Ролзу, дефицит – условие справедливости; следовательно, при упразднении дефицита средствами цифровой утопии принципы

справедливости вообще перестанут действовать, так как утратят свою необходимость [Sand, 2024, p. 141]. Как пишут отечественные авторы О.Н. Халуторных и Е.А. Егорова: «В рамках цифровой утопии высшее благо – информация; величайшее достижение социума – равенство, выражающееся в одинаковом доступе к информации и воздвигнутое на базе новых медиа» [Халуторных, Егорова, 2022].

Подобными средствами, по мнению Занда, может быть достигнута и всеобщая гармония в обществе. Пожалуй, на данный момент только благодаря Интернету, социальным сетям, виртуальным пространствам и видеоиграм представители всех возможных социальных, культурных, этнических и прочих групп способны создавать свои собственные сообщества, одновременно и достаточно изолированные друг от друга во избежание конфликтов на почве несовместимых ценностей, и не препятствующие свободному перемещению между ними. Более того, виртуальный опыт позволил бы некоторым людям заместить опыт реальный, недоступный им в силу ограничений здоровья или иных факторов. Таким образом, цифровая утопия становится большой платформой, «метаутопией» для утопий частных и более локальных. Естественно, цифровая утопия справедливо критикуется как способ эскапизма, избегания реальности. Тем не менее она также ставит вопросы о справедливости тех условий жизни человека, которые традиционно понимаются как данность, и ищет ответ на эти вопросы в области технологий.

Вместо заключения: как возможен утопический метод?

Проанализировав как гиперкритические позиции в отношении утопизма, так и набирающие все большую популярность подходы к рассмотрению утопии как метода и выявив некоторые преимущества последних, мы можем теперь задаться вопросом: каким образом эта тенденция исследования утопий как инструмента для осмысления и критики мира может быть продуктивно задействована в философии науки и эпистемологии применительно к научно-технологическому утопизму? Иными словами, можем ли мы не просто ограничиться признанием эвристической ценности утопического воображения и утопических проектов, но и превратить их самих в рабочий инструмент изучения науки, ее институтов и их взаимодействия с обществом? Ответ на этот вопрос, разумеется, не может быть сейчас однозначным и окончательным. Со своей стороны мы лишь предложим свое видение того, как данный тренд может быть взят «на вооружение» эпистемологами и философами науки.



На наш взгляд, «утопический метод» в социальной философии науки может начаться с подбора верной исследовательской оптики: вместо того чтобы рассматривать научно-технологические утопические проекты как угрозу научной рациональности, как опасные технократические и тоталитарные планы общественных преобразований, мы предлагаем вслед за упомянутыми нами ранее авторами отнестись к ним как к мысленным экспериментам, позволяющим выявить скрытые ценностные допущения науки, подвергнуть анализу те институциональные факторы науки как социального феномена, которые обычно не попадают в поле зрения исследователя. Задача такого исследования – не диктовать институту науки какие-либо чуждые ей ценности, а расширить пространство для рефлексии над целями и границами научно-технического развития.

Конкретные исследовательские операции, которые могли бы составить содержание утопического метода, можно представить, адаптировав предложенные Левитас модусы метода IROS. Археологический модус в применении к философии науки означал бы систематическое выявление утопических допущений, имплицитно присутствующих в научных программах, технологических стратегиях и политиках инноваций. Например, подобным образом можно подвергнуть анализу идеи «открытой науки» (выраженные как в часто используемой модели публикации научных материалов Open Access, так и в маргинальных проектах научного флибустьерства), выявить их взаимосвязь с утопией знания как всеобщего блага (берущую начало еще в утопическом проекте науки Ф. Бэкона), проследить их взаимоотношение и столкновение с коммерческими и политическими интересами. Онтологический модус потребовал бы анализа того, как различные эпистемические практики и институциональные порядки конституируют субъекта познания: какие добродетели (беспристрастность, открытость, критичность) и пороки (догматизм, конформизм, корысть) они культивируют и какую идентичность исследователя они предполагают. Так, можно показать, как современное требование непрерывного воспроизводства научного результата, описываемое афоризмом “publish or perish” («публикуйся или умри») формирует идентичность ученого-предпринимателя, подрывая особое экзистенциальное призвание ученого как искателя истины и «дарителя» знания. Наконец, архитектурный модус побудил бы разрабатывать альтернативные сценарии развития науки и техники, не ограничиваясь простой экстраполяцией текущих трендов. В контексте проблемы «быстрой науки» результатом работы архитектурного модуса может послужить разработка и выдвижение проекта «медленной науки», выступающей альтернативой безумной скорости производства научного знания в ущерб его качеству (см. например: [Касавин, 2026]). Такое исследование позволило бы выполнять одновременно эвристическую, критическую и регулятивную

функции, расширяя пространство рефлексии над проблемами научного знания и функционирования науки как института.

Подобное применение метода IROS в контексте исследования науки можно было бы сравнить, например, с идеями Пола Фейерабенда (методом пролиферации, принципами контриндукции и «все позволено») или онтологическим поворотом. Однако если Фейерабенд призывает к разработке максимального числа альтернативных точек зрения посредством отказа от догматического следования методологическим стандартам и правилам научной рациональности, позволяя ученым использовать любые стратегии для получения жизнеспособных гипотез, IROS предлагает вполне четкую и систематическую процедуру по конструированию и анализу альтернатив. В свою очередь, онтологический поворот, открывающий перспективы для описания множественных миров, все еще не побуждает исследователей разрабатывать миры еще-не-существующие; эту задачу как раз берет на себя «утопический метод».

Переход от проблем демаркации утопии и науки, критики утопических проектов за несостоятельность к «утопическому методу» позволил бы отказаться от вопроса «как защитить науку от утопии?» и использовать утопическое воображение для более глубокой и своевременной рефлексии над научно-технологическим развитием и проблемами ответственного проектирования будущего человечества. К попытке применить предложенный нами подход к анализу реальных научно-технологических утопических проектов мы непременно обратимся в последующих работах.

Список литературы

Блох, 1991 – *Блох Э.* Принцип надежды // Утопия и утопическое мышление: антология зарубежной литературы / Сост., общ. ред. и предисл. В.А. Чаликовой. М.: Прогресс, 1991. С. 49–78.

Герасимова, 2008 – *Герасимова И.А.* Аргументация презентизма // Эпистемология и философия науки. 2008. Т. 16. № 2. С. 188–198.

Касавин, 2026 – *Касавин И.Т.* «Быстрая наука»: свобода и качество исследований сквозь призму публикационной активности // Эпистемология и философия науки. 2026. Т. 63. № 1. С. 6–21. DOI: 10.5840/eps20266311.

Лакатос, 2008 – *Лакатос И.* Фальсификация и методология научно-исследовательских программ // *Лакатос И.* Избранные произведения по философии и методологии науки / Пер. с англ. И.Н. Веселовского, А.Л. Никифорова, В.Н. Поруса. М.: Академический проект; Трикста, 2008. С. 281–475.

Манхейм, 1994 – *Манхейм К.* Идеология и утопия // *Манхейм К.* Диагноз нашего времени. М.: Юрист, 1994. С. 7–276.

Масланов, 2024 – *Масланов Е.В.* Наука и техногенная цивилизация: к вопросу о гуманистическом проекте науки. М.: Русское общество истории и философии науки, 2024.

Паниотова, 2024 – *Паниотова Т.С.* Современные исследования утопии: Новые грани старых проблем // Утопические проекты в истории культуры: материалы V Всероссийской (с международным участием) научной конференции на тему «Социализм между утопией и антиутопией». Ростов-на-Дону; Таганрог: Южный федеральный университет, 2024. С. 5–10.

Поппер, 1992 – *Поппер К.Р.* Открытое общество и его враги: в 2 т. Т. 1: Чары Платона / Под ред. В.Н. Садовского. М.: Феникс: Международный фонд «Культурная инициатива», 1992.

Соколова, 2019 – *Соколова Р.И.* Реабилитация утопии как симптом кризиса российской и западной цивилизаций // Философские науки. 2019. Т. 62. № 5. С. 7–26. DOI: 10.30727/0235-1188-2019-62-5-7-26.

Трунов, 2020 – *Трунов А.А.* Интеллигенция, идеология и утопия в эпистемологическом проекте К. Манхейма // Nomothetika: Философия. Социология. Право. 2020. Т. 45. № 4. С. 711–718. DOI: 10.18413/2712-746X-2020-45-4-711-718.

Халуторных, Егорова, 2022 – *Халуторных О.Н., Егорова Е.А.* Цифровая утопия в современном обществе: социально-философский аспект // Гуманитарный вестник. 2022. Т. 94. № 2. DOI: 10.18698/2306-8477-2022-2-771.

Черткова, 2010 – *Черткова Е.Л.* Наука в контексте утопии: уроки истории // Философия науки и техники. 2010. Т. 15. С. 263–276.

Черткова, 2015 – *Черткова Е.Л.* Утопия как «философский проект» // Вопросы философии. 2015. № 5. С. 190–201.

References

Bloch, E. “Printsip nadezhdy” [The Principle of Hope], in: V.A. Chalikova (ed.). *Utopiya i utopicheskoe myshlenie: antologiya zarubezhnoi literatury* [Utopia and Utopian Thinking: An Anthology of Foreign Literature]. Moscow: Progress Publ., 1991, pp. 49–78. (Trans. into Russian)

Chertkova, E.L. “Nauka v kontekste utopii: uroki istorii” [Science in the Context of Utopia: History Lessons], *Filosofiya nauki i tekhniki*, 2010, vol. 15, pp. 263–276. (In Russian)

Chertkova, E.L. “Utopiya kak ‘filosofskii proekt’” [Utopia as a “Philosophical Project”], *Voprosy filosofii*, 2015, no. 5, pp. 190–201. (In Russian)

Estlund, 2014 – Estlund, D. “Utopophobia”, *Philosophy & Public Affairs*, 2014, vol. 42, no. 2, pp. 113–134.

Geertz, 1973 – Geertz, C. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books, Inc., 1973.

Gerasimova, I.A. “Argumentatsiya prezentizma” [The Argumentation of Presentism], *Epistemology & Philosophy of Science*, 2008, vol. 16, no. 2, pp. 188–198. (In Russian)

Habermas, 1987 – Habermas, J. *The Philosophical Discourse of Modernity: Twelve Lectures*, trans. by Frederick Lawrence. Massachusetts: MIT Press, 1987.

Kasavin, I.T. "Bystraya nauka": svoboda i kachestvo issledovanii skvoz' prizmu publikatsionnoi aktivnosti ["Rapid Science": Freedom and Quality of Research Through the Prism of Publication Activity], *Epistemology & Philosophy of Science*, 2026, vol. 63, no. 1, pp. 6–21. DOI: 10.5840/eps20266311. (In Russian)

Khalutornykh, O.N., and Egorova, E.A. "Tsifrovaya utopiya v sovremennom obshchestve: sotsial'no-filosofskii aspect" [Digital Utopia in Modern Society: A Socio-Philosophical Aspect], *Gumanitarnyi vestnik*, 2022, vol. 94, no. 2. DOI: 10.18698/2306-8477-2022-2-771. (In Russian)

Lakatos, I. "Fal'sifikatsiya i metodologiya nauchno-issledovatel'skikh programm" [Falsification and the Methodology of Scientific Research Programmes], in: I. Lakatos. *Izbrannye proizvedeniya po filosofii i metodologii nauki* [Selected Works on Philosophy and Methodology of Science], trans. by I.N. Veselovskii, A.L. Nikiforov, V.N. Porus. Moscow: Akademicheskii proekt; Triksta, 2008, pp. 281–475. (In Russian)

Levitas, 2013 – Levitas, R. *Utopia as Method: The Imaginary Reconstitution of Society*. London: Palgrave Macmillan, 2013.

Mannheim, K. "Ideologiya i utopiya" [Ideology and Utopia], in: K. Mannheim. *Diagnoz nashego vremeni* [Diagnosis of Our Time]. Moscow: Yurist Publ., 1994, pp. 7–276. (In Russian)

Martin, 2023 – Martin, L. *Alternative Societies: For a Plurist Socialism*. Bristol: Bristol University Press, 2023. DOI: 10.46692/9781529229691.

Maslanov, E.V. *Nauka i tekhnogennaya tsivilizatsiya: k voprosu o gumanisticheskom proekte nauki* [Science and Technogenic Civilization: On the Question of the Humanistic Project of Science]. Moscow: Russian Society for the History and Philosophy of Science Publ., 2024. (In Russian)

Paniotova, T.S. "Sovremennye issledovaniya utopii: novye grani starykh problem" [Contemporary Studies of Utopia: New Facets of Old Problems], in: *Utopicheskie proekty v istorii kul'tury* [Utopian Projects in the History of Culture]. *Proceedings of the 5th National Scientific Conference "Socialism between Utopia and Dystopia"*. Rostov-on-Don; Taganrog: Southern Federal University Publ., 2024, pp. 5–10. (In Russian)

Popper, K.R. *Otkrytoe obshchestvo i ego vragi: v 2 t.* [Open Society and Its Enemies: in 2 vols.], vol. 1: *Chary Platona* [The Spell of Plato], ed. by V.N. Sadovskii. Moscow: Feniks Publ.; Mezhdunarodnyi fond "Kul'turnaya initsiativa", 1992. (Trans. into Russian)

Sand, 2024 – Sand, M. *Technological Utopianism and the Idea of Justice*. Cham: Palgrave Macmillan, 2024.

Sokolova, R.I. "Reabilitatsiya utopii kak simptom krizisa rossiiskoi i zapadnoi tsivilizatsii" [Rehabilitation of Utopia as a Symptom of the Crisis of Russian and Western Civilizations], *Russian Journal of Philosophical Sciences*, 2019, vol. 62, no. 5, pp. 7–26. DOI: 10.30727/0235-1188-2019-62-5-7-26. (In Russian)

Trunov, A.A. "Intelligentsiya, ideologiya i utopiya v epistemologicheskom proekte K. Mankheima" [Intelligentsia, Ideology and Utopia in the Epistemological Project of K. Mannheim], *Nomothetika: Philosophy. Sociology. Law*, 2020, vol. 45, no. 4, pp. 711–718. DOI: 10.18413/2712-746X-2020-45-4-711-718. (In Russian)



Vohnsen, 2024 – Vohnsen, N.H. “Corporate Utopias: The Tech Elit’s Visions and the Global Crises of Democratic Problem-Solving”, *Journal of Business Anthropology*, 2024, vol. 13, no. 2, pp. 79–101. DOI: 10.22439/jba.v13i2.7141.

Wells, 1906 – Wells, H.G. “The So-Called Science of Sociology”, *The Sociological Review*, 1906, vol. 3, no. 1, pp. 357–369.

West, 1996 – West, J. “Utopianism and National Competitiveness in Technology Rhetoric: The Case of Japan’s Information Infrastructure”, *The Information Society*, 1996, vol. 12, no. 3, pp. 251–272. DOI: 10.1080/019722496129459.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 168–188
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 168–188
DOI: https://doi.org/10.5840/eps202663347

Объективность в зеркале языка: от лингвистического анализа к эпистемологическим интерпретациям

Сахарова Анна Владимировна – кандидат филологических наук, научный сотрудник. Институт философии РАН. Российская Федерация, 109240, г. Москва, Гончарная ул., д. 12, стр. 1; e-mail: hanna.lazareva@gmail.com



В статье представлен семантический анализ языковых практик употребления лексемы *объективный* и ее дериватов на материале корпуса естественно-научных статей. Цель исследования – выявить семантические и прагматические особенности лексемы, проанализировать полученные результаты в лингвистическом ключе и интерпретировать полученные результаты с опорой на философские концепции объективности. Лингвистический анализ контекстов употребления рассматриваемых лексем показал, что онтологическое понимание объективности как соответствия реальности сохраняется в научном дискурсе, однако занимает периферийное положение. Доминирующим является методологическое понимание, при котором объективность соотносится с характеристиками методов, процедур и измерительных практик. Дальнейший анализ включал интерпретацию полученных лингвистических данных с опорой на философские концепции объективности, что позволило выявить два аспекта проблематичности понятия. Первый связан с расхождением между пониманием объективности, выраженным в языке научных статей, и характеристиками самого этого языка. Это расхождение формулируется в контексте подхода к объективности Лоррейн Дастон и Питера Галисона. Употребление лексемы *объективный* и ее дериватов отражает идеал «механической объективности», который предполагает инструментальную независимость, формализацию и устранение субъекта. Однако язык научных статей одновременно воплощает нормативный идеал «тренированного суждения», включающий интерпретацию, оценку и риторическую организацию материала. Это может свидетельствовать о том, что постепенный переход между указанными эпистемическими нормами осуществляется, но не осознается самим научным сообществом. Второе расхождение возникает между интерпретацией объективности, фиксируемой в текстах научных статей, и ее трактовкой в философской литературе. В текстах научных статей объективность может быть интерпретирована в онтологическом и методологическом смысле. В философии науки, наряду с этими двумя типами, выделяется также интерсубъективное понимание объективности, предполагающее трактовку объективности как социальной разделяемости и воспроизводимости внутри сообщества. Третий тип не находит отражения в языковых практиках, однако является ключевым для ряда направлений философии науки, в частности для последователей «сильной программы». Возможно, именно это расхождение в понимании объективности и обусловило остроту полемики «научных войн». Также это расхождение ставит вопрос о степени соотнесенности философских концептуализаций науки с реальными дискурсивными формами научной деятельности.

Ключевые слова: объективность, семантика, прагматика, эпистемология, сильная программа, логический анализ языка, философия науки

© Сахарова А.В., 2026

OBJECTIVITY IN THE MIRROR OF LANGUAGE: FROM LINGUISTIC ANALYSIS TO EPISTEMOLOGICAL INTERPRETATIONS

Anna V. Sakharova –
PhD in Philology,
Research Fellow.
Institute of Philosophy,
Russian Academy of Sciences.
12/1 Goncharnaya St.,
Moscow 109240,
Russian Federation;
e-mail: hanna.lazareva@gmail.com

The article presents a semantic analysis of linguistic practices in the use of the lexeme *objective* and its derivatives, based on a corpus of natural science articles. The aim of the study is to identify the semantic and pragmatic features of the lexeme, analyze the obtained results from a linguistic perspective, and interpret these results with reference to philosophical concepts of objectivity. A linguistic analysis of the contexts in which the lexemes in question are used shows that the ontological understanding of objectivity as correspondence to reality is preserved in scientific discourse, but occupies a peripheral position. The dominant understanding is methodological, in which objectivity is associated with the characteristics of methods, procedures, and measurement practices. Further analysis involved interpreting the linguistic data obtained with reference to philosophical concepts of objectivity, which made it possible to identify two aspects of the problematic nature of the concept. The first is related to the discrepancy between the understanding of objectivity expressed in the language of scientific articles and the characteristics of that language itself. This discrepancy is formulated within the framework of Lorraine Daston and Peter Galison's approach to objectivity. The use of the lexeme *objective* and its derivatives reflects the ideal of "mechanical objectivity", which presupposes instrumental independence, formalization, and the elimination of the subject. However, the language of scientific articles simultaneously embodies the normative ideal of "trained judgment", which includes interpretation, evaluation, and the rhetorical organization of material. This may indicate that a gradual transition between these epistemic norms is taking place, but is not recognized by the scientific community itself. The second discrepancy arises between the interpretation of objectivity recorded in the texts of scientific articles and its treatment in the philosophical literature. In the texts of scientific articles, objectivity can be interpreted in an ontological and methodological sense. In the philosophy of science, alongside these two types, an intersubjective understanding of objectivity is also distinguished, which entails interpreting objectivity as social shareability and reproducibility within a community. The third type is not reflected in linguistic practices, yet it is crucial for a number of trends in the philosophy of science, in particular for adherents of the "strong program". It is possibly this discrepancy in the understanding of objectivity that gave rise to the intensity of the "science wars" polemics. This discrepancy also raises the question of the degree of correspondence between philosophical conceptualizations of science and the actual discursive forms of scientific activity.

Keywords: objectivity, semantics, pragmatics, epistemology, Strong Programme, logical analysis of language, philosophy of science

Введение

Понятие объективности традиционно рассматривается как ключевая категория и фундаментальная ценность науки. Однако его кажущаяся самоочевидность скрывает в себе внутреннюю неоднозначность и проблематичность. Проблематичность объективности во многом обусловлена тем, что данное понятие функционирует на пересечении различных типов дискурсов и форм познания – обыденного, научного и философского. Повседневный и научный образы мира, как правило, противопоставляются: обыденный дискурс отражает непосредственный опыт, тогда как научный язык фиксирует результаты специализированного познания: объективное в этих контекстах может значить разное. Когда физик, философ и человек, не связанный с наукой, говорят об объективности, за этим стоят не просто различные определения, но различные способы концептуализации мира. В результате объективность оказывается семантически и прагматически неоднородным понятием, особенно на стыке дискурсов и в ситуациях, когда один дискурс оказывает влияние на другой. Интуитивные определения – как отсутствие субъективности или соответствие реальности – оказываются лишь приближением, поскольку «для большого класса случаев – хотя и не для всех, – в которых мы используем слово “значение”, его можно определить так: значение слова – это его употребление в языке» [Витгенштейн, 2018, с. 43], т.е. конкретное содержание понятия определяется контекстом. Мы не можем сослаться и на словарные определения: определения из толкового и философского словаря будут разными, а словарь языка науки отсутствует как вид. Дополнительную сложность вносит разнообразие трактовок объективности внутри указанных дискурсов. Например, внутри философии науки это понятие становится проблемным даже в тех случаях, когда, не находясь в фокусе внимания, оно лежит в основании теоретических позиций, как, например, в дискуссиях, известных как «научные войны».

По-разному, вероятно, понятие объективности может осмысляться и внутри научного дискурса. Однако исследователи особенно в последние десятилетия редко делают само понятие объективности предметом явного эпистемологического анализа и обсуждения. Как отмечает Стивен Шейпин, практикующие ученые склонны дистанцироваться от эпистемологических обсуждений, воспринимая их как внешние по отношению к научной деятельности: «Практикующие ученые, по-видимому, почти не обращают внимания на дебаты или высказывания эпистемологов. Действительно, в естественных науках эпистемологические и метанаучные размышления, а также социологические размышления часто воспринимаются как знак заслуженного положения. Иногда их прямо осуждают как неуместные,



непрофессиональные или даже патологические: может быть, это то, что делают философы или социологи, но (как говорят) это не то, чем должны заниматься практикующие ученые» [Shapin, 2015, p. 674].

Перспективным для прояснения позиции самих ученых по отношению к понятию объективности и ее смыслам представляется рассмотрение его функционирования в научной практике – там, где ученые, обеспокоенные статусом своих объектов или результатами работы, так или иначе воспроизводят это понятие, не нуждаясь в его рефлексии. Иными словами, отношение ученых к объективности может выражаться не только в явных высказываниях о ней, но и в практиках исследовательской деятельности (на что, например, указывает [Соколова, 2022, с. 30]). Такое «неявное» присутствие объективности в языке науки позволяет поставить вопрос о том, является ли ее концептуальное прояснение необходимым условием научного мышления или же она функционирует как имплицитная установка, не требующая постоянной вербализации: «Не следует ли перестать заботиться о смысле этого концепта и предоставить объективности возможность скрытого присутствия в научных практиках и экспертной деятельности ученых» [Шиповалова, 2025, с. 7].

