

Оригинальная статья

УДК 165.2

<http://doi.org/10.32603/2412-8562-2022-8-2-17-27>

Машины и человеческая эпистемология

Тимофей Сергеевич Дёмин¹, Константин Геннадьевич Фролов^{2✉}

^{1, 2}Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

²НИУ «Высшая школа экономики», Москва, Россия

¹detimofei@gmail.com, <https://orcid.org/0000-0002-0973-2756>

²kgfrollov@etu.ru, <https://orcid.org/0000-0002-9071-6138>

Введение. В статье анализируется структура эпистемологии знаний слабого искусственного интеллекта в сопоставлении со структурой эпистемологии знаний человека. Актуальность проблематики подобного рода объясняется востребованностью одного из наиболее динамично развивающихся разделов современной философии – философии искусственного интеллекта. Новизна исследования связана с идеей применения инструментов анализа, разработанных в рамках аналитической эпистемологии, к области искусственного интеллекта (ИИ).

Методология и источники. Статья написана в рамках аналитической традиции в философии. Для исследования существенных аспектов понятия знания используется концептуальный анализ, предполагающий, что объяснение сложного явления требует выявления компонентов, из которых оно состоит. Этот метод лег в основу аналитической дискуссии о знании второй половины XX в. Также в статье применяется сравнительный анализ.

Результаты и обсуждение. Знания людей отличаются тремя существенными характеристиками – наличием информационно-несущих ментальных состояний, их надежностью и фактивностью. Было проанализировано, насколько информационно-несущие внутренние состояния слабого ИИ соответствуют этим качествам. Был сделан вывод, что внутренние состояния могут считаться убеждениями при слабой трактовке, иметь высокую степень надежности при определенных условиях и способны обладать фактивностью в случае, если принять, что слабый ИИ обладает убеждениями.

Заключение. Слабая трактовка понятия убеждения позволяет утверждать, что нейронные сети способны иметь убеждения. Более строгая трактовка понятия убеждения включает также требование понимания смысла. Однако в нашем распоряжении нет удовлетворительной теории понимания смысла. При этом условие надежности представляет собой единственный из критериев знания, которому могут соответствовать функциональные состояния машин в случае некоторых задач, в связи с чем особенно остро встает проблема генеральности.

Ключевые слова: эпистемология, знание, убеждение, искусственный интеллект, эпистемология искусственного интеллекта, условие надежности

Финансирование: работа выполнена при финансовой поддержке гранта Президента РФ для молодых ученых – кандидатов наук (проект МК-703.2021.2 «Натуралистическая эпистемология

© Дёмин Т. С., Фролов К. Г., 2022



Контент доступен по лицензии Creative Commons Attribution 4.0 License.
This work is licensed under a Creative Commons Attribution 4.0 License.

агентности и моральной ответственности при разработке и использовании технологий дополненного интеллекта»).

Для цитирования: Дёмин Т. С., Фролов К. Г. Машины и человеческая эпистемология // ДИСКУРС. 2022. Т. 8, № 2. С. 17–27. DOI: 10.32603/2412-8562-2022-8-2-17-27.

Original paper

Machines and Human Epistemology

Timofei S. Demin¹, Konstantin G. Frolov²

^{1,2}*Saint Petersburg Electrotechnical University, St Petersburg, Russia*

²*Higher School of Economics, Moscow, Russia*

¹*detimofei@gmail.com, <https://orcid.org/0000-0002-0973-2756>*

²*kgfrolov@etu.ru, <https://orcid.org/0000-0002-9071-6138>*

Introduction. The article analyzes the structure of the epistemology of knowledge of weak artificial intelligence in comparison with the structure of human knowledge. The article was written within the framework of one of the most demanded branch sections of philosophy – the philosophy of artificial intelligence. The novelty of the research is based on the idea of applying the analysis of knowledge developed within the framework of analytical epistemology to the field of artificial intelligence.

Methodology and sources. The article is written in the framework of analytical tradition in philosophy. We use conceptual analysis to analyze the most crucial aspects of knowledge. This method assumes that in order to explain a complex phenomenon, it is necessary to analyze its components. This method formed the basis of an analytical discussion about knowledge in the second half of the 20th century. The article also uses comparative analysis.

Results and discussion. People's knowledge is characterized by three essential characteristics – the presence of information-bearing mental states, their reliability, and factuality. We analyzed to what extent the information-carrying internal states of a weak AI correspond to these qualities. The authors concluded that internal states can be considered beliefs if they are weakly interpreted, they can have a high degree of reliability under certain conditions, and they can have factuality if we accept that a weak AI has beliefs.

Conclusion. A weak interpretation of the concept of belief allows us to argue that neural networks are capable of having beliefs. A more rigorous interpretation of the concept of belief also includes the requirement to understand the meaning. However, we do not have at our disposal a satisfactory theory of understanding meaning. In this case, the condition of reliability is the only criterion of knowledge to which the functional states of machines can correspond in the case of certain tasks, in connection with which the problem of generality arises especially acutely.

Keywords: epistemology, knowledge, belief, artificial intelligence, epistemology of artificial intelligence, safety condition

Source of financing: the work was supported by a grant of the President of the Russian Federation (project MK-703.2021.2 “Naturalistic epistemology of agency and moral responsibility in the development and use of augmented intelligence technologies”).