Но как получить доступ к этому понятию в ситуации, когда исследователи не формулируют его эксплицитно? Источники могут быть различными, однако наиболее распространенным подходом являются case studies. Вместе с тем они характеризуются высокой степенью интерпретативности и не обладают столь же явно фиксированным эмпирическим основанием, как лингвистические исследования, основанные на систематическом анализе языковых данных.

В качестве дополнительного источника, дополняющего «кейсы» («ситуации»), могут выступать тексты научных статей, поскольку они отражают имплицитные установки и неявное знание авторов естественно-научных исследований [Полани, 1995]. В этом случае текст рассматривается не только как носитель содержания, но и как материал для реконструкции стоящих за ним когнитивных структур. При этом лингвистический анализ может использоваться как самостоятельный метод изучения научных практик, а также как средство уточнения и эмпирической конкретизации результатов, полученных в рамках case studies.

Гипотеза

В основе исследования лежит предположение о том, что анализ языковых практик использования лексемы *объективный* и производных от нее (*объективность*, *объективизировать*) как языковых репрезентаций концепта «объективность» в естественно-научных текстах

позволяет выявить стоящие за ними имплицитные когнитивные практики – т.е. не проговариваемые прямо, но регулярно воспроизводимые способы концептуализации научного знания, его границ и оснований достоверности. Этот анализ может быть полезен для философских исследований научных практик, поскольку он позволяет реконструировать неявные эпистемические установки, регулирующие деятельность ученых, и уточнить, каким образом в реальной научной коммуникации присутствует представление об объективности. Тем самым лингвистический анализ выступает как инструмент доступа к «внутренней» перспективе науки, позволяя косвенно зафиксировать позицию самих исследователей без обращения к их эксплицитной методологической рефлексии. Понимание того, как эти когнитивные практики проявляются в языке, становится особенно значимым в сравнении с философскими моделями объективности, которые описывают нормативные представления о научном знании. Сопоставление выявленных языковых практик с философскими подходами позволят не только увидеть, где фактические способы концептуализации совпадают с философской рефлексией, но и выявить расхождения, а также проанализировать их причины и значение для понимания научной практики.

Материал и методы исследования

В настоящем исследовании рассматривается синхронический срез представлений об объективности в начале XXI в., реконструируемый на материале научных статей авторов естественно-научных и технических специальностей. Анализ основан на семантическом исследовании употреблений лексемы *объективный* и производных от нее (*объективность*, *объективизировать*), репрезентирующих концепт «объективность» в естественно-научных текстах.

Материалом исследования послужила выборка из 472 русскоязычных научных статей по физике, механике, материаловедению и металлургии (1999–2020 гг.), размещенных в электронных библиотеках eLIBRARY.RU и CyberLeninka. Все тексты имеют ненулевую цитируемость, что рассматривается как формальный показатель включенности в научную коммуникацию. Общий объем выборки составил 3 118 995 словоформ.

Довольно часто лингвистическому анализу подвергаются нормативные установки научного текста – прежде всего в контексте образовательной деятельности и обучения научному письму (см., например, [Lederman, Zeidler, 2023]). Однако сами эти установки являются частью более широкой риторической практики науки. Как отмечается в [Zerbe et al., 2025], сегодня это направление переживает двойной

кризис: наука утрачивает былое риторическое доминирование, а сама риторика науки остается узкоакадемической дисциплиной. В этих условиях особенно важным становится анализ того, как научные тексты используют риторические ресурсы для конструирования своих ключевых эпистемических категорий, что требует привлечения качественно ориентированных методов анализа языкового материала, основанных на интерпретации контекстов употребления. Для анализа объективности на русскоязычном материале наиболее адекватной представляется методология, разработанная в рамках направления «Логический анализ языка» [Арутюнова, 1988; Арутюнова, 1999; Радбиль, 2016; Радбиль, 2017]. В рамках этого подхода язык рассматривается не только как средство выражения уже готовых смыслов, но как система, в которой закреплены способы концептуализации действительности, в том числе – эпистемические и оценочные структуры. Соответственно, анализ лексемы *объективный* и ее дериватов предполагает фиксацию частотности и контекстов употребления, а также реконструкцию контекстуальных семантических оппозиций и сходств, в которые она встроена (например, объективное vs субъективное, объективное = естественное).

Лингвистический компонент исследования основывается на выявлении и типизации контекстов с указанными лексемами и их семантической интерпретации с опорой на идею о том, что «само собой разумеющееся» в языке представляет наибольшую аналитическую ценность: «Важно не то, что утверждают носители языка, а то, что они считают само собою разумеющимся, не видя необходимости специально останавливать на этом внимание» [Радбиль, 2017, с. 34]. В частности, после выделения контекстов и их типизации предполагается описание выявленных семантических признаков (например, независимость от субъекта, процедурная обоснованность, соответствие реальности), анализ деривационных (производных) значений, в которых происходит смещение или расширение исходной семантики и учет прагматических функций высказывания, в частности риторических употреблений, направленных на прагматическое воздействие на адресата.

Лингвисты группы «Логический анализ языка» не обращались к анализу представлений об объективности в языке, однако они подробно рассматривали вопрос о языковой рецепции истины и истинного [Логический анализ языка, 1991; Логический анализ языка, 1995]. В частности, особенности семантики лексем *истинный* и *ложный* анализировались Галиной Ивановной Кустовой [Кустова, 2008], метаязыковые показатели со значением истинности – Тимуром Беньюминовичем Радбилем [Радбиль, 2016], а вопросы концептуального анализа лексических показателей истинности – Ниной Давидовной Арутюновой [Арутюнова, 1991]. В результате исследований концептов истины и истинного была выработана лингвистическая

методология, применимая и к анализу лексемы *объективный* и ее дериватов, хотя сам концепт, как мы уже указали выше, специально не рассматривался.

В аспекте намеченного сопоставления философских подходов с позицией ученого, выраженной в текстах, наиболее значимыми представляются работы Лоррейн Дастон и Питера Галисона, представителей «сильной программы» (в частности, Дэвида Блура), а также исследования Лады Владимировны Шиповаловой.

В работе Л. Дастон и П. Галисона [Дастон, Галисон, 2018] «механическая объективность» (исключение субъекта) рассматривается не как универсальное и неизменное свойство научного знания, а как исторически изменчивая, существующая наряду с другими эпистемическими добродетелями и идеалами научного познания: «истиной-по-природе» (идеализация сущности) и «тренированным суждением» (экспертная интерпретация). Смена доминирующих эпистемических добродетелей отражает изменение представлений о роли субъекта в познании – от центральной роли ученого-гения – к его радикальному устранению – и далее к его дисциплинированному включению в виде обученного интерпретатора. Эта модель задает рамку для анализа внутреннего напряжения между декларацией объективности и фактическими практиками научного письма.

В философских подходах к объективности фиксируется ее многоуровневый характер, включающий онтологические, методологические и интерсубъективные измерения, что системно эксплицировано в работах Л.В. Шиповаловой, выделяющей три основных значения: негативное (как независимость от субъекта), позитивное онтологическое (как соответствие реальности) и интерсубъективное (как социальная значимость, разделяемость и устойчивость знания)» [Шиповалова, 2014, с. 307]. В рамках этого третьего уровня особое внимание получает «сильная программа» Д. Блура [Bloor, 1984], в которой объективность трактуется как результат интерсубъективных процессов: объективное знание понимается не как отражение независимой реальности и не как следствие устранения субъекта, а как эффект функционирования коллективных норм, конвенций и механизмов согласования внутри научного сообщества, благодаря которым обеспечивается социальная устойчивость и воспроизводимость убеждений.

Результаты лингвистического исследования

В исследованном материале выявлено 55 вхождений лексемы *объективный* и ее дериватов, распределенных по 17 текстам, что составляет 3,6% от общего числа проанализированных статей. Количество



употреблений лексемы *объективный* и ее дериватов на один миллион слов в корпусе текстов: IPM (instances per million) = 17,6.

Контексты употребления указанных лексем в научных текстах могут быть систематизированы по двум основным типам: **онтологическому** (контексты, характеризующие реальность и соответствие информации реальности) и **методологическому** (контексты, говорящие об объективности метода и условиях получения знания).

Таблица 1. Количественное распределение лексемы *объективный* и ее дериватов по типам

	Количество вхождений	Количество текстов
Онтологический тип	18	8
Методологический тип	37	12

Онтологический тип контекстов

Онтологический тип контекстов представлен 18 вхождениями в 8 из рассмотренных текстов. В рамках этого типа можно выделить два значения лексемы *объективный* и ее производных.

Первое – собственно онтологическое: в этих контекстах объективность выступает как дескриптивная характеристика самой реальности. Она понимается как существование «само по себе», независимое от того, фиксирует ли это состояние или процесс наблюдатель. Данный тип контекстов встречается в 9 случаях, из которых два имеют терминологический характер, один может быть интерпретирован как риторический, а один демонстрирует деривацию значения. Примеры такого употребления включают утверждения: «Это состояние **объективно** (здесь и далее выделено мной. – *А.С.*) и не зависит от наблюдателя и от процесса наблюдения» [Жмудь, 2015, с. 107] и «Тем самым автор парадокса признает, что в любой момент до наблюдения кот либо **объективно** жив, либо **объективно** мертв. То есть состояние кота может быть **объективно** лишь одним из двух, а никак не смешением двух» [Там же, с. 108]. Несмотря на невысокую частотность, данный семантический тип кажется нам важным для анализа, поскольку отражает интерпретацию объективного как соответствующего реальному положению вещей. В этом онтологическом понимании реальность не допускает степеней объективности, в отличие от эпистемологических контекстов (см. ниже), где возможна сравнительная оценка степени объективности. Так, кот либо объективно жив, либо объективно мертв; понятие «более объективно живой» здесь не имеет смысла.

Пример деривации этого типа значения связан с противопоставлением «объективная эволюция» и «субъективное модифицирование»: «Это может быть результатом **объективной эволюции** и/или **субъективного модифицирования** химического, минералогического, дисперсного и морфологического состава цементного камня» [Баженов, 2014, с. 8]. В данном контексте слово *объективный* используется как синоним естественности, отсутствия вмешательства человека или субъекта. Противопоставленный объективному «субъективно модифицированный состав», напротив, указывает на изменение, внесенное деятельностью человека.

Терминологическое употребление в контекстах «*покоящаяся объективно система отсчета*» и «*объективно существующее понятие покоя*» фиксируют соответствие реальности – но не физической, а теоретической и математической, идеализированной, в которой состояния покоя и движения объектов фиксированы независимо от наблюдателя.

Во втором типе онтологического значения лексемы *объективный* и ее производных в центре внимания оказывается не сама реальность как таковая, а соответствие знания (информации, данных) реальности. В этом случае свойство объективности приписывается уже полученному знанию: «Долг же научной общественности состоит в своевременном и по возможности наиболее **объективном** информировании населения» [Гичев, 1999, с. 67] – здесь акцент делается на том, чтобы передаваемая информация соответствовала реальному положению вещей, вне зависимости от методов ее получения и передачи.

Аналогично, выражение: «Это объясняется целым рядом **объективных** факторов» [Леонов, 2018, с. 104–105], – фиксирует факторы, которые, по мысли автора, соответствуют действительности. В контексте: «С момента появления первых результатов, свидетельствующих о влиянии ЭМП на состояние здоровья человека, в научной литературе не утихает полемика о достоверности и **объективности** полученных выводов» [Гичев, 1999, с. 20], – автор усиливает акцент на объективности, ставя ее в один синонимический ряд с достоверностью, т.е. соответствием информации реальности.

В обоих онтологических значениях встречаются риторические употребления лексем, в которых апелляция к объективности выполняет функцию усиления убедительности высказывания и не соотносится с верифицируемыми характеристиками знания или реальности. Риторические употребления относительно немногочисленны: 6 случаев в 5 текстах. В целом их можно назвать деривациями онтологического типа контекстов, поскольку апелляция происходит именно к реальности как к самому сильному риторическому аргументу: «Апелляция к реальной структуре мира именно потому и является максимально эффективным средством языкового воздействия –



трудно что-то противопоставить неотвратимым законам мироздания» [Радбиль, 2017, с. 267]. Апелляция к объективности придает утверждению дополнительную убедительность, например, в следующем контексте: «Дискретные модели, и **это объективность,** в недалеком будущем будут в количественном отношении значительно превалировать над непрерывными моделями» [Морозов, 2014, с. 184]. Риторическое употребление в данном случае связано с апелляцией к будущему состоянию науки («это завтрашний день»), что выводит высказывание за пределы собственно объективного описания. Объективность предполагает фиксацию уже существующих состояний, свойств или закономерностей, тогда как высказывания о будущем носят прогностический характер и неизбежно включают элементы оценки, предположения и теоретического моделирования. В силу этого объективность не может в строгом смысле приписываться будущим состояниям, поскольку они еще не реализованы и не поддаются непосредственной верификации.

Аналогично риторический потенциал фраз «объективные причины», «объективные факторы» также обусловлен их неверифицируемой оценочностью. Эти факторы и причины опираются на субъективную оценку автора. Например, в следующем контексте речь идет об объективных причинах, связанных с отсутствием комплексного подхода (а кто мешает?) и о недостаточном (для кого? Относительно чего?) финансировании: «Отставание отечественных разработок ВОД объясняется рядом **объективных причин**, среди которых определяющей является отсутствие комплексного подхода к решению этой проблемы и, как следствие, недостаточное финансирование этой области контрольно-измерительной техники» [Гармаш, 2005, с. 131].

Также в какой-то мере тавтологичное использование фразы «объективная реальность» служит скорее для усиления убедительности, чем для какого-то именно прояснения вопроса. Как, например, в следующем контексте присутствует риторическое усиление за счет противопоставления научного исследования и произвольного «рисования дендрограмм»: «…автоматизированная система кластеризации превращается в этих условиях из инструмента исследования структуры **объективной реальности** в автоматизированный инструмент рисования таких дендрограмм, какие больше нравятся пользователю» [Луценко, 2011, с. 5].

Примечательно, что риторические контексты размывают противопоставление объективного и субъективного: «Значительная доля остающегося недоумения обусловлена не только **объективными причинами**, связанными с недостатком необходимых сведений, но и **субъективными факторами**, в частности, широким распространением так называемой стандартной модели Солнца» [Веселовский, 2008]. Насколько причины, связанные с недостатком необходимых

сведений, более объективны, чем факторы, связанные с широким распространением так называемой стандартной модели Солнца?

Методологический тип контекстов

Методологический тип контекстов является доминирующим: зафиксировано 37 вхождений в 12 текстах.

Объективность в методологическом типе контекстов выступает прежде всего как характеристика способа получения знания, т.е. как свойство используемых методов, процедур и измерительных практик. В этом случае она соотносится не с онтологическим статусом изучаемых явлений, а с процедурными параметрами исследования. Акцент переносится на процесс получения данных: объективными называются методы, оценки, анализы и процедуры постольку, поскольку они обеспечивают формализуемый, проверяемый и независимый от индивидуальных особенностей исследователя результат.

Например, в контексте: «…отсутствие **объективных** методов контроля за результатами облучения, которые позволили бы давать оптимальную дозу облучения в каждом конкретном случае» [Владимиров, 1999, с. 2–3], – объективное прямо связывается с наличием или отсутствием инструментально и процедурно организованных способов фиксации объема воздействия и его результатов. Речь идет не о том, что сами эффекты лазерного воздействия являются или не являются «объективными» в онтологическом смысле, а о невозможности их надежной регистрации и количественной оценки. Тем самым критерий объективности локализуется в процедурах контроля и измерения, а не в соотнесенности знания с независимой реальностью.

В ряде контекстов методологическое понимание объективности видоизменяется и уточняется: объективное развивает более узкое значение, синонимичное инструментально и количественно измеренному, а также научному (также понятному инструментально). Например, в тексте [Коренбаум, 2004] *объективное* понимается как измеренное с использованием инструментальных, формализованных и, как правило, количественно исчислимых методов. Речь идет об измерении акустических характеристик при болезнях легких, которое обычно делается врачом «на слух»: «Аускультация легких, т.е. искусство выслушивания дыхательных и голосовых звуков, проведенных на стенку грудной клетки, имеет почти 200-летнюю медицинскую историю. Тем не менее до сих пор не удавалось **объективизировать** эту диагностическую процедуру, создать надежные акустические технологии, которые могли бы найти применение в современной медицинской практике» [Там же, с. 68]. Далее противопоставляются



два типа знания: субъективное, основанное на индивидуальном опыте врача («врачебное искусство»), и объективное – инструментально опосредованное, допускающее количественную фиксацию и воспроизводимость результатов («объективная наука»): «…эта диагностическая процедура остается более врачебным искусством, чем **объективной** наукой…» [Коренбаум, 2004, с. 69].

В тексте [Гичев, 1999] присутствует противопоставление субъективных способов измерения электромагнитных полей, имевшихся в прошлом (в статье упоминаются данные опросов), прямому, инструментальному измерению: «Совместить в себе **объективность** прямой дозиметрии и возможность применения для оценки экспозиции ЭМП, имевшейся в прошлом, был призван метод вычисления экспозиционной дозы, введенный в практику исследований с начала 90-х гг.» [Там же, с. 25]. Далее объективность ставится в один ряд с научной обоснованностью: «Таким образом, вопросы стандартизации методов исследования и выбора наиболее **объективных и научно обоснованных** критериев величины воздействия ЭМП на человека остаются на настоящее время определяющими для дальнейшего развития знаний в этой области» [Там же, с. 25–26].

Объективность в методологическом типе контекстов также закрепляется в составе специализированной терминологии: всего обнаружено 7 употреблений в 2 текстах. Так, выражение *объективная экстраполяция* [Сандуляк, 2010] обозначает процедуру, основанную на экспериментально установленной и теоретически обоснованной закономерности, при которой параметры выводятся из эмпирических данных и допускают оценку погрешности. Аналогичным образом, *объективная производная* в механике сплошных сред определяется как величина, не зависящая от выбора системы отсчета [Роговой, 2005].

Обсуждение

Анализ языковых практик позволяет реконструировать ряд имплицитных когнитивных установок, регулирующих научную деятельность. Проблематичный характер объективности выявляется непосредственно на языковом материале: по текстам можно реконструировать установку на объективность как на проблематичную, дефицитную характеристику, понимание и реализация которой требует методологических усилий. В относительно небольшой, но концептуально значимой группе контекстов объективность представлена как отсутствующая, дефицитная или проблематичная; всего выявлено 8 таких употреблений. В этих примерах объективность не выступает как данное или самоочевидное свойство знания, а, напротив, подвергается

проблематизации через указание на отсутствие необходимых параметров и методов, недостаточную разработанность процедур, ограниченность или несовершенство применяемых подходов, а также через сомнения в достоверности получаемых результатов: «…**объективная** сравнительная **оценка** возможности методик **затруднена** ввиду отсутствия во многих публикациях необходимых химикоаналитических и метрологических параметров» [Базель, 2012, с. 578], «Методологические неточности, связанные, в первую очередь, с **проблемой объективной оценки** экспозиционной дозы» [Гичев, 1999, с. 47], «Традиционный физико-математический подход **не всегда обеспечивает объективную** регистрацию динамики развития заболеваний» [Пятин, 2018, с. 147].

Эта проблематичность и напряженность в употреблении понятия объективности проявляется не как логическое противоречие в строгом смысле, а как различие уровней его функционирования в научном дискурсе. С одной стороны, фиксируется несоответствие между эксплицитными декларациями объективности и ее отсутствием на уровне риторики научного дискурса; с другой – возникают затруднения при попытке реконструировать само содержание понятия объективности на основе эмпирического материала и соотнести его с философскими интерпретациями, что связано с конкуренцией различных контекстуальных значений рассматриваемого понятия.

Объективность в языке науки и объективность языка науки – разные режимы

Традиционная концепция наивного реализма, предполагающая прямой доступ ученого к фактам окружающей действительности, довольно редко (см. онтологические контексты) проявляется в современных научных текстах. Напротив, наиболее распространены методологические контексты, в которых объективность понимается прежде всего как характеристика методов исследования и процедур получения знания, а не как свойство самой реальности. Речь идет о способах, с помощью которых знание стремится быть объективным, а не о непосредственной фиксации фактов, существующих вне наблюдателя. Таким образом, объективность выражена в языке науки не как прямо соответствующее внешнему миру, а как результат корректно организованной исследовательской процедуры, обеспечивающей надежность полученных данных. Это смещение акцента с онтологического статуса знания на условия его получения закономерно ведет к приоритету измеримости как ключевого критерия объективности.



Эта установка проявляется в методологических контекстах: объективность систематически связывается с измерительными, количественно исчислимыми и инструментальными способами получения знания. Особый акцент делается на использовании приборов, стандартизированных методик и формализованных процедур обработки данных. Мы наблюдем семантическую деривацию лексемы *объективный* и приобретение ей дополнительного «инструментального» оттенка значения: она понимается как эффект применения определенных технических и методологических средств. Предельным выражением такого подхода становится идеал «механической фиксации» результатов наблюдений или экспериментов, при которой участие субъекта должно быть сведено к минимуму, а наблюдение – передано приборам и измерительным устройствам.

В теоретическом плане этот идеал соотносится с подвидом механической объективности в концепции, предложенной Л. Дастон и П. Галисоном. Как мы уже отмечали выше, в рамках их подхода объективность (формализованная и инструментально обеспеченная регистрация без участия субъекта) рассматривается не как универсальное и неизменное свойство научного знания, а как исторически изменчивая эпистемическая добродетель, наряду с «истиной-по-природе» (идеализацией сущности) и «тренированным суждением» (экспертной интерпретацией). Переход между этими формами отражает изменение представлений о роли субъекта в познании – от центральной роли ученого-гения («истина-по-природе») – к его радикальному устранению («механическая объективность») – и далее к его дисциплинированному включению в виде обученного интерпретатора («тренированное суждение»).

Эксплицитные выражения объективности, которые мы анализировали – методологический тип, как раз соответствуют «механической объективности», поскольку делают акцент на инструментальность, исчислимость и процедурную строгость, предполагая минимизацию или устранение субъективного вклада.

Однако на уровне реального функционирования научного текста эта картина существенно усложняется. Даже в тех случаях, когда речь идет об «объективных» результатах, язык научных статей демонстрирует наличие интерпретативных стратегий, риторических приемов (даже в онтологическом типе контекстов) и эпистемических оценок. Как показывают исследования научного дискурса, в нем регулярно присутствуют способы организации материала, которые не сводятся к фиксации данных, а включают их отбор, акцентирование и интерпретацию (см. подробный анализ в [Сахарова, 2025]). В этом смысле научный текст оказывается не прозрачным отражением фактов, а сложной дискурсивной практикой, в которой субъект не устраняется, а лишь перераспределяется по различным уровням текста – от явных оценок до скрытых риторических структур.

Именно это соответствует тому, что Л. Дастон и П. Галисон описывают как режим «тренированного суждения»: здесь объективность предполагает не устранение субъекта, а его профессионально сформированную способность к интерпретации. Как отмечают авторы, эксперт «способен научиться – читать, интерпретировать, извлекать заметные и значимые структуры из мешанины неинтересных артефактов и фона» [Дастон, Галисон, 2018, с. 456]. Тем самым научный текст фиксирует не только результаты наблюдения, но и сам процесс их интерпретации, осуществляемый исследователем.

Тем самым в языке научных статей возникает внутреннее напряжение: на уровне явных деклараций воспроизводится идеал механической объективности, тогда как на уровне фактической организации текста имплицитно реализуется логика тренированного суждения. Это расхождение указывает на внутреннюю неоднородность понятия объективности и демонстрирует, что оно функционирует одновременно в нескольких несводимых друг к другу эпистемических режимах.

В более широком смысле это позволяет предположить, что между данными режимами – от механической объективности к тренированному суждению – постепенно происходит переход, затрагивающий не только теоретические модели науки, но и реальные языковые практики ее описания. Однако этот процесс, по-видимому, не фиксируется и не рефлексируется самим научным сообществом в явной форме: смена эпистемических установок осуществляется на уровне практик и дискурсивных привычек, оставаясь скрытой за сохранением традиционных формулировок и нормативных деклараций об инструментальной и механической объективности научного знания.

Три значения объективности

Проблема объективности научного знания во многом связана с неоднозначностью самого термина. В исследовании Л.В. Шиповаловой дается подробный обзор различных подходов объективности в философской литературе и фиксируется по меньшей мере три различных семантических компонента понятия объективности: объективность как соответствие реальности; как независимость от субъекта, достигаемая через процедуры и методы; и как разделяемость и социальная устойчивость знания [Шиповалова, 2014, с. 307].

Если соотнести эти философские различения с языковыми практиками науки, то получается, что первые два значения представлены в нем достаточно отчетливо. С одной стороны, хоть и в редких случаях, но сохраняется и ориентация на реальность: научные



высказывания претендуют на описание внешнего мира и тем самым апеллируют к онтологическому значению объективности как соответствия объекту. С другой стороны, объективность связывается с характеристиками исследовательских процедур – воспроизводимостью, стандартизированностью, контролируемостью, что соответствует методологическому пониманию.

Третье значение объективности, присутствующее в философской литературе, – интерсубъективное – наиболее последовательно разработано в рамках «сильной программы» социологии знания, прежде всего у Д. Блура: «Я полагаю, что объективное убеждение – это убеждение, которое не принадлежит ни одному отдельному человеку. Оно не колеблется, как субъективное состояние или личное предпочтение. Оно не мое и не ваше, но им можно поделиться. Оно имеет внешний, предметный аспект» [Bloor, 1984, p. 229]. В данной трактовке объективность связывается не с соответствием реальности и не с устранением субъекта, а с социальной разделяемостью и устойчивостью знания внутри научного сообщества.

Именно это значение практически не обнаруживается в анализируемых языковых практиках научных текстов. Его отсутствие позволяет частично объяснить напряжение между представителями естественных наук и сторонниками «сильной программы», проявившееся в ходе так называемых «научных войн». Учеными объективность мыслится в методологическом и онтологическом ключе – как процедурная надежность и соответствие реальности. Интерсубъективная трактовка, сводящая объективность к социальным механизмам согласия, оказывается для исследователей естественно-научных специальностей чуждой, поскольку не находит опоры в их собственных языковых практиках. Именно это расхождение в понимании объективности, возможно, и обусловило резкость полемики: речь шла не просто о разных теориях науки, а о *различных семантических режимах одного и того же ключевого понятия*. Это подтверждают и выводы исследования Л.В. Шиповаловой о том, что именно конфликт двух основных интерпретаций научной объективности является основанием современного противостояния эпистемологического рационализма (реализма) и релятивизма (конструктивизма) как двух подходов к исследованию научной деятельности [Шиповалова, 2014, с. 308].

Таким образом, три значения объективности – онтологическое, методологическое и интерсубъективное – не только концептуально различаются, но и неравномерно представлены в языке науки и философской рефлексии. Если первые два активно функционируют и структурируют научный дискурс, то третье остается внешним по отношению к нему, что и создает предпосылки для теоретических и методологических конфликтов.

В более общем плане это ставит вопрос о статусе ряда философских концепций науки (например, вышеупомянутой «сильной программы»), которые, несмотря на заявленную опору на анализ научной практики, фактически формируются на основе внешней по отношению к самой научной практике интерпретации объективности. В результате возникает ситуация, при которой теоретические модели описывают науку как целостный и концептуально упорядоченный объект, тогда как сами практики науки (в данном случае – языковые) демонстрируют более фрагментарную, контекстно зависимую и внутренне неоднородную организацию. Отсюда следует потенциальный разрыв между философской реконструкцией науки и тем, как научное сообщество фактически использует и осмысляет собственные базовые понятия, например в данном случае понятие объективности. В результате философская реконструкция может фиксировать скорее нормативные или идеализированные формы научной практики, а не их дескриптивные характеристики, что приводит к смещению от описания науки «как она есть» к ее концептуальной модели «как она должна быть» в рамках той или иной теоретической перспективы.

Выводы

Полученные результаты позволяют уточнить статус объективности в современной научной практике, сместив фокус с ее философского осмысления на реальные способы функционирования в языке науки. Относительно невысокая частотность лексемы *объективный* и производных от нее при их высокой нормативной значимости свидетельствует о том, что объективность не является предметом явной рефлексии, а функционирует как имплицитная эпистемическая установка, встроенная в научную коммуникацию.

Проведенный анализ позволил уточнить распределение различных значений объективности в научном дискурсе: онтологическое и методологическое значения устойчиво выявляются на уровне языковых практик. Доминирование методологического типа употреблений показывает, что в языковой рецепции ученых объективность в первую очередь соотносится не с онтологическим статусом реальности, а с характеристиками исследовательских процедур и результатов. При этом выявлены семантические расширения лексемы: объективность понимается из контекста как «научность», «инструментальность», «исчислимость», «естественность». Объективность в этих случаях перестает быть отдельной эпистемической характеристикой и начинает выполнять роль обобщенного маркера легитимного знания, происходит семантическое «растворение» объективности



в комплексе требований к научному знанию, при котором она уже не противопоставляется другим его свойствам, а выступает их интегральным обозначением.

Онтологический тип контекстов демонстрирует, что значение объективности как характеристики реальности или соответствия ей сохраняется в научном языке, но занимает периферийное положение и реализуется в ограниченном числе случаев. При этом внутри данного типа объективность как характеристика может относиться как к самой реальности, так и к знанию о ней – в смысле его соответствия действительному положению вещей. Из онтологического значения «вырастают» риторические употребления, что демонстрирует функцию объективности как ценностного маркера, обеспечивающего легитимацию высказываний. В этих случаях она утрачивает строгую определенность и выступает как элемент научного дискурса, направленный на усиление убедительности.

Представленное в философской (и социологической) рефлексии науки интерсубъективное измерение объективности отсутствует в научных текстах. Это расхождение смысла объективности между наукой и философией науки имеет принципиальное значение, поскольку указывает на существование напряжения между различными режимами понимания объективности: с одной стороны, закрепившимися в научной коммуникации представлениями о соответствии реальности и процедурной надежности, с другой – теоретическими моделями, акцентирующими социальную природу знания. Именно эта асимметрия приводит к напряжению между различными философскими интерпретациями науки, поскольку разные подходы акцентируют различные аспекты объективности, тем самым формируя не только конкурирующие теоретические модели, но и несовпадающие семантические режимы понимания природы научного знания.

Неоднозначный характер объективности может быть обнаружен не только на границе между наукой и ее философскими описаниями, но и внутри самой научной практики. На уровне прямого высказывания научный текст воспроизводит установку на механическую объективность (в терминологии Л. Дастон и П. Галисона) – через апелляцию к воспроизводимости, стандартизации и инструментальной независимости. Однако на уровне реальных языковых практик наука не может устранить интерпретативное участие исследователя: язык научных статей неизбежно включает элементы оценки, выбора и риторической организации материала. Эта риторичность и субъективность проявляются даже в употреблениях самой лексемы *объективный*. Это приводит к внутреннему напряжению между декларируемой безличностью знания и фактической природой его производства, в которой объективность оказывается не устранением субъекта, а результатом социально организованных практик.

Источники

Баженов и др., 2014 – *Баженов Ю.М., Чернышов Е.М., Коротких Д.Н.* Конструирование структур современных бетонов: определяющие принципы и технологические платформы // Construction materials. 2014. № 3. С. 6–14.

Веселовский, 2008 – *Веселовский И.С.* Происхождение солнечного ветра: астрофизический и плазменно-физический аспекты проблемы // Солнечно-земная физика. 2008. Т. 1. С. 93–98.

Владимиров, 1999 – *Владимиров Ю.А.* Лазерная терапия: настоящее и будущее // Соровский образовательный журнал. 1999. № 2. С. 2–8.

Гармаш и др., 2005 – *Гармаш В.Б., Егоров Ф.А., Коломиец Л.Н. и др.* Возможности, задачи и перспективы волоконно-оптических измерительных систем в современном приборостроении // Фотон-экспресс. 2005. № 6 (46). С. 128–140.

Гичев, Гичев, 1999 – *Гичев Ю.П., Гичев Ю.Ю.* Влияние электромагнитных полей на здоровье человека // Экология. Серия аналитических обзоров мировой литературы. 1999. № 52. С. 1–91.

Жмудь, 2015 – *Жмудь В.А.* Системный подход к доказательствам или опровержениям гипотез // Автоматика и программная инженерия. 2015. № 4 (14). С. 100–133.

Коренбаум и др., 2004 – *Коренбаум В.И., Почекутова И.А., Костив А.Е., Кулаков Ю.В., Тагильцев А.А.* Акустическая диагностика системы дыхания человека на основе объективного анализа дыхательных звуков // Вестник Дальневосточного отделения Российской академии наук. 2004. № 5 (117). С. 68–79.

Леонов и др., 2018 – *Леонов О.А., Шкаруба Н.Ж., Вергазова Ю.Г., Антонова У.Ю.* Метрологическое обеспечение контроля гильз цилиндров при ремонте дизелей // Вестник Барановичского государственного университета. Серия «Технические науки». 2018. № 6. С. 104–109.

Луценко, 2011 – *Луценко Е.В.* Методологические аспекты выявления, представления и использования знаний в асканализе и интеллектуальной системе «Эйдос» // Научный журнал Кубанского государственного аграрного университета. 2011. № 70. С. 1–49.

Морозов, 2014 – *Морозов А.В.* Качественная теория дифференциальных уравнений – основная составляющая теории динамических систем // Труды Военно-космической академии им. А.Ф. Можайского. 2014. Вып. 642. С. 177–184.

Роговой, 2005 – *Роговой А.А.* Определяющие соотношения для конечных упруго-неупругих деформаций // Прикладная механика и техническая физика. 2005. Т. 46. № 5. С. 138–149.

Сандуляк и др., 2010 – *Сандуляк А.А., Свистунов Д.И., Полисмакова М.Н. и др.* «Экстраполируемая цепочка» магнитных тест-фильтров как средство контроля ферропримесей // Законодательная и прикладная метрология. 2010. № 3. С. 26, 35–39.

Список литературы

Арутюнова, 1988 – *Арутюнова Н.Д.* Типы языковых значений: Оценка. Событие. Факт. М.: Наука, 1988. 339 с.

Арутюнова, 1999 – *Арутюнова Н.Д.* Язык и мир человека. М.: Школа «Языки русской культуры», 1999. 911 с.

Витгенштейн, 2018 – *Витгенштейн Л.* Философские исследования / Пер. с нем. Л. Добросельского. М.: АСТ, 2018. 352 с.

Дастон, Галисон, 2018 – *Дастон Л., Галисон П.* Объективность / Пер. с англ. Т. Вархотова, С. Гавриленко, А. Писарева. М.: НЛО, 2018. 584 с.

Кустова, 2008 – *Кустова Г.И.* Оценки истинности/ложности (на материале прилагательных «истинный» и «ложный») // Логический анализ языка. Между ложью и фантазией / Под ред. Н.Д. Арутюновой. М.: Индрик, 2008. С. 159–170.

Логический анализ языка, 1991 – *Логический анализ языка.* Культурные концепты. М.: Наука, 1991. 203 с.

Логический анализ языка, 1995 – *Логический анализ языка.* Истина и истинность в культуре и языке / Под ред. Н.Д. Арутюновой, Н.К. Рябцевой. М.: Наука, 1995. 202 с.

Полани, 1995 – *Полани М.* Личностное знание. На пути к посткритической философии / Пер. с англ. М.Б. Гнедовского, Н.М. Смирновой, Б.А. Старостина; ред. В.А. Лекторского, В.А. Аршинова. М., 1995. 344 с.

Радбиль, 2017 – *Радбиль Т.Б.* Язык и мир: парадоксы взаимоотражения. М.: Изд. дом ЯСК, 2017. 591 с.

Радбиль, 2016 – *Радбиль Т.Б.* Метаязыковые показатели со значением истинности в речевых стратегиях de re // Логический анализ языка. Информационная структура текстов разных жанров и эпох. М.: Гнозис, 2016. С. 137–148.

Сахарова, 2025 – *Сахарова А.В.* Лингвистическая одиссея науки: через сомнение к границам знания. М.: Межрегиональная общественная организация «Русское общество истории и философии науки», 2025. 219 с.

Соколова, 2022 – *Соколова Т.Д.* Объективность: регулятивный идеал или мнимая цель? // Цифровой ученый: лаборатория философа. 2022. Т. 5. № 4. С. 27–32. DOI: 10.32326/2618-9267-2022-5-4-27-32.

Шиповалова, 2014 – *Шиповалова Л.В.* Научная объективность в исторической перспективе: дис. … доктора философских наук: 09.00.08. СПб., 2014.

Шиповалова, 2025 – *Шиповалова Л.В.* Концепт объективности как проблема философии науки // Эпистемология и философия науки. 2025. Т. 62. № 1. С. 6–21. DOI: 10.5840/eps20256211.

References

Arutyunova, N. *Tipy yazykovykh znachenii: Otsenka. Sobytie. Fakt* [Types of Linguistic Meanings: Evaluation. Event. Fact]. Moscow: Nauka Publ., 1988, 339 pp. (In Russian)

Arutyunova, N. *Yazyk i mir cheloveka* [Language and the Human World]. Moscow: Shkola “Yazyki russkoi kul’tury” Publ., 1999, 911 pp. (In Russian)

Bloor, 1984 – Bloor, D. "A Sociological Theory of Objectivity", *Royal Institute of Philosophy Lecture Series*, 1984, vol. 17, pp. 229–245. DOI: 10.1017/S0957042X00002005.

Daston, L. & Galison, P. *Ob"ektivnost'* [Objectivity], trans. by T. Varkhotova, S. Gavrilenko, A. Pisareva. Moscow: NLO Publ., 2018, 584 pp. (In Russian)

Kustova, G. "Otsenki istinnosti/lozhnosti (na materiale prilagatel'nykh 'istinnyi' i 'lozhnyi')" [Evaluations of Truth/Falsity (Based on the Adjectives "True" and "False")], in: N. Arutyunova (ed.). *Logicheskii analiz yazyka. Mezhdu lozh'yu i fantaziei* [Logical Analysis of Language. Between Falsehood and Fantasy]. Moscow: Indrik Publ., 2008, pp. 159–170. (In Russian)

Lederman & Zeidler, 2023 – Lederman, N.G. & Zeidler, D.L. *Handbook of Research on Science Education: Volume III*, ed. by J.S. Lederman. London: Routledge, 2023.

Logicheskii analiz yazyka. Istina i istinnost' v kul'ture i yazyke [Logical Analysis of Language. Truth and Truthfulness in Culture and Language]. Moscow, 1995, 202 pp. (In Russian)

Logicheskii analiz yazyka. Kul'turnye kontsepty [Logical Analysis of Language. Cultural Concepts]. Moscow: Nauka Publ., 1991, 203 pp. (In Russian)

Polanyi, M. *Lichnostnoe znanie. Na puti k postkriticheskoi filosofii* [Personal Knowledge. Towards a Post-Critical Philosophy], trans. by M. Gnedovsky, N. Smirnova, B. Starostin, ed. by V. Lektorsky, V. Arshinov. Moscow, 1995, 344 pp. (In Russian)

Radbil, T. "Metayazykovye pokazateli so znacheniem istinnosti v rechevykh strategiyakh de re" [Metalinguistic Markers of Truth in de re Speech Strategies], in: *Logicheskii analiz yazyka. Informatsionnaya struktura tekstov raznykh zhanrov i epokh* [Logical Analysis of Language. Information Structure of Texts of Different Genres and Epochs]. Moscow: Gnozis Publ., 2016, pp. 137–148. (In Russian)

Radbil, T. *Yazyk i mir: paradoksy vzaimootrazheniya* [Language and the World: Paradoxes of Mutual Reflection]. Moscow: YaSK Publ., 2017, 591 pp. (In Russian)

Sakharova, A. *Lingvisticheskaya odisseya nauki: cherez somnenie k granitsam znaniya* [Linguistic Odyssey of Science: From Doubt to the Limits of Knowledge]. Moscow: Russian Society for the History and Philosophy of Science Publ., 2025, 219 pp. (In Russian)

Shapin, 2015 – Shapin, S. "Truth and Credibility in Science", *International Encyclopedia of the Social & Behavioral Sciences*, 2015, vol. 24, pp. 673–678.

Shipovalova, L. "Kontsept ob"ektivnosti kak problema filosofii nauki" [The Concept of Objectivity as a Problem of Philosophy of Science], *Epistemology & Philosophy of Science*, 2025, vol. 62, no. 1, pp. 6–21. DOI: 10.5840/eps20256211.

Shipovalova, L. *Nauchnaya ob"ektivnost' v istoricheskoi perspektive: dis. doktora filosofskikh nauk* [Scientific Objectivity in Historical Perspective: Doctoral Dissertation]. St. Petersburg, 2014. (In Russian)

Sokolova, T.D. "Ob"ektivnost': regulyativnyy ideal, ili mnimaya tsel'?" [Objectivity: A Regulative Ideal, Or an Imaginary Goal?], *The Digital Scholar: Philosopher's Lab*, 2022, vol. 5 (4), pp. 27–32. DOI: 10.32326/2618-9267-2022-5-4-27-32. (In Russian)

Wittgenstein, L. *Filosofskie issledovaniya* [Philosophical Investigations], trans. by L. Dobroselsky. Moscow: AST Publ., 2018, 352 pp. (In Russian)

Zerbe, Hanganu-Bresch & Cutrufello, 2025 – Zerbe, M.J., Hanganu-Bresch, C., Cutrufello, G. "The Rhetoric of Science in (Times of) Crisis", *Rhetoric Society Quarterly*, 2025, vol. 55, no. 3, pp. 207–222. DOI: 10.1080/02773945.2025.2493481.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 189–206
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 189–206
DOI: https://doi.org/10.5840/eps202663348

Адаптивный ответ: рецепция взглядов Ф. Энгельса на наследственность и антропогенез советскими биологами (1920–1950-е гг.)

Винарский Максим Викторович – доктор биологических наук, профессор.
Санкт-Петербургский государственный университет.
Российская Федерация, 199034, г. Санкт-Петербург, Университетская наб., д. 7/9.
Ведущий научный сотрудник.
Санкт-Петербургский филиал Института истории естествознания и техники им. С.И. Вавилова РАН.
Российская Федерация, 199034, г. Санкт-Петербург, Университетская наб., д. 5;
e-mail: radix.vinarski@gmail.com



На основе изучения опубликованных и архивных текстов в статье рассматривается многообразие подходов советских биологов к рецепции высказываний Энгельса (и других «классиков марксизма»), относящихся к проблеме происхождения человека и наследованию приобретенных признаков (неоламаркизму). Показано, что в условиях советской идеократии догматический подход к пониманию и использованию канонизированных текстов «основоположников марксизма» часто сочетался с явно селективным и интерпретирующим подходом, что позволяло до некоторой степени согласовать высказывания «классиков» с современными научными данными. При незыблемости философского и научного авторитета Энгельса в целом, в советское время замалчивались его откровенно расистские высказывания о врожденных различиях когнитивных способностей «высших» и «низших» рас, а его убеждение в реальности ламаркистского механизма «наследования приобретенных признаков» трактовалось ситуативно, в зависимости от собственной научной позиции интерпретатора и существовавшей в данный момент времени официальной оценки научной состоятельности и философской приемлемости ламаркизма. Такое разнообразие приемов цитирования и интерпретации рассматривается как «адаптивный ответ» советского биологического сообщества на требование государственной власти придерживаться «единственно правильного» марксистского мировоззрения, канонизированного в виде корпуса доктринальных текстов Маркса, Энгельса, Ленина и Сталина.

Ключевые слова: «Диалектика природы», диалектический материализм, антропогенез, неоламаркизм, наследование приобретенных признаков, советская идеология

Adaptive response: the reception of F. Engels' views on heredity and anthropogenesis by Soviet biologists (1920s–1950s)

Maxim V. Vinarski – DSc in Biology, Professor.

Based on a study of published and archival texts, the article examines the diversity of approaches among Soviet biologists

© Винарский М.В., 2026

Saint Petersburg State University.
7/9 Universitetskaya Emb., Saint Petersburg 199034, Russian Federation.
Saint Petersburg branch of the S.I. Vavilov Institute of the History of Science and Technology, Russian Academy of Sciences.
5 Universitetskaya Emb., Saint Petersburg 199034, Russian Federation;
e-mail: radix.vinarski@gmail.com

to the reception of statements by Engels (and other "classics of Marxism") concerning the problem of the origin of man and the inheritance of acquired characteristics (neo-Lamarckism). It is shown that under the conditions of Soviet ideocracy, the dogmatic approach to the understanding and use of the canonized texts of the "founders of Marxism" was often combined with a distinctly selective and interpretive approach, which made it possible to reconcile the statements of the "classics" with contemporary scientific data to a certain degree. Despite the inviolability of Engels' philosophical and scientific authority in general, his overtly racist statements about the innate differences in cognitive abilities between "higher" and "lower" races were hushed up during the Soviet period, while his conviction in the reality of the Lamarckian mechanism of "inheritance of acquired characteristics" was interpreted situationally, depending on the interpreter's own scientific position and the official assessment at any given time of the scientific validity and philosophical acceptability of Lamarckism. This diversity of citation and interpretive practices is considered as an "adaptive response" of the Soviet biological community to the demand of state authorities to adhere to the "only correct" Marxist worldview, canonized in the form of a corpus of doctrinal texts by Marx, Engels, Lenin, and Stalin.

Keywords: "Dialectics of Nature", dialectical materialism, human evolution, neo-Lamarckism, inheritance of acquired traits, Soviet ideology

Введение

Трактовка советского государственного устройства как квазирелигиозного по своей природе давно стала общим местом. Н.А. Бердяев [Berdyaev, 1937] был едва ли первым, но, вероятно, самым влиятельным из авторов, проводивших аналогию между классической теократией и советской идеократией. Параллели так очевидны, что в литературе на них указывалось многократно. Действительно, уже в начале 1920-х гг. революционная власть создала в СССР свою, секулярную, обрядность и культ «святых» (революционные мыслители, мученики за идею), сформировала догматику и сословие профессиональных толкователей догмы. Наконец, существовало и «священное писание» – корпус сочинений классиков марксизма. Следование этому канону считалось обязательным условием для работы в области науки и философии. Уже в 1922 г. В.И. Ленин в директивной форме потребовал от каждого советского естествоиспытателя «быть современным материалистом, сознательным сторонником того материализма, который представлен Марксом, то есть… быть диалектическим материалистом» [Ленин, 1922, с. 9–10]. От ученых-гуманитариев столь же безоговорочно требовалось следовать положениям исторического материализма.



Принятие в качестве источника мировоззренческой истины свода канонизированных текстов неизбежно привело к известному в богословском дискурсе феномену начетничества и цитатности. Прагматика и тактика использования цитат из «классиков марксизма» советскими учеными подробнее всего изучены на примере сообщества историков [Сидорова, 2008; Юрганов, 2019; Крих, 2020]. В наиболее прямолинейной форме их суммировал в своих воспоминаниях Арон Гуревич: «Неотъемлемой стороной научного труда историка являлось тогда обильное цитирование "классиков", высказывания которых воспринимались как истина в последней инстанции» [Гуревич, 2004, с. 89]. В науку вошло понятие о «ритуальном цитировании», целью которого был не поиск научной истины, а демонстрация авторами собственной политической лояльности и знакомства с доктринальными философскими текстами [Сидорова, 2008; Винарский, 2025].

Но если для гуманитариев основными источниками цитат служили сочинения (причем не только опубликованные при жизни, но и рукописные тексты и частная переписка) Маркса, Ленина и (в 1930–1950-е гг.) Сталина, то в области естественных наук важнейшим авторитетом были тексты Фридриха Энгельса. Прочие «классики» вопросами философии естествознания занимались мало (исключение – «Материализм и эмпириокритицизм» Ленина). Энгельс же рассматривался как в советской [Кедров, 1947], так и в западной [Jordan, 1967] литературе как автор, специально и глубоко обращавшийся к достижениям современных ему естественных наук и потому внесший решающий вклад в «кодификацию» диалектического материализма. В 1920-е гг. Н. Гредескул (бывший член ЦК кадетской партии, ставший лояльным советским профессором) обосновал это в виде микронарратива: Маркс и Энгельс, желая «сомкнуть свое учение с естествознанием», организовали «своего рода научное разделение труда: Маркс продолжал работать в области обществознания, а Энгельс был как бы откомандирован от имени марксизма на фронт естествознания. И он в течение десятков лет изучал естествознание…» [Гредескул, 1928, с. 23].

Особенности обращения советских ученых-естественников к работам «классиков» изучены далеко не достаточно. В этой статье я рассматриваю историю «творческой» апроприации биологами СССР взглядов Энгельса на происхождение человека и связанных с этой проблемой идей о наследственности. Представляется, что, в силу своей первостепенной идеологической актуальности, проблема антропогенеза была той сферой, где ритуалистический (а также «творческий») подход к заимствованию из марксистского канона мог проявляться в наиболее обнаженном виде.

Фридрих Энгельс и «ламарксизм»

Советская литература о происхождении человека начала 1920-х гг., как научная, так и научно-популярная, находилась под сильным влиянием воззрений Томаса Хаксли и в еще большей степени Эрнста Геккеля, «немецкого Дарвина» – наиболее влиятельного на рубеже столетий «наследника» и интерпретатора дарвиновского учения.

Для Хаксли [Huxley, 1863] и Геккеля [Haeckel, 1891; Геккель, 1909] двумя ведущими факторами антропогенеза были возникновение двуногости (бипедализма) и членораздельной речи. Оба этих фактора имели чисто биологическую природу. Ранние советские авторы, например В.В. Лункевич [Лункевич, 1923], послушно следовали этому представлению.

Все изменилось в 1925 г., когда в СССР под названием «Диалектика природы» были изданы оставшиеся в рукописи заметки Ф. Энгельса, работавшего в конце жизни над капитальным трудом о философии естествознания. Хотя эти рукописные материалы не были подготовлены к изданию и невозможно судить о том, какой вид бы они приняли, успей Энгельс довести свой труд до печатного станка, в советском идеоцентризме они получили такой же канонический статус, как опубликованные при жизни Энгельса работы по естествознанию («Анти-Дюринг» и др.). В состав «Диалектики природы» вошел довольно обширный раздел, в котором Энгельс излагал свои взгляды на происхождение человека [Энгельс, 1952, с. 132–144]. В отличие от профессиональных биологов, он подчеркивал роль социальных факторов в антропогенезе, в первую очередь – роль труда. Эта так называемая «трудовая теория» очень быстро стала обязательной для советских философов, социологов и антропологов. Одним из первых ее ввел в научный оборот упомянутый выше Гредескул [Гредескул, 1928].

Сам Энгельс считал, что эта идея стала новым пониманием антропогенеза, превзошедшим теории его предшественников, так как даже «материалистически мыслящие естествоиспытатели из школы Дарвина (Геккель? – *М.В.*) не могут себе составить ясного представления о происхождении человека, так как в силу влияния… идеалистического миросозерцания они не видят роли, которую играл при этом труд» [Энгельс, 1952, с. 139].

Однако при изложении фактологии Энгельс во многом следовал как последнему слову науки того времени построениям Геккеля, принимая, к примеру, его гипотезу о прародине человечества – материке Лемурия, позднее затонувшем в водах Индийского океана. Следуя Геккелю, Энгельс признавал в качестве значимого эволюционного фактора неоламаркистский механизм «наследования приобретенных признаков» (НПП), который в позднейшей литературе

часто трактуется как антидарвинистский (хотя сам Дарвин признавал НПП в качестве одного из факторов эволюции). В то время, до работ А. Вейсмана, подвергшего сомнению реальность НПП, это было общим местом в биологии. Наконец, Энгельс не отрицал значения бипедализма и артикулированной речи, но, в отличие от Хаксли и Геккеля, придавал этим факторам антропогенеза лишь второстепенное значение. Так, человеческую речь он понимал как производное от трудовой деятельности, результат расширения «господства над природой» и сплочения социальных связей, в результате чего у «формировавшихся людей» «явилась потребность что-то сказать друг другу» [Энгельс, 1952, с. 134–135]. НПП рассматривалось Энгельсом как реальный движитель антропогенеза, позволявший сделать наследственными навыки и привычки, возникавшие у архаичных людей в ходе трудовой деятельности.

Для советских авторов все эти высказывания «классика» имели двоякое значение. С одной стороны, ими можно было пользоваться как очень авторитетным аргументом в научных спорах. С другой стороны, в «Диалектике природы» было зафиксировано состояние науки конца XIX в., включая положения, отвергнутые уже в начале XX столетия. Например, Геккель в 1902 г. отказался от своей лемурийской гипотезы, «перенеся» прародину человечества в Южную Азию [Геккель, 1909, с. 337–339]. Вопрос о том, как относиться к подобным «реликтам», разделил советское научное сообщество на два лагеря.

Некоторые естественники-марксисты считали, что научный факт, будучи признан классиками марксизма, тем самым становится не подлежащим ревизии абсолютом. Такой подход в 1920-е гг. проявился в дискуссиях о «физическом идеализме», в частности о теории относительности. Известно, что Эйнштейн своей теорией опроверг существование мирового эфира. Тем не менее некоторые физики в СССР считали возможным усомниться в этом, приводя цитаты из Энгельса и Ленина, в которых упоминается мировой эфир [Сонин, 2017]. Лженаучной и «поповской» по схожим соображениям была объявлена и концепция нестационарной, расширяющейся Вселенной. Диамат унаследовал от немецкой классической философии ньютонианское представление о бесконечной и неизменной Вселенной, что в советских условиях немедленно было догматизировано – по утверждению борцов с «идеализмом» в науке, гипотеза нестационарной Вселенной находится в «непреодолимом противоречии с материализмом» [Там же, с. 101]. Идеалистические же теории в СССР рассматривались как лженаучные по одному факту их реальной или предполагаемой «идеалистичности».

В довоенное время вопрос о реальности НПП как фактора эволюции был одним из наиболее остро дебатировавшихся среди биологов и философов. В то время в классической генетике уже устоялось

выдвинутое А. Вейсманом и развитое Т. Морганом положение о том, что наследование приобретенных признаков в принципе невозможно. Оппоненты генетиков – механоламаркисты – настаивали на том, что НПП доказано экспериментально, и с удовольствием ссылались на «Диалектику природы» как на авторитетный источник своей веры.

В 1927 г. в Коммунистической академии в Москве состоялась дискуссия о проблемах евгеники, основным докладчиком на которой выступил антрополог М. Волоцкой, ламаркист по убеждениям. В полном согласии с Энгельсом он утверждал, что трудовые навыки, «положительное влияние труда», могут наследоваться, что спасает пролетариат от вырождения, выступая в качестве контрбаланса вредному влиянию капиталистической эксплуатации и плохих условий жизни [Волоцкой, 1927, с. 226]. Как выразился один из сторонников Волоцкого, «процесс возникновения нового общественного человека» возможен только в том случае, если приобретенные признаки наследуются [Там же, с. 241]. Известная максима Энгельса о том, что человека создал труд, понималась (и, пожалуй, справедливо) как отражение его ламаркистских убеждений.

По аналогии с социал-дарвинизмом, такую концепцию можно обозначить как *социал-ламаркизм* (в 1920-е гг. бытовало также словечко «ламарксизм», употреблявшееся, насколько могу судить, иронически).

Однако в той же дискуссии участвовали и антиламаркисты. Один из них, М. Местергази, констатировал: «По отношению к не признающим наследования приобретенных признаков применялись разные малопривлекательные эпитеты. Однажды даже ставилось на голосование: признавать ли марксистам [такое] наследование или не признавать. Хорошо, что это голосование не прошло. Прямо говорили, что ежели кто не признает наследования, тот не марксист. Марксист, мол, должен быть обязательно ламаркистом» [Волоцкой, 1927, с. 234]. Действительно, еще в 1925 г. на страницах ведущего теоретического журнала «Под знаменем марксизма» В.Н. Слепков (в будущем профессиональный генетик, а пока – начинающий диаматчик) весьма развязным тоном поучал своих оппонентов: «…очень не мешало бы нашим евгенистам поучиться социологической грамоте, чтобы не совать биологические факторы и куда нужно и куда не нужно»; им надо знать «Тезисы о Фейербахе» Маркса, в которых сказано, что «люди представляют собой продукт обстоятельств и воспитания и что, следовательно, изменившиеся люди являются продуктом изменившихся обстоятельств и другого воспитания» [Слепков, 1925, с. 113]. Именно на этих основаниях – задолго до печально известной «августовской» сессии ВАСХНИЛ 1948 г. – в СССР начались нападки на генетику как на «буржуазную» и «идеалистическую» науку.

Для советских противников ламаркизма высказывания Энгельса представляли серьезный вызов. Применительно к антропогенезу



в «Диалектике природы» прямо говорилось, что «[с]овременное естествознание признает наследственность приобретенных свойств и этим расширяет субъект опыта, распространяя его с индивида на род: теперь уже не считается необходимым, чтобы каждый отдельный индивид лично испытал все на своем опыте; его индивидуальный опыт может быть до известной степени заменен результатами опыта его предков» [Энгельс, 1952, с. 213]. Не имея возможности отвергнуть «авторитет» классика, антиламаркисты в СССР часто выбирали тактику «творческого» подхода к его наследию, пытаясь разделить собственно философское и фактологическое содержание сочинений Энгельса. Противники НПП доказывали, что, вопреки букве этих текстов, можно оставаться марксистом, не становясь сторонником механоламаркизма. Жаркая полемика по этому поводу развернулась весной 1931 г. на дискуссии о «механистическом материализме и меньшевиствующем идеализме» в биологии, подводившей черту под научно-философскими спорами 1920-х гг.[1]

На этой дискуссии один из противников ламаркизма публично заявлял, что позиция Энгельса по вопросу НПП отражает общее мнение натуралистов того времени (включая Дарвина), которое впоследствии было признано ошибочным. Это, по его словам, нисколько не умаляет ни величия Энгельса, ни гениальности его теории происхождения человека. Поэтому, «если мы в этом с Энгельсом разойдемся, никакой беды не будет, и напрасно некоторые лица хотят нас заставить придерживаться каждой буквы Энгельса. Для меня Энгельс тоже авторитет, но я вовсе не считаю, что я должен держаться за каждую букву Энгельса» [Против…, 1931, с. 63].

Данная дискуссия шла уже на весьма изменившемся идеологическом фоне. В первой половине 1929 г. «механистическое» направление в советском марксизме было признано идеалистическим, а вместе с ним под удар попал и механоламаркизм. Теория Ламарка и даже само его имя на некоторое время стали в СССР одиозными. Перед биологами встала задача «очистить» Энгельса от обвинений в ламаркизме, что породило очередной виток дебатов. Пытались даже отрицать очевидное – явное признание Энгельсом НПП, для чего приходилось прибегать к несколько казуистическим аргументам. Например, М. Левин предлагал считать, что на самом деле Энгельс имел в виду не НПП в смысле Ламарка, а говорил о «социальной традиции… социальной преемственности» в антропогенезе [Там же, с. 37]. Генетик А.С. Серебровский доказывал, что Энгельса нельзя

[1] Еще одна подобная дискуссия, на которой сторонники и противники НПП обвиняли друг друга в попытках ревизовать Энгельса, имела место в апреле 1929 г. на 2-й Всесоюзной конференции марксистско-ленинских научных учреждений. Ограниченный объем статьи не позволяет мне подробно остановиться на ее рассмотрении.

называть ламаркистом, потому что он не разделял идеалистических положений теории Ламарка, а что касается НПП, то в эпоху Энгельса оно признавалось и дарвинистами, и только потом «этот вопрос подвергся пересмотру на основе новейших научных данных» [Против…, 1931, с. 40]. Серебровский был прав в том отношении, что в годы, когда Энгельс работал над фрагментами «Диалектики природы», реальность НПП никем не отрицалась и даже не ставился вопрос о необходимости ее экспериментальной проверки. Первым, кто поставил такой вопрос и попытался его решить опытным путем, был А. Вейсман. Но его первая работа, посвященная опровержению НПП, опубликована только в 1883 г. В «Диалектике природы» она не упоминается, вероятно, Энгельс ее не знал.

Результаты споров 1931 г. отразились в хрестоматии «Маркс, Энгельс, Ленин о биологии», готовили которую победители в дискуссии об «идеализме» в биологии. В сопроводительном тексте к высказываниям Энгельса об антропогенезе было сказано, что «механоламаркистское крыло в биологии» присвоило себе имя Энгельса, «наделив его представлениями о наследственности и изменчивости в духе пошлых метафизических ламаркистских спекуляций по вопросам наследования благоприобретенных “признаков” и т.д.». Механоламаркисты истолковали мысли Энгельса как подтверждение НПП, хотя, по мнению составителей антологии, Энгельс «никогда не солидаризировался с Ламарком и не мог солидаризироваться с ним» [Маркс, Энгельс…, 1933, с. 175]. В этом была определенная правота, потому что исторический ламаркизм – чрезвычайно эклектичная доктрина, в которую, помимо НПП, входило много других объясняющих гипотез. Споры историков биологии о том, что считать подлинным ламаркизмом (ламаркизмом в смысле Ламарка), ведутся до сих пор [Винарский, 2024]. Поэтому Энгельс мог признавать НПП, не считая себя связанным интеллектуальным родством с Ламарком.

В этом контексте понятно более позднее высказывание генетика Н.П. Дубинина, относящееся уже к 1936 г. Полемизируя с «мичуринскими» биологами, он указывал, что один из них, С. Перов, «позорил» «великое имя» Энгельса, позволяя себе его «смешивать с ламаркистами» (цит. по: [Дубинин, 1992, с. 63]).

Впрочем, уже к середине 1930-х гг. споры об интерпретации наследия Энгельса практически потеряли всякий смысл, поскольку возобладал узкодогматический подход к высказываниям «классиков» как к вневременной истине, не подлежащей никакой ревизии. Прерогатива истолкования наследия Маркса, Энгельса и Ленина была делегирована единственному живому «классику», а полифонизм мнений сменился показным единодушием. Такое изменение интеллектуальной ситуации в замечательно экспрессивных выражениях обрисовал один советский историк: «…пришел хозяин, стукнул кулаком по столу и сказал: довольно валять дурака, надо работать. <…>



До сих пор у нас были бесконечные и бесплодные дискуссии. Сколько голов – столько умов. Масса отсебятины…» (цит. по: [Юрганов, 2019, с. 68]). Другому историку той эпохи принадлежит точное определение аподиктического значения цитат из «классиков»: «…если мы возьмем классиков марксизма, то они, хорошо зная исторические факты, дают такие точки зрения, *которые никогда не меняются*, потому что *эта точка зрения есть научная истина*, потому что они даны мыслителями-диалектиками» [Там же, с. 172–173] (курсив мой. – *М.В.*).

«Читать, читать и бесконечно читать…»

Несмотря на агрессивную антиламаркистскую риторику начала 1930-х гг., уже к середине десятилетия механоламаркистская доктрина вошла в качестве очень важного элемента в «мичуринскую» теорию Т.Д. Лысенко. Лысенкоисты, принимавшие НПП, открещивались от ярлыка «ламаркистов» [Kouprianov, 2011] и продолжали делать это до 1946 г., когда история с ламаркистскими симпатиями «классиков марксизма» получила новый поворот.

Марксистский канон расширялся. В 1946 г. вышел первый том собрания сочинений И.В. Сталина [Сталин, 1946], в котором были впервые на русском языке опубликованы его ранние статьи в грузинских газетах, составившие серию «Анархизм или социализм?». В этих статьях, относящихся к 1906–1907 гг., Сталин довольно специфически и весьма упрощенно (если не сказать вульгарно) интерпретировал вопросы эволюционной теории, в которых профессионально не разбирался. Тем не менее его высказывания о дарвинизме, ламаркизме, а также о происхождении человека были тут же признаны гениальными и почти на десятилетие обязательными для цитирования в соответствующих контекстах.

Для моей темы важно то, что высказывания молодого Сталина входили в прямое противоречие с текстами Энгельса. Это дало Лысенко и его сторонникам возможность «обойти» высказывания последнего в «Диалектике природы» и предпочесть им мнения другого классика, в то время гораздо более актуального. Здесь проявилась издревле знакомая христианским богословам коллизия – наличие взаимно противоречивых утверждений в тесте Писания, что требовало особых экзегетических усилий, чтобы либо показать мнимость этих противоречий, либо достичь синтетического понимания их смысла.

«Мичуринцы» полагали, что в своем труде Сталин вскрыл «коренную ошибку» Дарвина, которой оказалась даже не мальтузианская «подкладка» его теории, а так называемый «плоский эволю-

ционизм». Так именовалось присущее якобы Дарвину понимание эволюции как процесса постепенного и плавного видоизменения, лишенного «скачков», которые соответствовали бы моментам быстрого перехода количества в качество. Сталин полагал, что, отрицая «скачки» в эволюции, на необходимости которых настаивает диалектическое мышление, дарвинизм отрицает революционный аспект развития. Эта теория представляет собой антипод теории катастроф Ж. Кювье, объяснявшей развитие живой природы исключительно «скачками», пароксизмами. И то, и другое – метафизично. Таким образом, автор «Анархизма или социализма?» неявно полемизировал с Энгельсом, увидевшим в дарвиновской теории триумф «рациональной диалектики», после которого вся «старая телеология пошла к черту» [Энгельс, 1952, с. 154]. Конечно, в 1906 г. Сталину не была известна «Диалектика природы», но он вполне мог читать «Анти-Дюринг» Энгельса, впервые опубликованный в русском переводе в 1904 г., где говорилось, что Дарвин нанес решительный удар «метафизическому взгляду на природу» [Маркс, Энгельс…, 1933, с. 27]. Таким образом, трактовка дарвинизма и проблем наследственности Сталиным оказалась более актуальной в послевоенном СССР, чем взгляды Энгельса.

Еще одну выгоду, которую извлекли лысенкоисты из опубликования труда молодого Джугашвили, была поддержка им принципа НПП. Хотя в тексте «Анархизма или социализма?» нет прямого обсуждения этого принципа, автор его недвусмысленно симпатизировал теории «неоламаркизма, которой уступает место неодарвинизм» [Сталин, 1946, с. 300]. Рассуждения о происхождении человека, содержавшиеся в этих ранних сталинских статьях, также носили – в полном согласии с Энгельсом – ламаркистский характер. Светлое имя Ламарка было реабилитировано, а «Анархизм или социализм?» вошел в обязательный круг чтения любого советского биолога.

Чрезвычайно наглядно это проявилось в 1947 г. на дискуссии о реальности внутривидовой конкуренции (дарвинистский постулат, который Лысенко решительно отвергал). Видный «мичуринец» А.А. Авакян обвинял своих противников в том, что они ревизуют наследие Маркса и не учитывают ставший доступным на русском языке текст Сталина. Он говорил: «…нельзя выступать с утверждениями, что у Маркса в его время было мало фактического материала по внутривидовой конкуренции. Это не к лицу любому ученому, независимо от чина нашего масштаба (так! – *М.В.*). И нельзя отрицать, что внутривидовая конкуренция это… учение Мальтуса». Он указывал своему оппоненту, что тому «нужно читать Сталина I-й том, читать, читать и бесконечно читать». О себе Авакян говорил, что он «человек маленький, рядовой, уважающий наших академиков, но просящий более серьезно относиться к мнению



основоположников марксизма-ленинизма, что поможет им убедиться в отсутствии внутривидовой конкуренции»[2].

Однако уже в 1953 г. цитаты из Сталина стали утрачивать респектабельность [Сидорова, 2008]. Прошло совсем немного лет, и философ Г.В. Платонов, сторонник лысенкоизма, был вынужден дезавуировать высказывания из «Анархизма и социализма?». По его словам, Сталин писал свои статьи, будучи «молодым, начинающим марксистом», и только в условиях культа личности это произведение, не отличающееся «большой точностью и ясностью», могло стать основой марксистской оценки дарвинизма [Платонов, 1958, с. 34]. Поэтому в середине 1950-х гг. лысенкоисты, доказывая реальность НПП, вновь обратились к авторитету Энгельса (см., например, [Фейгинсон, 1955, с. 134]).

В свое время Л.А. Сидорова [Сидорова, 2008], анализируя работу советских историков с цитатами из основоположников марксизма, выделила три «группы приемов» обращения с ними: 1. Догматическая (цитата как неопровержимый аргумент в научной дискуссии). 2. Формальная (ритуальная), т.е. приведение цитат без прямого отношения к содержанию самого научного труда[3]. 3. Творческая, когда цитата используется как отправная точка для собственного исследования. По ее мнению, третья группа получила распространение в основном уже после смерти Сталина, но на протяжении всей советской истории господствовал все-таки догматический подход к цитированию как следствие «методологического монизма советской исторической науки» [Там же, с. 219].

На мой взгляд, многие из приведенных выше примеров соответствуют еще одному, четвертому, подходу к цитированию «классиков», который можно назвать оппортунистическим (или даже негативистским). Он использовался либо для снятия противоречия между конфликтующими по смыслу цитатами (путем выбора одной из них), либо же для обхода «неудобных» высказываний путем их игнорирования. Последнее можно также продемонстрировать на примере взглядов Энгельса на происхождение человека.

Большинство изводов ламаркизма, признающих ведущее влияние условий среды на наследственность, естественным образом приводят к мысли о врожденном неравенстве отдельных человеческих рас или наций ввиду разности их исторических судеб. Энгельс, как и Геккель (и Дарвин, и большинство натуралистов конца XIX в.), признавал существование «высших» и «низших» рас

[2] «Справка о дискуссии, вызванной высказываниями Лысенко об отсутствии в природе борьбы за существование». Архив РАН. Ф. 1521. Оп. 1. Д. 290. Л. 9.

[3] Замечу, что в биологии круг ритуально цитируемых авторов помимо «классиков» марксизма включал в 1930–1950-е гг. также Дарвина, Тимирязева, Мичурина, Лысенко [Kouprianov, 2011; Винарский, 2025].

и наличие генетически обусловленных различий в их интеллектуальных способностях. «Диалектика природы» говорит об этом безо всяких обиняков: «Если, например, у нас математические аксиомы представляются каждому восьмилетнему ребенку чем-то само собой разумеющимся, не нуждающимся ни в каком опытном доказательстве, то это является лишь результатом “накопленной наследственности”. Бушмену же или австралийскому негру вряд ли можно втолковать их [даже] посредством доказательства» [Энгельс, 1952, с. 213–214].

В своем дискурсе о «высших» и «низших» расах, как и в некоторых других высказываниях об эволюции человека, Энгельс, несомненно, следовал Геккелю – самому влиятельному в те годы автору, развивавшему эволюционную теорию антропогенеза. Геккель выстраивал расы человека на лестнице совершенства, от высших к низшим. Последние представляют собой последнюю ступень, ниже которой располагаются уже «обезьянолюди». Сопоставьте с этим слова Энгельса о том, что «самые низшие дикари, даже такие, у которых приходится предположить возврат к звероподобному состоянию с одновременным физическим вырождением», все же стоят выше «вымерших обезьянолюдей» [Там же, с. 133]. В текстах Геккеля найдется и прямая параллель к рассуждениям Энгельса о врожденной неспособности «дикаря» к математике: «Очень многие дикие народы умеют считать только до десяти или двадцати, между тем как умных собак заставляли добиваться счисления до 40 и даже свыше 60». Поэтому невозможно «насадить человеческую культуру там, где еще нет для этого надлежащей почвы, т.е. определенной степени развития мозга» [Геккель, 1909, с. 375].

В 1920-е гг. откровенно расистское (с нынешней точки зрения) учение о «высших» и «низших» расах не было чем-то необычным в советской литературе. Так, один из основоположников «механистического» течения в отечественном марксизме, И. Скворцов-Степанов, писал, что «в устройстве костей современного дикаря, в относительной силе передних и задних конечностей, в особенностях челюстей, скуловых костей и в высоких надбровных дугах… а вместе с тем и в психических свойствах, еще больше сходства с человекообразными обезьянами, чем у культурных народов: китайцев, японцев, татар, европейцев, индусов, арабов» [Степанов, 1925, с. 28]. Он же утверждал, что «китайский кули (чернорабочий) сильно уступает американскому рабочему по своей работоспособности и для некоторых работ, требующих изощренной, напряженной внимательности и высокой нервной возбудимости, *пока вообще непригоден*. Экономисту при выяснении этих явлений очень полезно заглянуть в *хорошее университетское руководство по физиологии*» [Там же, с. 41] (курсив мой. – *М.В.*). При этом автор отмежевывался от социал-дарвинистского тезиса о том, что есть люди, от природы



осужденные «на жалкое прозябание» и повиновение – «все это в усиленной степени относится к туземному населению Африки, Азии, Австралии: к неграм, китайцам, индусам, арабам и т.д.» [Степанов, 1925, с. 58].

Противники ламаркизма энергично опровергали это представление и также апеллировали к математическим способностям, стараясь при этом не задевать имя Энгельса. Упомянутый выше М. Местергази, оппонируя Волоцкому, говорил: «Ламаркистам лучше не касаться вопроса о способностях малокультурных народностей. <…> Если взять московского обывателя и какого-нибудь тунгуса или самоеда, то никакой разницы в одаренности нет. Чем московский обыватель отличается от тунгуса? Только тем, что он рождается в более богатой культурной среде. <…> Машинка его, его нейроны, по количеству и разнообразию, может быть, не отличаются от тунгусских, но условные связи устанавливаются в гораздо большем количестве. <…> В своей бедной жизни тунгус считает до десяти – не больше, и предки его больше десяти не считали, а сюда приезжают ребята и прекрасно математикой занимаются. Нисколько не хуже нас. Значит, никаких наследственных тяжестей от прошлого их существования они на себе не несут» [Волоцкой, 1927, с. 234]. Это – полемика по отношению к «Диалектике природы», но полемика скрытая.

В 1930-е гг. любые намеки на расизм и идею расовой неравноценности в СССР были табуированы, что было обусловлено в том числе приходом к власти национал-социалистов в Германии. Общепринятым мнением в советской антропологии стало то, что что биологически детерминированных различий в умственном и психическом развитии рас не существует. «Успешное национальное развитие народов Индии, Индонезии, молодых государств Африки после их освобождения от колониализма служит ярким свидетельством, что не “плохая психика”, а колониальное господство было тормозом на пути их исторического развития» [Якимов, 1963, с. 360].

Итак, Энгельс был повинен в грехе признания наследственного неравенства рас, врожденности и неизбывности интеллектуальных различий между ними, что находило свое объяснение в рамках неоламаркистской парадигмы. То, что советские идеологи не обвиняли Энгельса в том, что́ они поставили бы на вид любому другому западному автору (тому же Эрнсту Геккелю, например), показывает, что их подход к цитированию «классиков» имел селективный (или оппортунистический) характер. Неудобные высказывания канонизированных авторов просто игнорировались. Так, никто не писал всерьез про «прародину» человечества Лемурию, существование которой в XX в. было убедительно опровергнуто геологическими данными.

Это проявлялось в отношении не только к расизму. Вот пример, взятый из переписки биолога и философа А.А. Любищева, который

был хорошо знаком с гебраистом И.Д. Амусиным – видным советским специалистом по происхождению христианства. По свидетельству Любищева, Амусину «с трудом удалось от редактора добиться, чтобы вставить цитату из Энгельса, восстанавливающую старое положение, что христианство возникло в Палестине и именно в начале первого века, так как наши марксиствующие ослы и холуи считали, что "научно доказано", что Христа не было, а христианство возникло в диаспоре лет на 200–300 позже. Я знаю, многие верующие христиане очень довольны этой книгой»[4]. В данном случае автору пришлось не игнорировать, а «пробивать» в печать цитаты из «классика», оказавшиеся неуместными в данный момент развития советской идеологии.

Заключение

Я полагаю, что многообразие приемов цитирования и интерпретирования цитат «классиков» советскими учеными, показанное в этой работе, а также другими исследователями (Л.А. Сидорова, А.Л. Юрганов, С.Б. Крих), свидетельствует о том, что научное сообщество СССР давало эффективный «адаптивный ответ» (используя термин эволюционной экологии) на требование власти сообразовывать научные убеждения с канонизированными взглядами «классиков марксизма». Хотя принятие диалектического (или исторического) материализма в качестве методологической базы и личного мировоззрения далеко не всегда было вынужденным и/или показным, реальной исследовательской мысли зачастую было тесно в установленных идеологическими надзирателями (А.А. Любищев называл их «философскими нянюшками», призванными опекать правоверие советских ученых) рамках. Сохраняя в массе своей политическую лояльность власти, ученые были вынуждены маневрировать, и в этих разнообразных маневрах проявлялась «научная позиция» конкретного ученого, «присущий ему исследовательский стиль» [Сидорова, 2008, с. 188]. Поэтому «существование рациональных элементов в отношении советских [ученых] к марксистско-ленинскому наследию на фоне преобладающего догматического характера его использования отмечалось» на всем протяжении 1930–1950-х гг. [Там же, с. 196]. По наблюдению С.Б. Криха [Крих, 2020, с. 166], каждый советский ученый умел «аккуратно обходить или специфично трактовать любого

[4] Письмо А.А. Любищева В.Л. Циопкало, 15.11.1961. Санкт-Петербургский филиал Архива РАН. Ф. 1033. Оп. 2. Д. 521. Л. 130. Имеется в виду книга «Рукописи Мертвого моря» [Амусин, 1960]. Автор ссылается на высказывания Энгельса на стр. 262 и 263.



из “классиков”… и такого рода операции вовсе не были нелигитимными – пока не переходилась грань, которую негласно чувствовало научное сообщество».

Известно, что даже в обществах с развитым каноническим мышлением и жесткой нормативизацией творчества наличие канона, например эстетического, не всегда приводит к тотальной догматизации. Каноническая схема в известных случаях «провоцирует талантливого мастера на конкретное преодоление ее внутри нее самой системой мало заметных, но художественно значимых отклонений от нее» [Бычков, 2020, с. 288]. Это высказывание известного специалиста по эстетике может быть, *mutatis mutandis*, применено и к рассмотренным выше перипетиям.

Список литературы

Амусин, 1960 – *Амусин И.Д.* Рукописи Мертвого моря. М.: Изд-во АН СССР, 1960. 272 с.

Бычков, 2020 – *Бычков В.В.* Эстетика. М.: Академический проект, 2020. 452 с.

Винарский, 2024 – *Винарский М.В.* «Новое учение о виде» Трофима Лысенко: тексты и контексты. СПб.: Реноме, 2024. 304 с.

Винарский, 2025 – *Винарский М.В.* Ритуальное цитирование в советской биологической литературе: на примере «Зоологического журнала» (1930–1950-е гг.) // Проблемы деятельности ученого и научных коллективов. 2025. Т. 11. С. 68–79.

Волоцкой, 1927 – *Волоцкой М.* Спорные вопросы евгеники [с прениями по докладу] // Вестник Коммунистической академии. 1927. Кн. 20. С. 212–254.

Геккель, 1909 – *Геккель Э.* Естественная история миротворения. СПб.: Научная мысль, 1909. Кн. 2. 384 с.

Гредескул, 1928 – *Гредескул Н.А.* Происхождение человека от обезьяны. Л.: Красная газета, 1928. 45 с.

Гуревич, 2004 – *Гуревич А.Я.* История историка. М.: РОССПЭН, 2004. 284 с.

Дубинин, 1992 – *Дубинин Н.П.* История и трагедия советской генетики. М.: Наука, 1992. 384 с.

Кедров, 1947 – *Кедров Б.М.* Энгельс и естествознание. М.: Политиздат, 1947. 480 с.

Крих, 2020 – *Крих С.Б.* Другая история: «Периферийная» советская наука о древности. М.: НЛО, 2020. 314 с.

Ленин, 1922 – *Ленин В.И.* О значении воинствующего материализма // Под знаменем марксизма. 1922. № 3. С. 5–12.

Лункевич, 1923 – *Лункевич В.В.* Откуда и как пошел род людской? Происхождение человека. Симферополь: Пролетарское дело, 1923. 66 с.

Маркс, Энгельс…, 1933 – Маркс, Энгельс, Ленин о биологии / Сост.: М.И. Айзупет, П.А. Исаков, Б.П. Токин. М.: Партиздат, 1933. 240 с.

Платонов, 1958 – *Платонов Г.В.* Мичуринское учение и дарвинизм // Философские вопросы естествознания. М.: Изд-во МГУ, 1958. Вып. 1. С. 7–65.

Против…, 1931 – Против механистического материализма и меньшевиствующего идеализма в биологии / Под ред. П.П. Бондаренко, В.С. Брандгендлера, М.С. Мицкевича, Б.П. Токина. М.; Л.: Медгиз, 1931. 104 с.

Сидорова, 2008 – *Сидорова Л.А.* Советская историческая наука середины XX века: синтез трех поколений историков. М.: ИРИ РАН, 2008. 293 с.

Слепков, 1925 – *Слепков В.Н.* Наследственность и отбор у человека // Под знаменем марксизма. 1925. № 4. С. 102–122.

Сонин, 2017 – *Сонин А.С.* Физический идеализм. Драматический путь внедрения революционных идей физики начала XX века. М.: ЛЕНАНД, 2017. 320 с.

Сталин, 1946 – *Сталин И.В.* Сочинения. М.: ОГИЗ-Госполитиздат, 1946. Т. 1. 425 с.

Степанов, 1925 – *Степанов И.* Исторический материализм и современное естествознание. М.: Госиздат, 1925. 84 с.

Фейгинсон, 1955 – *Фейгинсон Н.И.* Основные вопросы мичуринской генетики. М.: Изд-во МГУ, 1955. 454 с.

Энгельс, 1952 – *Энгельс Ф.* Диалектика природы. М.: Политиздат, 1952. 328 с.

Юрганов, 2019 – *Юрганов А.Л.* Русское национальное государство: Жизненный мир историков эпохи сталинизма. М.: РГГУ, 2019. 765 с.

Якимов, 1963 – *Якимов В.П.* От обезьяны к человеку, от животного стада к обществу // Очерк диалектики живой природы / Под ред. Г.В. Платонова. М.: Соцэкгиз, 1963. С. 319–368.

References

Amusin, I.D. *Rukopisi Mertvogo morya* [The Dead Sea Scrolls]. Moscow: The USSR Academy of Sciences Press, 1960. (In Russian)

Berdyaev, 1937 – Berdyaev, N. *The Origin of Russian Communism*. London: Geoffrey Bles, 1937.

Bychkov, V.V. *Estetika* [Aesthetics]. Moscow: Akademicheskiy Proekt, 2020. (In Russian)

Dubinin, N.P. *Istoriya i tragediya sovetskoy genetiki* [The History and Tragedy of Soviet Genetics]. Moscow: Nauka, 1992. (In Russian)

Engels, E. *Dialektika prirody* [Dialectics of Nature]. Moscow: Politizdat, 1952. (In Russian)

Feiginson, N.I. *Osnovnyye voprosy michurinskoy genetiki* [The Basic Questions of Michurinist Genetics]. Moscow: The Moscow State Univerity Press, 1955. (In Russian)

Gredeskul, N.A. *Proiskhozhdeniye cheloveka ot obezyany* [The Origin of Man from an Ape]. Leningrad: Krasnaya gazeta, 1928. (In Russian)

Gurevich, A.Y. *Istoria istorika* [The History of an Historian]. Moscow: ROSSPEN Publ., 2004. (In Russian)

Haeckel, 1891 – Haeckel, E. *Anthropogenie; oder Entwickelungsgeschichte des menschen. Keimes- und stammesgeschichte*. Leipzig: W. Engelmann, 1891.

Haeckel, E. *Estestvennanya istoriya mirotvoreniya* [The Natural History of the World Creation]. St. Petersburg: Nauchnaya mysl', 1909. Vol. 2. (In Russian)

Huxley, 1863 – Huxley, T.H. *Evidence as to Man's Place in Nature*. London: Williams & Norgate, 1863.

Jordan, 1967 – Jordan, Z.A. *The Evolution of Dialectical Materialism*. London etc.: Macmillan, 1967.

Kedrov, B.M. *Engels i estestvoznaniye* [Engels and Natural Sciences]. Moscow: Politizdat, 1947. (In Russian)

Kouprianov, 2011 – Kouprianov, A.V. "The 'Soviet Creative Darwinism' (1930s–1950s): From the selective reading of Darwin's works to the transmutation of species", *Studies in the History of Biology*, 2011, vol. 3, pp. 8–31.

Krikh, S.B. *Drugaya istoriya: "Periferiinaya" sovetskaya nauka o drevnosti* [The Other History: The "Marginal" Soviet Science of Antiquity]. Moscow: Novoye Literaturnoye Obozreniye, 2020. (In Russian)

Lenin, V.I. "O znachenii voinstvuyushchego materializma" [On the Significance of Militant Materialism], *Pod znamenem marksizma*, 1922, no. 3, pp. 5–12. (In Russian)

Lunkevich, V.V. *Otkuda i kak poshel rod lyudskoy. Proiskhozhdenie cheloveka* [Where and How Did the Human Race Come from? The Origin of Man]. Simferopol: Proletarskoye delo, 1923. (In Russian)

Marks, Engels, Lenin o biologii [Marx, Engels, Lenin on Biology]. Moscow: Partizdat, 1933. (In Russian)

Platonov, G.V. "Michurinskoye ucheniye i darwinizm" [The Michurinist Doctrine and Darwinism]. In: *Filosofskiye voprosy estestvoznaniya*. Moscow: The Moscow State University Press, 1958, vol. 1, pp. 7–65. (In Russian)

Protiv mekhanisticheskogo materializma i men'shevistvuyushchego idealizma v biologii [Against Mechanistic Materialism and Menshevik-style Idealism in Biology]. Moscow: Medgiz, 1931. (In Russian)

Sidorova, L.A. *Sovetskaya istoricheskaya nauka serediny XX veka: sintez tryokh pokoleniy istorikov* [Soviet historiography of the middle of the 20th century: a synthesis of three generations of historians]. Moscow: Institute of Russian History, 2008. (In Russian)

Slepkov, V.N. "Nasledstvennost' i otbor u cheloveka" [Heredity and selection in humans], *Pod znamenem marksizma*, 1925, no. 4, pp. 102–122. (In Russian)

Sonin, A.S. *Fizicheskiy idealizm. Dramaticheskiy put' vnedreniya revolyutsionnykh idey fiziki nachala XX veka* [Physical Idealism. The Dramatic Path of Implementation of Revolutionary Ideas of Physics at the Beginning of the 20th Century]. Moscow: LENAND, 2017. (In Russian)

Stalin, I.V. *Sochineniya* [Works]. Moscow: OGIZ-Gospolitizdat, 1946. Vol. 1. (In Russian)

Stepanov, I. *Istoricheskii materializm i sovremennoe estestvoznanie*. [Historical Materialism and Modern Science]. Moscow: Gosizdat, 1925. (In Russian)

Vinarskiy, M.V. *"Novoe uchenie o vide" Trofima Lysenko: teksty i kontekst*y ["The New Doctrine of Species" by Trofim Lysenko: Texts and Contexts]. St. Petersburg: Renome Publ., 2024. (In Russian)

Vinarskiy, M.V. "Ritual'noye tsitirovaniye v sovetskoy biologicheskoy literature: na primere 'Zoologichskogo zhurnala' (1930–1950-e gg.) [Ritual Citation in Soviet Bio-

logical Literature: On the Example of "Zoologicheskiy zhurnal", 1930–1950s], *Problemy deyatel'nosti uchenogo i nauchnykh kollektivov*, vol. 11, pp. 68–79. (In Russian)

Volotskoi, M. "Spornyye voprosy eugeniki (s preniyami po dokladu)" [Disputable Questions of Eugenics (with a Discussion on the Lecture)], *Vestnik Komakademii*, 1927, vol. 20, pp. 212–254. (In Russian)

Yakimov, V.P. Ot obezyany k cheloveku: ot zhivotnogo stada k obshchestvu [From Ape to Man, from Animal Herd to Society]. In: *Ocherki dialektiki zhivoy prirody* [Essays on Dialectics of Living Nature]. Moscow: Sotsekgiz, 1963, pp. 319–368. (In Russian)

Yurganov, A.L. *Russkoe natsional'noe gosudarstvo: Zhiznennyi mir istorikov epokhi stalinizma* [The Russian National State: The Lifeworld of Historians of the Stalinist Era]. Moscow: RGGU Publ., 2019. (In Russian)

Эпистемология и философия науки
2026. Т. 63. № 3. С. 207–219
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 207–219
DOI: https://doi.org/10.5840/eps202663349

Мыслить с машинами: автоматизм, автономия и историческая эпистемология интеллекта

Беляева Ульяна Павловна – кандидат философских наук, доцент.
Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского.
Российская Федерация, 398020, г. Липецк, Ленина ул., д. 42;
e-mail: ulyana-sin@yandex.ru



Статья представляет собой аналитический обзор книги Дэвида У. Бейтса «Искусственная история естественного интеллекта: мышление с машинами от Декарта до цифровой эпохи» (2024), посвященной историческим условиям формирования современного понятия интеллекта. В центре обзора находится анализ исследовательской стратегии Бейтса, позволяющей проследить, как оппозиция естественного и искусственного интеллекта формировалась внутри длительной истории машинных, физиологических, кибернетических и технических моделей мышления. Методологически обзор опирается на проблемно-концептуальный анализ монографии, историко-философскую реконструкцию аргументации и сопоставление с исследованиями кибернетики, философии техники, распределенной когниции и предиктивной обработки. Показано, что Бейтс отказывается от линейной истории искусственного интеллекта и строит траекторию проблем: от картезианской физиологии, раннемодерновых теорий автомата и кантовской проблематики суждения к Бэббиджу, Лавлейс, психофизике, термодинамике, философской антропологии, кибернетике, Тьюрингу, теориям технической экстериоризации и моделям машинного обучения. Основной результат обзора состоит в выявлении эпистемологического значения тезиса Бейтса о том, что оппозиция естественного и искусственного интеллекта исторически производна. Машина здесь предстает не инструментом готового субъекта, а эпистемической фигурой, через которую разум делал наблюдаемыми собственные операции, ограничения и формы автономии. В заключение обосновывается, что книга Бейтса продуктивна для философии науки, поскольку позволяет рассматривать познание как историю материальных, телесных и технических моделей мышления. Одновременно отмечаются проблемные стороны работы: избирательность исторической композиции, недостаточная проясненность различия между человеческой нормотворческой автономией и машинной адаптивностью, а также ограниченность евроамериканской рамки исследования.

Ключевые слова: Дэвид У. Бейтс, историческая эпистемология, естественный интеллект, искусственный интеллект, автоматизм, автономия, кибернетика

© Беляева У.П., 2026

THINKING WITH MACHINES: AUTOMATISM, AUTONOMY, AND THE HISTORICAL EPISTEMOLOGY OF INTELLIGENCE

Ulyana P. Belyaeva –
PhD in Philosophy,
Associate Professor.
Lipetsk State Pedagogical University named after P.P. Semenov-Tyan-Shansky.
42 Lenin St., Lipetsk 398020, Russian Federation;
e-mail: ulyana-sin@yandex.ru

The article presents an analytical review of David W. Bates's book *Artificial History of Natural Intelligence: Thinking with Machines from Descartes to the Digital Age* (2024), which examines the historical conditions for the formation of the modern concept of intelligence. The review focuses on an analysis of Bates's research strategy, which makes it possible to trace how the opposition between natural and artificial intelligence took shape within the long history of machine, physiological, cybernetic, and technical models of thought. Methodologically, the review draws on a problem-conceptual analysis of the monograph, a historical-philosophical reconstruction of its argumentation, and a comparison with studies in cybernetics, philosophy of technology, distributed cognition, and predictive processing. It is shown that Bates rejects a linear history of artificial intelligence and instead constructs a trajectory of problems: from Cartesian physiology, early modern theories of the automaton, and the Kantian problematic of judgment, to Babbage, Lovelace, psychophysics, thermodynamics, philosophical anthropology, cybernetics, Turing, theories of technical exteriorization, and models of machine learning. The main result of the review consists in identifying the epistemological significance of Bates's thesis that the opposition between natural and artificial intelligence is historically derivative. The machine here appears not as an instrument of a ready-made subject, but as an epistemic figure through which the mind made its own operations, limitations, and forms of autonomy observable. In conclusion, it is argued that Bates's book is productive for the philosophy of science, insofar as it allows cognition to be examined as a history of material, bodily, and technical models of thought. At the same time, problematic aspects of the work are noted: the selectivity of the historical composition, the insufficient clarification of the difference between human normative autonomy and machine adaptability, as well as the limitations of the Euro-American framework of the study.

Keywords: David W. Bates, historical epistemology, natural intelligence, artificial intelligence, automatism, autonomy, cybernetics

Книга Дэвида У. Бейтса *An Artificial History of Natural Intelligence: Thinking with Machines from Descartes to the Digital Age* [Bates, 2024], вышедшая в 2024 г., принадлежит к тому типу исследований, которые трудно отнести к одной дисциплинарной рубрике. Формально перед нами история идей, связанная с философией сознания, историей техники, кибернетикой и современными когнитивными науками. По существу же это историко-эпистемологическое исследование о том, как интеллект становился объектом знания через машинные, физиологические, технические и символические модели. Бейтс



пишет не историю искусственного интеллекта в привычном смысле и не очередное рассуждение о том, способна ли машина мыслить. Его предмет значительно сложнее: он реконструирует длительную историю взаимного моделирования человеческого разума и машины, в ходе которой сама оппозиция «естественного» и «искусственного» интеллекта обнаруживает свою производность.

Дэвид У. Бейтс, профессор риторики Калифорнийского университета в Беркли, давно работает на пересечении истории науки, исследований техники и истории когнитивных представлений. Его более ранние книги *Enlightenment Aberrations: Error and Revolution in France* [Bates, 2002] и *States of War: Enlightenment Origins of the Political* [Bates, 2011] уже показывали устойчивый интерес Бейтса к внутренним напряжениям рациональности модерна – к тому, как ошибка, порядок, политическое решение, суверенность и техническая организация знания становятся философски значимыми формами описания человека и общества. Новая монография продолжает эту линию, но переносит ее в область исторической реконструкции разума. Бейтс рассматривает интеллект как исторически изменчивую конфигурацию, в которой соединяются телесная автоматичность, техническая экстериоризация человеческих функций и автономия. Последнюю он понимает как способность приостанавливать сложившиеся схемы действия и вырабатывать новые нормы в ситуации неопределенности.

Для эпистемологического дискурса эта книга важна потому, что она смещает разговор об искусственном интеллекте из сферы технoэтической реакции в область анализа условий познания. Бейтс показывает, что интеллект не может быть адекватно понят, если рассматривать его либо как внутреннее свойство субъекта, либо как функциональную способность системы к вычислению. Машина в его книге оказывается особой эпистемической фигурой, через которую разум описывает собственную организацию, границы и возможность автономного действия. В этом смысле монография соотносится с линией исследований, представленной Ж.-П. Дюпюи, Н. Кэтрин Хейлз, Э. Кларком, Д. Чалмерсом, Б. Стиглером, Ж. Симондоном и А. Леруа-Гураном, но отличается от них более длинной историко-философской дугой: от картезианской физиологии и раннемодерновой проблемы автомата до машинного обучения и предиктивной обработки. Одновременно Бейтс не скрывает, что работает преимущественно внутри западной традиции философии и истории науки – от картезианской и кантовской проблематики разума до англо-американской кибернетики и французской философии техники, – и прямо позиционирует свою книгу как внутреннюю критику западной рациональности [Ibid., p. 9–12].

Бейтс подчеркивает, что предлагает «не историю автоматичности, не историю ИИ и вообще не историю чего-либо» [Bates, 2024, p. 6].

Эта формула не является риторическим отказом от историчности. Скорее, она обозначает отказ от линейной генеалогии, в которой современный искусственный интеллект выводился бы из ряда технических предшественников. Бейтс строит не хронологию изобретений, а траекторию проблем. Его интересует не то, как машины постепенно стали «умнее», а то, как менялись сами описания человеческого мышления: в разные исторические моменты оно понималось как машинное, органическое, автоматическое, пластичное, технически расширенное, а также как способное в критический момент приостановить собственные автоматизмы и перейти к новой норме действия. Поэтому книга движется через философские тексты, физиологические теории, психофизику, термодинамику, антропологию техники, кибернетику и когнитивную науку, не превращая ни одну из этих областей в единственное объяснительное основание.

Центральная напряженность монографии задается парой понятий «автономия» и «автоматичность». Бейтс исходит из того, что современная когнитивная наука и нейронауки все чаще описывают человека как систему автоматических процессов, информационной обработки, нейронного предсказания и бессознательной регуляции. В этой перспективе автономия субъекта кажется либо иллюзией, либо вторичным эффектом более глубоких автоматизмов. Гуманитарная же критика, как показывает Бейтс, зачастую развивает две достаточно уязвимые позиции – либо начиная защищать «человеческое» так, будто оно только теперь лишь подверглось угрозе, либо обращается к языку ценностей, который сам нуждается в дополнительном обосновании. Бейтс выбирает другой путь. Его интересует не спасение человеческой автономии от автоматизма, а возможность увидеть автономию как эффект сложной и многогранной истории автоматических, телесных, технических и символических связей. Отсюда и возникает радикальная и достаточно полемическая мысль: «естественного интеллекта не существует; весь интеллект искусственен» [Bates, 2024, p. 13]. Но здесь важно не считывать ее слишком буквально. Бейтс предлагает более тонкую интерпретацию человеческого, снимая представление о чистом природном разуме. Человеческое мышление, с точки зрения автора, с самого начала организовывалось множеством посредников-медиаторов, и поэтому оно не имеет какого-то «естественного ядра», которое можно было бы противопоставить машине.

Первый смысловой пласт книги связан с рассмотрением идей Декарта. Однако здесь автор уходит от ожидаемой апелляции к «картезианскому дуализму», фокусируя внимание на его описании тела как сложного самодействующего автомата, напоминающего систему внутренней регуляции. В трактате о человеке и в физиологических сочинениях Декарт описывает тело как сложный автомат, чья нервная система перерабатывает внешние воздействия, кодирует их



и организует ответные действия. Здесь Бейтс обнаруживает раннюю форму того, что можно назвать кибернетической интуицией: тело не просто приводится в движение внешними силами, оно управляет собой через внутренние потоки сигналов, память, привычку и интеграцию сенсорных данных [Bates, 2024, p. 20–25]. Такой ход можно соотнести с историко-научной реконструкцией Дж. Рискин, для которой раннемодерновая машина никогда не была только образом мертвой механики: вокруг автоматов, животных и живых тел постоянно велся спор о том, может ли механизм обладать внутренней активностью, чувствительностью и подобием агентности [Riskin, 2016].

Именно поэтому картезианская душа у Бейтса не выступает простым сверхмеханическим остатком. Она важна как фигура прерывания. Если тело уже мыслится как «корпоральная когниция», то чистый интеллект не управляет им извне, как суверен управляет механизмом. Его функция состоит в возможности разорвать автоматические сцепления ощущения, памяти, воображения и аффекта. В анализе воска, в суждении о людях за окном, в описании страстей Бейтс видит не отвлеченную метафизику субъекта, а эпистемологическую сцену, где мышление появляется как способность схватывать отношение, не данное телесному коду непосредственно [Bates, 2024, p. 26–29]. Отсюда возникает важная для всей книги мысль: автономия не предшествует автоматичности, она появляется внутри автоматической организации тела как возможность задержки, переописания и решения.

Гоббс и Спиноза позволяют Бейтсу показать, что раннемодерновая механизация разума не сводилась к редукции мышления к вычислению. В случае Гоббса обычно вспоминают тезис о рассуждении как счете, но Бейтса интересует более широкая логика искусственности. Имена, знаки, политическое тело, гражданский порядок образуют искусственные системы, в которых человеческая рациональность получает устойчивость. Государство как «искусственный человек» у Гоббса становится не просто метафорой власти, а моделью того, как разум нуждается в материально-знаковой организации, выходящей за пределы индивидуального мозга [Ibid., p. 33–36]. У Спинозы, напротив, ключевым становится отказ от представления о внешней свободе. Мышление связано с телесной организацией, а метод превращает ум в «духовный автомат», действующий по собственным законам. Это не уничтожает свободу, но переводит ее в плоскость адекватности, формирования и переориентации аффективно-телесных автоматизмов.

Лейбниц в этой реконструкции оказывается фигурой «автоматической гармонии». Бейтс показывает, что знаменитое учение о предустановленной гармонии можно читать не только как метафизическую доктрину, но и как попытку мыслить согласованность органического, механического и духовного порядков. Лейбницевская

монада не нуждается во внешнем механическом воздействии, но это не делает ее антимашинной. Напротив, душа и тело оказываются двумя способами организации порядка, каждый из которых обладает собственным автоматизмом. Человек, в отличие от простой машины, способен понимать порядок и создавать искусственные системы, в которых этот порядок получает материальное выражение [Bates, 2024, p. 49–60]. Здесь формируется важнейшая для Бейтса тема: техника не возникает после мышления как его инструмент, она становится способом, с помощью которого мышление выводит собственные операции в мир.

Юм и Кант задают следующую фазу этой истории. У Юма Бейтс видит решительный подрыв раннемодерновой уверенности в том, что разум способен обнаружить в мире порядок, аналогичный порядку человеческого замысла. В *Dialogues Concerning Natural Religion* мысль и интеллект оказываются лишь одними из «пружин» универсума, а не привилегированным ключом к его устройству. Юмовская критика причинности и субстанциального субъекта тем самым получает у Бейтса физиолого-эпистемологическое измерение: разум слаб, привычка сильна, воображение связывает идеи до всякого рационального контроля. Но эта слабость разума не просто скептический тупик. Она выявляет искусственный характер рациональности, которая должна быть сформирована, поддержана и внешне закреплена, чтобы противостоять естественной инерции аффектов и ассоциаций [Ibid., p. 69–73].

Кант у Бейтса занимает пороговое положение. С одной стороны, критическая философия систематизирует автоматические условия познания: формы чувственности, категории рассудка, синтетическое единство апперцепции, регулятивные функции разума. С другой стороны, именно у Канта особенно остро встает вопрос о спонтанности, суждении и нормативности. Бейтс обращает внимание на кантовскую формулу человека как, возможно, «самого искусственного» существа [Ibid., p. 80]. Это чрезвычайно показательно. Кантовский субъект не может быть сведен к машине, но и не является природной непосредственностью. Его познание формируется через сложную систему способностей, а рефлектирующая способность суждения действует там, где нет заранее данной закономерности. Поэтому кантовская автономия для Бейтса не отменяет автоматизм познания, а обнаруживает возможность нормы там, где автоматизм не дает готового правила [Ibid., p. 83–98]. В этой точке книга особенно близка Кангилему, у которого организм способен устанавливать новые нормы в ответ на изменение среды и собственные патологические состояния [Canguilhem, 1991]. Поэтому кантовская проблема суждения у Бейтса получает биофилософский смысл: автономия оказывается связанной не с выходом из автоматизма, а со способностью живой системы переопределять условия собственного функционирования.



Вторая большая линия монографии связана с XIX в., индустриальной машиной и экспериментальными науками о теле. Бэббидж и Лавлейс появляются у Бейтса не как предшественники компьютера в популярной истории техники, а как фигуры новой эпистемологии операций. Разностная машина Бэббиджа возникает из проблемы ошибки в математических таблицах. Механизация вычисления должна устранить ненадежное человеческое звено, но тем самым она делает видимым сам процесс рациональной операции. Аналитическая машина важна еще больше, поскольку отделяет вычисление от конкретной задачи и превращает операцию в программируемую структуру. Бейтс показывает, что здесь машина становится моделью не столько ума вообще, сколько формальной организации анализа [Bates, 2024, p. 108–115].

Лавлейс у Бейтса играет решающую роль потому, что она видит в машине «науку операций». Ее знаменитая формула о том, что аналитическая машина «ткет алгебраические узоры», у Бейтса получает глубокий смысл. Машина связывает материальное и символическое, порядок движения и порядок знака. Но Лавлейс идет дальше технического описания. Ее мечта об «исчислении нервной системы» показывает, что машинная модель анализа почти сразу возвращается к вопросу о мозге, нервах и физиологии мышления. Так Бейтс фиксирует характерный для XIX в. круг: машина создается как инструмент рационализации, затем становится моделью для понимания мыслительного процесса, а после этого стимулирует новые попытки описать мозг как систему операций [Ibid., p. 115–116].

Психофизика, физиология, логические машины и термодинамика позволяют Бейтсу расширить эту линию. В центре оказывается проблема бессознательного основания сознательного рассуждения. Если логические машины Джевонса или Марканда демонстрируют возможность автоматизации отдельных процедур вывода, то творческая догадка, гипотеза и внезапное понимание остаются не вполне механизируемыми. Поэтому у Бейтса важны не только машины логического вывода, но и Пирс с его абдукцией, Гельмгольц с анализом научного прозрения, психофизиологи, пытавшиеся объяснить переход от нервного процесса к интеллектуальной форме. Машина здесь не закрывает проблему мышления, а делает ее более точной: если рассуждение можно частично автоматизировать, то где возникает разрыв, благодаря которому появляется новая гипотеза?

Термодинамический материал книги особенно продуктивен для эпистемологии. Бейтс связывает энергетические модели, теорию регуляции, телеграфные и промышленные системы с вопросом о точках неопределенности. У Максвелла и в более широком термодинамическом контексте его интересуют «сингулярные точки», где малое вмешательство способно изменить ход процесса [Ibid., p. 125–130]. Эта тема явно перекликается с более ранней проблемой решения.

Человеческий разум не противопоставляется природе как чистая свобода; он действует в тех местах, где автоматический порядок допускает развилку. Поэтому интеллект в книге Бейтса все чаще мыслится как способность распознавать момент, когда система еще не закрыта.

Физиология мозга и эволюционные теории XIX в. усиливают эту перспективу. Бейтс подробно показывает, как локализационные теории мозга, исследования пластичности, идеи компенсации повреждений, понятия привычки и высших центров создают новый образ человека. Мозг мыслится как динамическая, обучаемая и уязвимая система, чья организация не исчерпывается врожденной структурой. В этом контексте культура и техника становятся не надстройкой над природой, а условием формирования человеческих способностей. Капповская теория органопроекции, дарвиновская тема орудия, размышления о письме и печати подводят Бейтса к идее технической экстериоризации: человек познает себя через вынесение своих функций наружу, но это «наружу» обратно перестраивает его внутреннюю организацию [Bates, 2024, p. 141–151].

Третья проблемная зона книги связана с кризисом органического порядка в первой половине XX в. Здесь появляются фон Икскюль, Лотка, Жак Лафитт, философская антропология, Кассирер, Шелер, Хайдеггер, Плеснер. Бейтс не превращает этот массив в отдельную историю биологии или антропологии. Его интересует вопрос о том, как живое и техническое начинают пониматься через общую проблему организации. Так, у фон Икскюля живое существо не укладывается в образ механизма, потому что оно существует не просто в физическом окружении, а в мире значимостей – в собственном Umwelt. Одновременно Лотка позволяет Бейтсу иначе сформулировать сам вопрос о технике: человеческие орудия и машины теперь выступают продолжением эволюции уже за пределами собственно организма, образуя особый экзосоматический режим существования. И уже у Лафитта эта идея получает развитие. Машина, выходя за границы инструментальной функции, перестраивает саму среду человеческого существования.

Философская антропология нужна Бейтсу не как готовый ответ на вопрос «что есть человек», а в качестве дискурса, где этот вопрос заново формулируется в ситуации разрушения прежних метафизических оснований. У Кассирера человек определяется через символическую форму; у Шелера – через дистанцию по отношению к среде; у Хайдеггера – через различие мира и окружения; у Плеснера – через эксцентрическое положение живого существа, которое не совпадает с самим собой. Существенно, что уже во всех этих теоретических конструкциях человек мыслится как существо, способное вынести себя в актах существования за пределы непосредственной исключительно биологической регуляции. Но Бейтс и здесь добавляет



новую контекстуальную оптику считывания традиции философской антропологии, которая сопрягается им с историей машин, знаков, орудий и технических посредников. Поэтому человеческая «исключительность» у него перестает быть только антропологическим свойством. Она формируется в долгом диалоге человека с тем, что он сам изготавливает.

Обращение Бейтса к Витгенштейну выполняет в композиции книги переходную функцию: после анализа философской антропологии оно подводит рассуждение к кибернетике через вопрос о критериях, по которым мы приписываем мышление тому или иному агенту. Рассматривая отдельные фрагменты Zettel, Бейтс показывает, что мысль не является автономным внутренним актом, а укоренена во внешней практике и формах употребления. Соответственно, мышление оказывается распределено между телом, речью, навыком и ситуацией. Это позволяет сформулировать вопрос о критериях мышления: какие практики дают основание приписывать кому-либо или чему-либо способность мыслить.

Кибернетика в реконструкции Бейтса не сводится к редукции организма к машине. Напротив, она оказывается областью, где машина начинает мыслиться через адаптацию, обратную связь, целеполагание, ошибку и патологию. В этом отношении автор оказывается близок к Э. Пикерингу, который описывал кибернетику как особую экспериментальную культуру работы с непредсказуемыми, адаптивными и не полностью контролируемыми системами [Pickering, 2010]. Норберт Винер, Уоррен Маккаллок, Уолтер Питтс, У. Росс Эшби, Джон фон Нейман образуют у Бейтса единое поле напряжений. Кибернетика, отмечает автор, действительно стирает различия между животным, человеком и машиной на языке информации, но при этом вынуждена заниматься сбоями, пределами, перегрузками, ошибками и способностью системы удерживать переменные в жизнеспособных пределах [Bates, 2024, p. 234–255]. Здесь особенно уместна параллель с Н. Кэтрин Хейлз, показавшей, что постгуманистская конфигурация субъекта формируется именно тогда, когда информация начинает мыслиться как относительно независимая от конкретного телесного носителя [Hayles, 1999]. Однако у Бейтса этот кибернетический сдвиг не ведет к исчезновению тела: напротив, тело, машина и знак образуют исторически изменчивую систему, внутри которой заново ставится вопрос об автономии.

Тьюринг занимает в книге место, далекое от привычной схемы «машина может мыслить». Бейтс подчеркивает, что Тьюринг не сводил мышление к вычислению в грубом смысле. Его интересовала возможность построить машину, способную обучаться, ошибаться, менять поведение и выходить за пределы дисциплинированной процедуры. Бейтс цитирует важную формулу Тьюринга: интеллектуальное поведение, вероятно, состоит в «отклонении от полностью

дисциплинированного поведения, вовлеченного в вычисление, но в довольно небольшом отклонении» [Bates, 2024, p. 266]. Это одно из ключевых мест монографии. Интеллект определяется не чистой свободой и не полной программностью, а малым, но существенным отклонением внутри формализованного порядка вычислительной процедуры. Поэтому машина Тьюринга, обучение, случайность, ошибка заключения, внешнее вмешательство и формирование привычек становятся для Бейтса более важными, чем имитационная игра сама по себе [Ibid., p. 268–287].

Последняя часть книги выводит этот анализ к экзосоматической и техногенетической перспективе. Ликлайдер и Энгельбарт у Бейтса важны в качестве мыслителей симбиоза человека и машины. Так, проект усиления интеллекта у Энгельбарта описывается как система, где человек, язык, артефакты и методы образуют функциональное целое. В таком контексте «расширенный разум» Кларка и Чалмерса получает более глубокое историческое обоснование [Clark, Chalmers, 1998]. Бейтс фактически показывает, что расширение разума не является поздним открытием аналитической философии сознания. Оно имеет длинную историю, в которой письмо, печать, логические машины, вычислительные устройства и социальные техники постепенно формировали само представление о том, где проходит граница мышления.

Леруа-Гуран становится одной из центральных фигур этой финальной траектории. Его теория жеста, речи и технической экстериоризации [Leroi-Gourhan, 1993] позволяет Бейтсу описать человека как существо, чья когнитивная история неотделима от орудий, языка, памяти и внешних носителей. Освобождение руки, развитие орудия, символизация и внешнее хранение опыта переводят человеческую эволюцию из чисто биологического времени в историческое [Bates, 2024, p. 311–321]. Разумность здесь формируется в самой связке мозга, руки, орудия, знака и коллективной памяти. В этом моменте можно увидеть идейную перекличку с теорией материального вовлечения Л. Малафуриса, согласно которой вещи участвуют в самой сборке когнитивных операций [Malafouris, 2013]. Кроме того, становится очевидно сближение Бейтса со Стиглером, для которого техника является условием антропогенеза, поскольку внешние носители памяти создают особый слой наследования, не сводимый ни к биологии, ни к индивидуальному опыту [Stiegler, 1994]. Посвящение книги памяти Стиглера в этом отношении не случайно: вся монография может быть прочитана как историческое развертывание тезиса о технической конституции человеческого времени и мышления.

В финале Бейтс обращается к сетевой эпохе, нейронным сетям, машинному обучению и предиктивной обработке. Но и здесь он избегает привычной драматургии «новой эпохи ИИ». Современное машинное обучение для него – не разрыв с прежней историей,



а очередной момент, когда человек через техническую систему переосмысляет собственный интеллект. В версии предиктивной обработки А. Кларка восприятие и действие предстают как динамика ожиданий, коррекции ошибок и активного взаимодействия организма со средой [Clark, 2016]. Именно здесь Бейтс делает существенную оговорку: интеллект не исчерпывается минимизацией ошибки, так как человеческое мышление, определив какую-либо ситуацию как кризисную, оказывается способно сменить сам модус действия и тем самым изменить набор действующих норм.

Вообще Бейтсу удается успешно балансировать между гуманистической идеей абсолютно свободного разума и вычислительной редукцией мышления. С одной стороны, он показывает, что, к примеру, машины выступали одним из исторических условий возникновения мысли. С другой стороны, автор не растворяет в них мысль, которая может выходить из «автоматической связки», создавая новые формы.

Сильной стороной монографии является также отказ от простой критики технологического разума. Бейтс не стремится показать, что современный ИИ является ложным, опасным или антигуманным потому, что машина «не человек». Его аргументация тоньше. Он утверждает, что сама идея «человека» уже исторически сформирована через машины, поэтому защита человеческого от машинного часто опирается на фикцию чистой природности. Но отсюда не следует капитуляция перед автоматизацией. Напротив, если человеческая автономия всегда формировалась внутри технических систем, то критика цифровой эпохи должна спрашивать не о том, как вернуться к дотехническому субъекту, а о том, какие формы технической организации сохраняют способность к решению, ошибке, воображению и созданию новых норм. В этом пункте Бейтс оказывается близок не столько к популярной философии ИИ, сколько к традиции Кангилема, Симондона и Стиглера, где техника и жизнь анализируются через норму, индивидуацию и историческую открытость [Canguilhem, 1991; Simondon, 2017; Stiegler, 1994].

Однако разговор о книге требует и ряда критических оговорок. Первая связана с методологической самоидентификацией книги. Бейтс настойчиво заявляет, что пишет «не историю», а траекторию сцеплений, и эта установка освобождает его от линейности. Но временами она оставляет не до конца проясненными критерии отбора материала. Читатель понимает, почему в книге необходимы Декарт, Кант, Бэббидж, Тьюринг и Леруа-Гуран, но отдельные переходы между фигурами выглядят скорее концептуально убедительными, чем историографически доказанными. Это не разрушает аргументацию, поскольку Бейтс честно признает фрагментарность и избирательность своего проекта [Bates, 2024, p. 12]. Тем не менее для историка науки такой способ выстраивания исследовательского нарратива может показаться слишком свободным.

Вторая оговорка касается нормативной стороны книги. Бейтс убедительно показывает, что автономия должна быть переосмыслена как способность устанавливать нормы внутри автоматических систем. Но вопрос о том, как практически различать человеческую нормотворческую автономию и сложную адаптивность машинных систем, остается открытым. Ссылка на ошибку, прерывание, кризис и решение философски продуктивна, но она нуждается в более строгом различении. Современные системы машинного обучения также работают с ошибкой, перестройкой весов, адаптацией и неполной предсказуемостью. Поэтому требуется дополнительный критерий, позволяющий отличить статистическую перенастройку от нормативного решения. Бейтс опосредованно подходит к этому различию через темы кризиса, будущего и отказа от уже действующей нормы, но оставляет его скорее в форме авторской интуиции, чем развернуто определенного критерия.

Итак, «Искусственная история естественного интеллекта: мышление с машинами от Декарта до цифровой эпохи» является значимым историко-эпистемологическим исследованием. Дэвид У. Бейтс, опираясь на вехи рационального дискурса модерна и значимые антропологические концепты, переосмысляет положение человека и само понятие «естественной интеллектуальности». Книга дает методологический инструментарий для осмысления как ныне модного технофеномена искусственного интеллекта, так и оснований человеческого мышления, позволяя по-новому, через призму генезиса исторических форм знания, взглянуть на его структуру.

Список литературы / References

Bates, 2002 – Bates, D.W. *Enlightenment Aberrations: Error and Revolution in France*. Ithaca; London: Cornell University Press, 2002.

Bates, 2011 – Bates, D.W. *States of War: Enlightenment Origins of the Political*. New York: Columbia University Press, 2011. https://doi.org/10.7312/columbia/9780231158053.001.0001

Bates, 2024 – Bates, D.W. *An Artificial History of Natural Intelligence: Thinking with Machines from Descartes to the Digital Age*. Chicago; London: University of Chicago Press, 2024. https://doi.org/10.7208/chicago/9780226832111.001.0001

Canguilhem, 1991 – Canguilhem, G. *The Normal and the Pathological*. New York: Zone Books, 1991.

Clark, 2016 – Clark, A. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press, 2016. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Clark, Chalmers, 1998 – Clark, A. & Chalmers, D.J. “The Extended Mind”, *Analysis*, 1998, vol. 58, no. 1, pp. 7–19. https://doi.org/10.1093/analys/58.1.7



Hayles, 1999 – Hayles, N.K. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago; London: University of Chicago Press, 1999. https://doi.org/10.7208/chicago/9780226321394.001.0001

Leroi-Gourhan, 1993 – Leroi-Gourhan, A. *Gesture and Speech*. Cambridge, MA; London: MIT Press, 1993.

Malafouris, 2013 – Malafouris, L. *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge, MA; London: MIT Press, 2013. https://doi.org/10.7551/mitpress/9476.001.0001

Pickering, 2010 – Pickering, A. *The Cybernetic Brain: Sketches of Another Future*. Chicago; London: University of Chicago Press, 2010. https://doi.org/10.7208/chicago/9780226667928.001.0001

Riskin, 2016 – Riskin, J. *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*. Chicago; London: University of Chicago Press, 2016. https://doi.org/10.7208/chicago/9780226303086.001.0001

Simondon, 2017 – Simondon, G. *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing, 2017.

Stiegler, 1994 – Stiegler, B. *La technique et le temps, 1: La faute d'Épiméthée*. Paris: Galilée, 1994.

Эпистемология и философия науки
2026. Т. 63. № 3. С. 220–233
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 220–233
DOI: https://doi.org/10.5840/eps202663350

Наука в цифровую эпоху: как ИИ меняет представление о научном знании и его создании

Лисенкова Анастасия Алексеевна – доктор культурологии,
директор.
Высшая школа общественных наук
Гуманитарного института
Санкт-Петербургского политехнического университета
Петра Великого.
Российская Федерация, 195251, г. Санкт-Петербург, ул. Политехническая, д. 29;
e-mail: oskar46@mail.ru

Быльева Дарья Сергеевна – кандидат политических наук, доцент.
Санкт-Петербургский политехнический университет
Петра Великого.
Российская Федерация, 195251, г. Санкт-Петербург, ул. Политехническая, д. 29;
e-mail: bylieva_ds@spbstu.ru

В современной философии науки и техники стремительное развитие технологий искусственного интеллекта (ИИ) выдвигает на первый план фундаментальные вопросы о природе научного познания и его трансформации под влиянием алгоритмических систем. Обзор философских дискуссий об интеллектуальных системах в науке демонстрирует, что за три десятилетия произошел фундаментальный сдвиг точки консенсуса: если на начальном этапе они рассматривались исключительно как инструмент усиления человеческих когнитивных способностей, то в 2022–2023 гг. философское сообщество оказалось перед необходимостью легитимизировать ИИ в качестве нетривиального эпистемического участника. Традиционный вопрос о том, можно ли считать ИИ когнитивным агентом, дополнился проблемами, связанными с практиками его использования: под влиянием больших данных и нейросетей происходит трансформация критериев научности, кризис субъект-объектного различения перед лицом алгоритмической непрозрачности и поиск новых моделей эпистемической ответственности. ИИ утвердился в качестве функционально нередуцируемого участника познавательных практик, однако ни одна из предложенных концепций – от расширенного инструментализма до распределенной или коллаборативной агентности – не устраняет полностью концептуальное напряжение, заданное разрывом между эффективным выполнением ИИ когнитивных операций и отсутствием у него интенциональных состояний. Это делает невозможным прямое применение традиционных критериев обоснованности. В то же время основные дискуссии остаются в рамках эпистемологии первого порядка, сосредоточенной на вопросе о (квази)субъектности машин. Однако данная исследовательская программа, при всей ее продуктивности, упускает из виду фундаментальную трансформацию: приобретение ИИ метанаучной роли, в которой он не только активно участвует в познании, но и во многом перестраивает конфигурацию исследовательского поля.

Ключевые слова: искусственный интеллект, научное познание, нейросети, алгоритмическая прозрачность, когнитивные акторы, этические вызовы, автоматизация науки, искусственный интеллект в науке





© Лисенкова А.А., 2026
© Быльева Д.С., 2026



SCIENCE IN THE DIGITAL AGE: HOW AI IS CHANGING THE UNDERSTANDING OF SCIENTIFIC KNOWLEDGE AND ITS CREATION

Anastasia A. Lisenkova –
DSc in Cultural Studies.
Head of the Graduate School of Social Sciences at the Institute of Humanities, Peter the Great St. Petersburg Polytechnic University.
29 Politekhnicheskaya St., Saint Petersburg 195251 Russian Federation;
e-mail: oskar46@mail.ru

Darya S. Bilyeva –
PhD in Political Science, Associate Professor.
Graduate School of Social Sciences at the Institute of Humanities, Peter the Great St. Petersburg Polytechnic University.
29 Politekhnicheskaya St., Saint Petersburg 195251 Russian Federation;
e-mail: bylieva_ds@spbstu.ru

In contemporary philosophy of science and technology, the rapid development of artificial intelligence (AI) technologies brings to the forefront fundamental questions about the nature of scientific cognition and its transformation under the influence of algorithmic systems. A review of philosophical discussions on intelligent systems in science demonstrates that over three decades there has been a fundamental shift in the point of consensus: whereas at the initial stage they were considered exclusively as a tool for enhancing human cognitive abilities, in 2022–2023 the philosophical community found itself facing the need to legitimize AI as a nontrivial epistemic participant. The traditional question of whether AI can be regarded as a cognitive agent has been supplemented by problems related to the practices of its use: under the influence of big data and neural networks, the criteria of scientificity are being transformed, the subject – object distinction is facing a crisis in the face of algorithmic opacity, and new models of epistemic responsibility are being sought. AI has established itself as a functionally irreducible participant in cognitive practices; however, none of the proposed conceptions – from extended instrumentalism to distributed or collaborative agency – completely eliminates the conceptual tension created by the gap between AI's effective performance of cognitive operations and its lack of intentional states. This makes the direct application of traditional criteria of justification impossible. At the same time, the main discussions remain within the framework of first-order epistemology, focused on the question of the (quasi-)subjectivity of machines. However, this research program, for all its productivity, overlooks a fundamental transformation: AI's acquisition of a metascientific role, in which it not only actively participates in cognition but also largely restructures the configuration of the research field.

Keywords: artificial intelligence, scientific cognition, neural networks, algorithmic transparency, cognitive actors, ethical challenges, automation of science, artificial intelligence in science

Введение

Цифровая эпоха поставила перед философией науки вопрос, который еще полвека назад казался сугубо футурологическим: каким образом системы, не обладающие сознанием, участвуют в производстве того, что мы привыкли считать знанием? Искусственный интеллект (ИИ), первоначально рассматривавшийся как вычислительный инструмент, постепенно занял позицию, которую трудно описать в терминах классической эпистемологии.

Современное развитие технологий искусственного интеллекта (ИИ) представляет собой не просто новый этап в научно-техническом прогрессе, а глубинную трансформацию самих основ научной деятельности, затрагивающую ее эпистемологические, методологические и социальные аспекты. ИИ перестал быть вспомогательным инструментом для обработки больших массивов данных или автоматизации рутинных операций – он становится активным участником познавательных процессов. Этот качественный сдвиг требует переосмысления традиционных представлений о природе научного знания, его методах и критериях валидности, поскольку алгоритмические системы вносят в познавательный процесс новые инструменты, основанные на вероятностных моделях, машинном обучении и обработке данных в масштабах, недоступных человеческому разуму. Целью статьи является выявление основных линий дискуссий о цифровых технологиях и ИИ в науке и демонстрация трансформации представлений, связанных с расширением роли последнего.

Роли ИИ в научном исследовании

Практическое применение ИИ в науке уже сегодня демонстрирует как его потенциал, так и сопутствующие риски: возникают вызовы, связанные с надежностью и интерпретируемостью результатов, полученных с помощью ИИ. Для анализа ИИ необходимо более четко представить, какие функциональные роли он может играть в науке.

Наиболее ранней ролью ИИ в науке можно считать обработку данных и выявление паттернов (классификация, кластеризация, регрессия, извлечение признаков из больших массивов данных). Системы ИИ такого типа выявляют корреляции, структуры и аномалии, которые затем интерпретируются человеком и способствуют развитию новых исследовательских парадигм, таких как data-driven science, где данные и вычислительные модели становятся центральными элементами научного метода. С помощью нейросетевой обработки данных Аргоннской национальной лаборатории удалось сделать ускоренное, масштабируемое и воспроизводимое обнаружение гравитационных волн, которые были предсказаны еще Эйнштейном, но не были обнаружены [Huerta et al., 2021]. Благодаря ИИ значительно продвинулось исследование белков. Обученная на сотнях тысяч экспериментально определенных последовательностей белка нейросеть AlphaFold смогла предсказывать трехмерные структуры с точностью, сопоставимой с экспериментальными методами и значительно превосходящей все иные известные до этого методы [Jumper et al., 2021]. Темп последующих открытий в области молекулярной биологии был также беспрецедентным. Открытие тысячи

потенциально новых семейств белков стало стимулом прогресса в генетических, эволюционных, медицинских исследованиях [Callaway, 2022, p. 236]. В гуманитарных науках нейросети используются для анализа исторических документов, литературных произведений, философских трактатов, выявляя скрытые лексические паттерны, структуру текста, его стилистику и даже подтексты. Так, например, ИИ позволил осуществить анализ литературных произведений на основе применения теории графов и NER-алгоритмов (в т.ч. выявив, в каких отношениях находились свыше 880 наименованных сущностей в книгах Д.Р. Толкина) [Yashchenko, Makarov, 2022]. В исторической науке, кроме оцифровки архивных документов и фондов, технологии ИИ успешно используются для расшифровки древних текстов. Например, Transkribus или языковые модели Google AI позволяют восстанавливать шумерские и ассирийские тексты. Проект Сары Паркак (“National Geographic”) использует машинное обучение для анализа спутниковых изображений и поиска археологических объектов, что помогло обнаружить сотни скрытых поселений в Египте. Проект “Pelagios” связывает древние карты с современными географическими данными. В культурологии и искусствоведении предобученные модели выявляют стилистические особенности, характерные для отдельных художников или художественных школ, что меняет процесс атрибуции и аутентификации произведений, сводя к минимуму субъективность.

Среди новых продуктов, создаваемых с помощью ИИ, можно в первую очередь отметить лекарства и новые материалы. Так, в результате анализа библиотеки химических соединений нейросеть обнаружила новые антибиотики, наиболее интересным из которых является халицин, воздействующий на бактерии, вызывающие псевдомембранозный колит и туберкулез [Stokes et al., 2020, p. 694]. Понятно, что для создания лекарственных средств недостаточно открытия активного вещества, требуется найти структуру лекарства, провести доклинические и клинические испытания. Благодаря работе нейросетей, одна из которых ищет, на что должно быть направлено воздействие для лечения заболевания, а вторая подбирает для этого молекулы, была найдена молекула INS018_055 для борьбы с легочным фиброзом. Искусственный интеллект активно способствует созданию материалов с уникальными свойствами, которые находят применение в широком спектре отраслей – от биомедицины до космической промышленности, эффективно прогнозируя их свойства (биосовместимость, прочность, электро- и теплопроводность и т.д.). Например, нейросеть PauliNet способна определить основные свойства вещества для произвольных молекул [Hermannet al., 2020, p. 893], сочетая подходы квантовой механики с алгоритмами глубокого обучения, что позволяет предсказывать энергетические и структурные характеристики молекул с высокой точностью.

Все более активно ИИ используется для генерации гипотез и новых теорий – в том числе в математических исследованиях [Davies et al., 2021, p. 73] для создания теорий принятия решений [Peterson et al., 2021, p. 1211] и др. Одно из направлений – это автоматизированное доказательство теорем с использованием модели трансформеров. Используя технологию трансформеров, система перебирает все возможные математические операции и построения дерева решений, сокращая время на проверку сложных доказательств и позволяя обнаруживать новые математические теории и решения. Сложные системы способны проектировать эксперименты, выполнять их в автоматизированных лабораториях и итеративно изменять результаты – так называемые «самоуправляемые лаборатории».

В целом для ИИ не существует разницы между прошлым и будущим, между тем, что известно человечеству, и новым. Индифферентность ИИ по отношению к той точке развития науки, на которой находится человечество, означает возможность «знать все». Для чего достаточно, как в фантастических произведениях, задать правильный вопрос или, в контексте научных исследований, вектор приложения сил и нужные базы данных [Быльева, 2025]. Поскольку сегодня ИИ генерирует гипотезы, разрабатывает эксперименты, собирает и интерпретирует большие наборы данных и открывает то, что, «вероятно, было бы невозможно получить с использованием только традиционных научных методов» [Wang et al., 2023, p. 47], можно предположить, что искусственные системы способны к выявлению закономерностей в областях, недоступных для человека, в частности обнаружению причинно-следственных связей между далеко отстоящими друг от друга событиями в исторической перспективе [Анохин и др., 2022, с. 97], а также выявлению прогнозных моделей и закономерностей без проведения реальных опытов.

При этом на более ранних этапах ИИ мог использоваться только институциональными акторами, крупными исследовательскими инфраструктурами. Эпистемическая агентность оставалась институционально опосредованной: доступ к инструментам, данным и вычислительным мощностям определял то, кто может выступать в роли исследователя. Распространение больших языковых моделей (БЯМ) привело к появлению новой массовой роли ИИ – автора, когда любой пользователь может использовать ИИ для написания статей, рецензий и т.п. При этом не существует возможности с абсолютной уверенностью определить, является ли текст сгенерированным, существуют только технические оценки вероятности. Статистика использования слов позволяет предположить, что не менее 10% научных статей созданы с помощью БЯМ. PubMed ежегодно публикует около 1,5 млн статей, из которых не менее 150 000 статей произведены с помощью искусственного интеллекта [Liang et al., 2024]. В марте 2024 г. в издательстве Elsevier была издана статья, начинавшаяся



со слов, хорошо известных всем пользователям ChatGPT: “Certainly, here is a possible introduction for your topic…” («Конечно, вот возможное введение для вашей темы…»), а в статье от июня 2024 г. можно было найти заявление: «Мне очень жаль, но у меня нет доступа к информации в реальном времени или данным о пациенте, поскольку я являюсь языковой моделью искусственного интеллекта». БЯМ могут быть полезны при выполнении рутинных интеллектуальных функций, однако расширяющееся использование ИИ для написания научных текстов значительно повышает риск наличия в статьях так называемых «галлюцинаций» и сфальсифицированных данных. В результате, с одной стороны, повышая скорость обработки данных и помогая составлять структуры, аннотации и подбирая списки литературы, ИИ сокращает время на работу со статьей на 40% [Alkaissi, McFarlane, 2023], с другой, это приводит к стандартизации и стилистической однотипности текстов на основе шаблонности построения фраз ИИ, а в целом к стремительному увеличению доли генерированных текстов в научном дискурсе.

Форматы использования ИИ демонстрируют функциональный плюрализм, не позволяющий зафиксировать его в какой-либо одной эпистемологической категории. Массовизация доступа к генеративным моделям дополнительно размывает институциональные границы научной деятельности. Генерация научных текстов с помощью ИИ приводит к «эпистемической инфляции», когда возможности проверки (экспертная оценка, институциональный контроль и т.п.) не соответствуют скорости создания.

Философская дискуссия до и в «эпоху генеративных моделей»

Философские размышления о науке, основанной на больших данных и интеллектуальных машинах, сформировались до «бума» БЯМ. По мере развития вычислительных систем все чаще возникает вопрос об их статусе и о том, можно ли рассматривать их как когнитивных агентов. Дискуссии о «разрушении субъекта» в науке встраиваются в более широкий контекст философии коллективного познания [List, Pettit, 2011] и споры об «искусственном субъекте». В рамках science and technology studies (STS) агентность понимается как эффект сетевых конфигураций, где нечеловеческие акторы участвуют в производстве знания наравне с людьми, перераспределяя эпистемические функции. Согласно функционализму в философии информации [Bynum, Moor, 1998; Floridi, 2013] любая система, демонстрирующая когнитивные функции (распознавание образов, вывод, обобщение и т.п.), может считаться когнитивным агентом.

Критики функционализма (например, Т. Мецингер) утверждают, что функциональное поведение не подразумевает интенциональности. ИИ не обладает самосознанием, не понимает смыслов и не способен к подлинной рефлексии [Metzinger, 2022]. В конкретно научном контексте вопрос об интеллектуальной деятельности и агентности машин разбивается на подмножество вопросов, связанных с возможностью полноценно участвовать в тех или иных научных процессах, традиционно закрепленных за человеком. В частности, уже с 1990-х гг. появилось понятие «компьютерное открытие» (например, [Langley, 1987; Thagard, 1988]). Однако существовал философский консенсус относительно цифровых технологий как инструментов, которые лишь усиливают когнитивные способности человека, не изменяя при этом структуру эпистемической активности.

Философы науки предпринимали попытки встроить использование «больших данных» в традиционные представления о научном методе. Вероятностная природа машинного обучения, в которой выводы основаны на статистических корреляциях, а не детерминированной логике, внесла дополнительную неопределенность в научные исследования. Одним из эпизодов в философии ИИ в науке стала дискуссия о машинном обучении как форме индукции (в русле старых дебатов между Карлом Поппером и Рудольфом Карнапом). Д. Гиллис утверждал, что существование успешных программ машинного обучения демонстрирует реальность индукции [Gillies, 1996]. Критика тезиса Гиллиса строилась на том, что системы машинного обучения работают на принципах, отличающихся от классических представлений об индукции, опираясь на оптимизацию, статистический вывод и, в случае глубокого обучения, автоматическое извлечение признаков исследуемого объекта. Далее дискуссия о вычислительном подходе к индуктивному выводу перешла от исследования, основанного на гипотезах, к исследованию, основанному на данных [Gillies, Gillies, 2022]. Сторонники «четвертой парадигмы» в науке выдвинули предположение, что корреляция может заменить причинно-следственное объяснение [Hey, 2009] – утверждение, вызвавшее резкую философскую критику [Leonelli, 2012; Pietsch, 2016]. Позже традиционные критерии, такие как воспроизводимость и фальсифицируемость, стали дополняться ИИ-специфическими метриками, такими как вариативность модели при повторном обучении или чувствительность к возмущениям входных данных [Jumper et al., 2021].

С появлением глубоких нейронных сетей, принципы функционирования которых недоступны даже их разработчикам, начались активные дискуссии об алгоритмической непрозрачности и проблеме «черного ящика». «Черный ящик» породил этические противоречия, связанные с ответственностью за решения, принимаемые ИИ. В философии науки в этой связи стали говорить о разделении понимания

и объяснения [Sullivan, 2022; Zednik, 2021], о «коллективной эпистемической агентности» с использованием ИИ [Tollefsen, 2015], «эпистемической непрозрачности» [Humphreys, 2009], стремясь предложить новый ракурс субъектности ученого в производстве научного знания с участием ИИ.

Бум нейросетей в 2022–2025 гг. стал не просто технологическим прорывом, но и катализатором изменений в представлениях об основах человеческого познания. Как пишут К. Митчем и Т. Лю, ИИ может заменить некоторые функции ученых, тогда как ученые не могут заменить некоторые функции ИИ (например, обработать петабайт данных) [Liu, Mitcham, 2024, p. 92]. Реальность в науке изменила точку консенсуса по поводу роли ИИ, «утвердив» его способность создавать новое [Stoliarova, 2025]. Генеративные модели не только классифицируют или предсказывают, но и продуцируют новые данные. Д. Хосрови и Ф. Финн утверждают, что при определенных условиях они могут выполнять функцию de novo synthetic evidence – свидетельства, сопоставимого по эпистемическому статусу с эмпирическими данными [Khosrowi, Finn, 2025].

Тем не менее наиболее дискуссионными остаются вопросы о том, может ли ИИ-система быть субъектом или квазиагентом научного знания, а также каковы эпистемологические последствия и этические аспекты ее участия в научной деятельности. Если первоначально этические принципы требовали отказа от использования ИИ-генерации текстов, то современные исследования предлагают разнообразные модели «эпистемической ответственности» [Lloyd, 2025]. Исследования границ применения ИИ в научном исследовании [Введенская, 2025; Зубковская, 2024; Шалак, 2024] сопровождаются концепциями ИИ как эпистемического посредника или артефактного агента, участвующего в производстве знаний и при этом не обладающего преднамеренными состояниями [Cruz-Aguilar, 2025]. Вместе с тем, не будучи сознательными, действия ИИ имеют эпистемические последствия и значение [Barassi, 2024]. В связи с этим современные исследователи развивают концепцию «масштабно-разделенной агентности» [Palermos, 2025], рассматривают ИИ-системы как эпистемические коллаборации [Greco, 2026], предлагая концепцию эмерджентной когнитивной структуры, в которой ИИ не является ни инструментом, ни автономным агентом, а границы субъекта временно размываются. Значимой проблемой в дискуссиях является феномен «галлюцинаций» ИИ, который, с одной стороны, может служить фактическим подтверждением безразличия генеративных моделей к истине [Herrmann, Levinstein, 2024], с другой, указывать на то, что ИИ создают ощущение ложной активности [Шевченко, 2025] и псевдокреативности нового порядка [Álvarez et al., 2025].

Дискуссии отражают противоречия между эффективным применением ИИ в научных практиках и невозможностью концептуализи-

ровать это участие в привычных эпистемологических категориях. В условиях, когда производство знания оказывается распределенным, гетерогенным и частично непрозрачным для познающих субъектов, трансформация научной рациональности требует особого внимания исследователей.

Метанаучная роль ИИ

Технологии ИИ анализируют современную науку как сложный объект с разветвленной структурой исследований, собственными уникальными социальными практиками и дискурсивными сетями. На основе выявленных трендов они генерируют новые гипотезы, отслеживают перспективные направления исследований, формируют исследовательские программы. ИИ-методы используются для анализа научной литературы в целях выявления тематических кластеров, динамики перехода между различными научными парадигмами, «горячих» и «заброшенных» областей, а также разрывов и лакун в имеющемся знании. Технологии NLP, глубокого обучения и графовые нейросети позволяют строить динамические графы знаний и цитат [Baghel et al., 2025], повышая эффективность прогностических научных моделей. Однако их применение создает замкнутую систему, ограничивающую интеллектуальный поиск, а наукометрические базы «алгоритмически» сужают многообразие подходов и результатов [Бажанов, 2023, с. 200].

Хасок Чанг представляет метанаучный уровень как пространство для критического осмысления науки, где история и философия науки работают в связке, позволяя видеть науку как динамичный процесс решения проблем [Chang, 2004, p. 235–250]. В роли метанаучного наблюдателя ИИ отслеживает изменения в программах исследований, сдвиги в нормативных рамках и накопленных аномалиях, фиксирует структуру современного научного дискурса и практики. Научный ландшафт представляется как гиперграф, учитывающий людей, методы, объекты, цели исследований в качестве основания прогнозной цепи для последующих открытий [Sourati, Evans, 2023]. По мнению Х. Китано, такие возможности анализа и прогнозирования должны существенно расширить возможности и ограничения, сдерживаемые пределами когнитивных возможностей человека [Kitano, 2021]. ИИ-ученый сможет генерировать и проверять огромное множество гипотез, независимо от их значимости и востребованности в научном дискурсе, так как может анализировать структуру цитирований и терминологическую близость для поиска скрытых связей в междисциплинарных исследованиях. С. Фуллер считает, что ИИ способен выявлять «неоткрытое публичное знание» (по Д. Свонсону)



[Delikatnyi, 2025], выступая не как (квази)агент, участвующий в познании наряду с человеком, а как конструктор исследовательского поля.

Заключение

Эволюция роли ИИ в науке от инструмента обработки данных и выявления паттернов до генератора гипотез, соавтора текстов и, в пределе, конструктора исследовательских полей представляет собой функциональную экспансию, которая не укладывается в классическую дихотомию «инструмент vs субъект» и делает несостоятельными попытки зафиксировать ИИ в рамках одной эпистемологической категории.

Философская рефлексия сталкивается с необходимостью пересмотреть базовые эпистемологические различения в ситуации, когда знание производится системами, не обладающими сознательными состояниями, оперирующими «нечеловеческими» способами получения знания и не подпадающими под действие привычных критериев обоснованности. Если классические критерии научности формировались в расчете на интенционально контролируемые процедуры, то в условиях, когда научный результат генерируется частично непрозрачными алгоритмами, ИИ становится не только участником познавательных практик, но и средством анализа науки на метауровне. Однако эта метанаучная функция сопряжена с фундаментальным ограничением: опора на исторические данные и сложившиеся паттерны цитирования создает риск алгоритмического сужения эпистемического многообразия, подмены плюрализма оптимизацией. Представляется, что дальнейшее движение философии науки лежит не в плоскости уточнения статуса ИИ в существующих категориях, а в пересмотре самих этих категорий. Необходим понятийный аппарат, способный описывать распределенные, гетерогенные и частично непрозрачные конфигурации акторов, участвующих в производстве знания.

Список литературы

Анохин и др., 2022 – *Анохин К.В., Новоселов К.С., Смирнов С.К., Ефимов А.Р., Матвеев Ф.М.* Искусственный интеллект для науки и наука для искусственного интеллекта // Вопросы философии. 2022. № 3. С. 93–105. DOI: 10.21146/0042-8744-2022-3-93-105

Бажанов, 2023 – *Бажанов В.А.* Искусственный интеллект, технологии Big Data (больших данных) и особенности современного политического процесса // Философия. Журнал Высшей школы экономики. 2023. Т. 7. № 3. С. 193–210.

Быльева, 2025 – *Быльева Д.С.* Философский вызов и онтолитический эффект искусственного интеллекта. М.: Инфра-М, 2025. 240 с.

Введенская, 2025 – *Введенская Е.В.* Критика искусственного интеллекта: Философская оптика // Технологос. 2025. № 1. С. 87–97. DOI: 10.15593/perm.kipf/2025.1.07.

Зубковская, 2024 – *Зубковская А.А.* Искусственный интеллект и научная объективность // Науковедческие исследования. 2024. № 2. С. 42–49.

Шалак, 2024 – *Шалак В.И.* Избавление от иллюзий ИИ на примере ChatGPT // Technology and Language. 2024. Т. 5. № 2. С. 26–39. DOI: 10.48417/technolang.2024.02.

Шевченко, 2025 – *Шевченко А.А.* «Галлюцинации» ИИ как новая форма эпистемической ошибки // Respublica Literaria. 2025. Т. 6. № 4. С. 93–98.

References

Alkaissi, McFarlane, 2023 – Alkaissi, H., McFarlane, S.I. “Artificial Hallucinations in ChatGPT: Implications in Scientific Writing”, *Cureus*, 2023. https://doi.org/10.7759/cureus.35179

Álvarez et al., 2025 – Álvarez, A., Villalba, M., Dumit, J. “Speculative Trainers: Large Language Models and Techniques of Affirmative Speculation”, *Technology and Language*, 2025, no. 18 (1), pp. 223–251. https://doi.org/10.48417/technolang.2025.01.15

Anokhin, K.V., Novoselov, K.S., Smirnov, S.K., Efimov, A.R., Matveev, F.M. “Iskusstvennyj intellekt dlya nauki i nauka dlya iskusstvennogo intellekta” [Artificial Intelligence for Science and Science for Artificial Intelligence], *Voprosy filosofii*, 2022, no. 3, pp. 93–105. https://doi.org/10.21146/0042-8744-2022-3-93-105. (In Russian)

Baghel et al., 2025 – Baghel, R.R.R.S., Rajput, P.S., Patel, M.K. “The Rise of AI–Driven Scientometrics: A Decade of Transformative Growth (2012–2023)”, *Annals of Library and Information Studies*, 2025, no. 72 (3). https://doi.org/10.56042/alis.v72i3.18487

Barassi, 2024 – Barassi, V. “Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI”, *Harvard Data Science Review*, 2024, no. Special Issue 5. https://doi.org/10.1162/99608f92.ad8ebbd4

Bazhanov, 2023 – Bazhanov, V.A. “Iskusstvennyj intellekt, tehnologii Big Data (bol’shih dannyh) i osobennosti sovremennogo politicheskogo processa” [Artificial Intelligence, Big Data Technologies and Peculiarities of the Modern Political Process], *Filosofija. Zhurnal Vysshej shkoly jekonomiki*, 2023, no. 7 (3), pp. 193–210. (In Russian)

Byl’eva, D.S. *Filosofskij vyzov i ontoliticheskij effekt iskusstvennogo intellekta* [Philosophical Challenge and Ontolythic Effect of Artificial Intelligence]. Moscow: Infra-M Publ., 2025. (In Russian)

Bynum, Moor, 1998 – Bynum, T.W., Moor, J.H. (eds.) *The Digital Phoenix: How Computers Are Changing Philosophy*. Oxford: Blackwell Publ., 1998.

Callaway, 2022 – Callaway, E. “What’s Next for AlphaFold and the AI Protein-Folding Revolution”, *Nature*, 2022, no. 604 (7905), pp. 234–238. https://doi.org/10.1038/d41586-022-00997-5

Chang, 2004 – Chang, H. *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press, 2004, 304 pp.

Cruz-Aguilar, 2025 – Cruz-Aguilar, M.A. “The Epistemic Revolution of AI: Reconfiguring the Foundations of Scientific Knowledge”, *AI & SOCIETY*, 2025. https://doi.org/10.1007/s00146-025-02658-3

Davies et al., 2021 – Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., … Kohli, P. “Advancing Mathematics by Guiding Human Intuition with AI”, *Nature*, 2021, no. 600 (7887), pp. 70–74. https://doi.org/10.1038/s41586-021-04086-x

Delikatnyi, 2025 – Delikatnyi, K. “Interview with Steve Fuller on Social Epistemology and the Future of Education in an Age of Artificial Intelligence”, *Social Epistemology Review and Reply Collective*, 2025, no. 14 (5), pp. 48–51.

Floridi, 2013 – Floridi, L. *The Philosophy of Information*. Oxford: Oxford University Press, 2013.

Gillies, 1996 – Gillies, D. *Artificial Intelligence and Scientific Method*. Oxford: Oxford University Press, 1996, 432 pp.

Gillies, Gillies, 2022 – Gillies, D., Gillies, M. “Artificial Intelligence and Philosophy of Science from the 1990s to 2020”, in: W.J. Gonzalez (ed.). *Current Trends in Philosophy of Science*. Cham: Springer Publ., 2022, pp. 65–79. https://doi.org/10.1007/978-3-031-01315-7_4

Greco, 2026 – Greco, J. “The Transmission of Knowledge via Large-Scale Technology: A Shared Agency Account”, *Social Epistemology*, 2026, no. 40 (1), pp. 87–100. https://doi.org/10.1080/02691728.2025.2463054

Hermann et al., 2020 – Hermann, J., Schätzle, Z., Noé, F. “Deep-Neural-Network Solution of the Electronic Schrödinger Equation”, *Nature Chemistry*, 2020, 12, pp. 891–897. https://doi.org/10.1038/s41557-020-0544-y

Herrmann, Levinstein, 2024 – Herrmann, D.A., Levinstein, B.A. “Standards for Belief Representations in LLMs”, *Minds and Machines*, 2024, no. 35 (1), art. 5. https://doi.org/10.1007/s11023-024-09709-6

Hey et al., 2009 – Hey, T., Tansley, S., Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research, 2009.

Huerta et al., 2021 – Huerta, E.A., Khan, A., Huang, X., Tian, M., Levental, M., Chard, R., … Foster, I. “Accelerated, Scalable and Reproducible AI–Driven Gravitational Wave Detection”, *Nature Astronomy*, 2021, no. 5 (10), pp. 1062–1068. https://doi.org/10.1038/s41550-021-01405-0

Humphreys, 2009 – Humphreys, P. “The Philosophical Novelty of Computer Simulation Methods”, *Synthese*, 2009, no. 169 (3), pp. 615–626. https://doi.org/10.1007/s11229-008-9435-2

Jumper et al., 2021 – Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. “Highly Accurate Protein Structure Prediction with AlphaFold”, *Nature*, 2021, no. 596 (7873), pp. 583–589. https://doi.org/10.1038/s41586-021-03819-2

Khosrowi, Finn, 2025 – Khosrowi, D., Finn, F. “Can Generative AI Produce Novel Evidence?”, *Philosophy of Science*, 2025, pp. 1–12. https://doi.org/10.1017/psa.2025.10123

Kitano, 2021 – Kitano, H. “Nobel Turing Challenge: Creating the Engine for Scientific Discovery”, *npj Systems Biology and Applications*, 2021, no. 7 (1), art. 29. https://doi.org/10.1038/s41540-021-00189-3

Langley et al., 1987 – Langley, P., Simon, H.A., Bradshaw, G.L., Zytkow, J.M. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: MIT Press, 1987.

Leonelli, 2012 – Leonelli, S. “Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies”, *International Studies in the Philosophy of Science*, 2012, no. 26 (1), pp. 47–65. https://doi.org/10.1080/02698595.2012.653119

Liang et al., 2024 – Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., … Zou, J.Y. “Mapping the Increasing Use of LLMs in Scientific Papers”, *arXiv*, 2024, no. 2404.01268.

List, Pettit, 2011 – List, C., Pettit, P. *Group Agency: The Possibility, Design and Status of Corporate Agents*. Oxford: Oxford University Press, 2011.

Liu, Mitcham, 2024 – Liu, T., Mitcham, C. “Toward Practical Hermeneutics of Fourth Paradigm AI for Science”, *Technology and Language*, 2024, no. 14 (1), pp. 89–105. https://doi.org/10.48417/technolang.2024.01.07

Lloyd, 2025 – Lloyd, D. “Epistemic Responsibility: Toward a Community Standard for Human-AI Collaborations”, *Frontiers in Artificial Intelligence*, 2025, no. 8. https://doi.org/10.3389/frai.2025.1635691

Metzinger, 2022 – Metzinger, T. “Towards a Global Artificial Intelligence Charter”, in: S. Voeneky, P. Kellmeyer, O. Mueller, W. Burgard (eds.). *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*. Cambridge: Cambridge University Press, 2022, pp. 167–175. https://doi.org/10.1017/9781009207898.013

Palermos, 2025 – Palermos, S.O. “Knowledge from AI”, *Social Epistemology Review and Reply Collective*, 2025, no. 14 (11).

Peterson et al., 2021 – Peterson, J.C., Bourgin, D.D., Agrawal, M., Reichman, D., Griffiths, T.L. “Using Large-Scale Experiments and Machine Learning to Discover Theories of Human Decision-Making”, *Science*, 2021, no. 372 (6547), pp. 1209–1214. https://doi.org/10.1126/science.abe2629

Pietsch, 2016 – Pietsch, W. “The Causal Nature of Modeling with Big Data”, *Philosophy & Technology*, 2016, no. 29 (2), pp. 137–171. https://doi.org/10.1007/s13347-015-0202-2

Shalak, V.I. “Izbavlenie ot illyuzij II na primere ChatGPT” [Delusion of AI on the Example of ChatGPT], *Technology and Language*, 2024, no. 5 (2), pp. 26–39. https://doi.org/10.48417/technolang.2024.02. (In Russian)

Shevchenko, A.A. “Gallyutsinatsii II kak novaya forma epistemicheskoj oshibki” [AI ‘Hallucinations’ as a New Form of Epistemic Error], *Respublica Literaria*, 2025, no. 6 (4), pp. 93–98. (In Russian)

Sourati, Evans, 2023 – Sourati, J., Evans, J.A. “Accelerating Science with Human-Aware Artificial Intelligence”, *Nature Human Behaviour*, 2023, no. 7 (10), pp. 1682–1696. https://doi.org/10.1038/s41562-023-01648-z



Stokes et al., 2020 – Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., … Collins, J.J. "A Deep Learning Approach to Antibiotic Discovery", *Cell*, 2020, no. 180 (4), pp. 688-702, e13. https://doi.org/10.1016/j.cell.2020.01.021

Stoliarova, 2025 – Stoliarova, O. "Techno-Contexts and the Birth of Novelty: Questioning the AI on Hermeneutics", *Technology and Language*, 2025, no. 6 (2), pp. 151–160. https://doi.org/10.48417/TECHNOLANG.2025.02.13

Sullivan, 2022 – Sullivan, E. "Understanding from Machine Learning Models", *British Journal for the Philosophy of Science*, 2022, no. 73 (1), pp. 109–133. https://doi.org/10.1093/bjps/axz035

Thagard, 1988 – Thagard, P. *Computational Philosophy of Science*. Cambridge, MA: MIT Press, 1988, 240 pp.

Tollefsen, 2015 – Tollefsen, D. *Groups as Agents*. Cambridge: Polity Press, 2015, 167 pp.

Vvedenskaya, E.V. "Kritika iskusstvennogo intellekta: Filosofskaya optika" [Critique of Artificial Intelligence: A Philosophical Optics], *Tekhnologos*, 2025, no. 1, pp. 87–97. https://doi.org/10.15593/perm.kipf/2025.1.07 (In Russian)

Wang et al., 2023 – Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., … Zitnik, M. "Scientific Discovery in the Age of Artificial Intelligence", *Nature*, 2023, no. 620 (7972), pp. 47–60. https://doi.org/10.1038/s41586-023-06221-2

Yashchenko, Makarov, 2022 – Yashchenko, A., Makarov, I. "Knowledge Retrieval and Relation Mining from Tolkien's History of Middle Earth", in: *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI–MACRo)*. New York: IEEE, 2022, pp. 257–264. https://doi.org/10.1109/CINTI-MACRo57952.2022.10029506

Zednik, 2021 – Zednik, C. "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence", *Philosophy & Technology*, 2021, no. 34 (2), pp. 265–288. https://doi.org/10.1007/s13347-019-00382-7

Zubkovskaya, A.A. "Iskusstvennyj intellekt i nauchnaya ob"ektivnost'" [Artificial Intelligence and Scientific Objectivity], *Naukovedcheskie issledovaniya*, 2024, no. 2, pp. 42–49. (In Russian)

Памятка для авторов

- Автор гарантирует, что текст, представленный для публикации в журнале, не был опубликован ранее или сдан в другое издание. При использовании материалов статьи в последующих публикациях ссылка на журнал «Эпистемология и философия науки» обязательна.
- Автор берет на себя ответственность за точность цитирования, правильность библиографических описаний, транскрибирование имен и фамилий.
- Рукописи принимаются исключительно в электронном виде в формате MS Word (шрифт – Times New Roman; размер – 12; междустрочный интервал – одинарный; абзацный отступ – 0,9; выравнивание – по левому краю; поля – 2,5 см) на сайте журнала https://journal.iphras.ru
- Объем статьи – от 0,75 до 1,2 а.л. (включая ссылки, примечания, список литературы, аннотации). Объем аналитического обзора – 0,5 а.л.
- Примечания оформляются как постраничные сноски со сквозной нумерацией. Библиографические сведения, отсылающие к Списку литературы, даются в основном тексте и в примечаниях в квадратных скобках; например: [Сидоров, 1994, с. 25–26]. На все источники из цитируемой литературы должны быть ссылки в тексте статьи.
- Помимо основного текста статьи, рукопись должна включать в себя следующие **сведения на английском и русском языках:**

  1) Ф.И.О. автора; ученую степень и ученое звание; место работы; полный адрес места работы (включая страну, индекс, город); адрес электронной почты автора;

  2) название статьи;

  3) аннотацию (1500–2000 знаков);

  4) ключевые слова (до 10 слов и словосочетаний);

  5) список литературы.
- Рукописи на русском языке должны содержать два варианта списка литературы:

  1. «**Список литературы**», выполненный в соответствии с требованиями ГОСТа. В начале списка в алфавитном порядке указываются источники на русском языке.

  2. Список «**References**», составленный в соответствии с требованиями международных библиографических баз данных (Scopus и др.). Все библиографические ссылки на русскоязычные источники приводятся в латинском алфавите по следующей схеме:

  – автор (имена отечественных авторов – в транслитерации латиницей, имена зарубежных авторов – в оригинальном или англоязычном написании);

  – заглавие статьи (транслитерация);

  – [перевод заглавия статьи на английский язык в квадратных скобках];

– название русскоязычного источника (транслитерация);

– [перевод названия источника на английский язык в квадратных скобках];

– выходные данные на английском языке (включая общее количество страниц в источнике или номера страниц, на которых размещен текст в сборнике/журнале/монографии).

- Для транслитерации необходимо использовать сайт http://translit.net/ (формат BSI).
- Подробные рекомендации по оформлению текстов содержатся на странице журнала: https://journal.iphras.ru/forcontributors
- К рукописи также должна прилагаться фотография автора.
- Рисунки и формулы должны быть продублированы в графическом режиме и записаны отдельным файлом. Тексты, содержащие специфические символы и неевропейские шрифты, должны быть продублированы в формате pdf.
- Решение о публикации материала принимается членами редколлегии, главным редактором и рецензентами в течение полугода с момента поступления текста в редакцию.
- Плата за публикацию материалов не взимается, гонорар авторам не выплачивается.
- Адрес редакции: Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1, оф. 315. Тел.: +7 (495) 697-95-76; e-mail: journal@iphras.ru; сайт: https://journal.iphras.ru

Научно-теоретический журнал

Epistemology & Philosophy of Science / Эпистемология и философия науки
2026. Том 63. Номер 3

Учредитель и издатель: Федеральное государственное бюджетное учреждение науки Институт философии Российской академии наук

Свидетельство о регистрации СМИ: ПИ № ФС77-57113 от 3 марта 2014 г.

Главный редактор *И.Т. Касавин*
Зам. главного редактора: *А.Ю. Антоновский, И.А. Герасимова, П.С. Куслий*
Ответственный секретарь *Л.А. Тухватулина*

Художники: *Ч.Р. Кантов, С.Ю. Растегина*
Корректор *И.А. Мальцева*
Технический редактор *Е.А. Морозова*

Подписано в печать с оригинал-макета 07.09.26
Формат 60×100 1/16. Печать офсетная. Гарнитура IPH Astra Serif
Усл. печ. л. 16,37. Уч.-изд. л. 15,85. Тираж 1 000 экз. Заказ № 16

Оригинал-макет изготовлен в Институте философии РАН
Компьютерная верстка: *Е.А. Морозова*

Отпечатано в ЦОП Института философии РАН
109240, г. Москва, ул. Гончарная, д. 12, стр. 1

Информацию о журнале «Эпистемология и философия науки»
см. на сайте: https://journal.iphras.ru