For citation: Demin, T.S. and Frolov, K.G. (2022), “Machines and Human Epistemology”, *DISCOURSE*, vol. 8, no. 2, pp. 17–27. DOI: 10.32603/2412-8562-2022-8-2-17-27 (Russia).

Введение. Целью данного исследования является определение условий применимости понятия знания в отношении слабого искусственного интеллекта. Анализ предполагаемых знаний машин будем проводить путем сопоставления возможностей слабого искусственного интеллекта с возможностями людей и с тем, как употребляется понятие знания применительно к людям. Чтобы провести этот анализ, сначала потребуется показать, как вообще возможно познание явлений одного порядка через явления другого порядка, т. е. познание путем сопоставления разнородного. Так, природу людей можно понимать через природу животных или машин. Однако справедлив и обратный ход: исследуя то, чем обладают люди, мы можем лучше понять, на каком уровне эпистемического развития находится слабый искусственный интеллект. В данной статье с помощью эпистемологии людей будет проанализирована эпистемология машин.

Методология и источники. Человек может понимать себя через сущности другого порядка. Эти сущности выступают в роли зеркала, через которое можно увидеть себя. Непрямой путь превращения инаковой по отношению к человеку сущности в зеркало – это метафора или аналогия. Другой важный инструмент для анализа и понимания реальности – метонимия. Так, Лакофф и Джонсон полагают, что метафора и метонимия – это то, из чего в существенной мере состоит все наше познание мира [1, 2]. Искусственный интеллект с самого начала стал зеркалом для человека, предлагая образы, аналогии и прямые соответствия между человеком и машиной от простой метафоры мозга как компьютера до разнообразных по форме воплощений сильного искусственного интеллекта в литературе, играх и кино.

Подобно животному машина в роли зеркала может давать ответы на философские вопросы и через отсутствующие у нее человеческие качества. Исходя из того, чего у машины нет, мы лучше понимаем, что характеризует человека как особенное существо. Так, Джон Серль через мысленный эксперимент с китайской комнатой демонстрирует отсутствие у машины понимания, которое у человека оказывается сложным и не до конца понятным феноменом [3].

Наиболее типичный пример слабого искусственного интеллекта представляют собой нейронные сети, которые реализуют некоторый навык. В некотором роде слабый ИИ знает, как водить автомобиль, отвечать осмысленно на некоторые вопросы в роли голосового ассистента, осуществлять поиск в Интернете, опознавать объекты на картинках, играть в Го, шахматы и другие игры. Сложности начинаются с пропозициональным знанием. Чего не хватает машине, чтобы знать, что Эверест – самая высокая гора или что кот сидит на ковре?

Если машина будет разумной в том же смысле, что и человек, то нет никакой проблемы на пересечении эпистемологии естественного и искусственного типов интеллекта: эпистемология разумной машины в таком случае аналогична эпистемологии человека. Машина сможет обладать знанием в том же смысле, в котором им обладают люди, когда у нее будет сознание, которое обладает достаточным уровнем понимания.

Однако кажется правдоподобным сценарий, при котором разум машины окажется по своей природе, своему устройству и своим проявлениям радикально отличен от человеческого [4]. Такие машины будут плохим зеркалом для понимания человека. Хотя остается неясным, в чем будет заключаться их инаковость кроме совершенной памяти, высокого уровня владения навыками (вроде игры в шахматы или распознавания визуальных объектов) и сверх-

быстрой скорости мыслительных процессов. Эта инаковость хотя и существенна, но все же не выглядит достаточно радикальной, чтобы сопоставление разума человека и машины оказалось совершенно невозможным и не имеющим под собой никаких оснований.

Даже если машины достигнут такого уровня развития, этому уровню будут предшествовать многочисленные этапы развития искусственного интеллекта. Если согласиться, что машины получают интеллектуальные способности, сопоставимые и схожие с человеческими, то, скорее всего, через этапы, которые релевантны для человеческой эпистемологии. Рост понимания машин будет соответствовать усложнению внутренних состояний машин, и каждый новый уровень развития будет приближать их к знанию, которым обладают люди. Какие этапы в развитии познания релевантны для машины? Чем эпистемология машин, скорее всего, будет отличаться от эпистемологии людей? Как будет выглядеть знание машин, если они превзойдут людей? Можно ли предложить классификацию уровней приближения к знанию у машин? Как стоит дополнить эпистемологию людей исходя из того, чем важным не обладают существующие уже сейчас слабые версии искусственного интеллекта? Наконец, чем эпистемология машин может быть полезна для эпистемологии людей? Все эти вопросы являются крайне актуальными для современных философских дискуссий в данной области.

Результаты и обсуждение. В зарубежной литературе традиционно признается, что понятие знания включает в себя три существенных аспекта: наличие определенного ментального состояния, истинность и условие надежности [5]. Далее, по отдельности рассмотрим, как эти аспекты проявляются у слабого искусственного интеллекта на уровне аналогии и что требуется машинам, чтобы их эпистемология стала сопоставима или превосходила по сложности своей структуры эпистемологию людей. Мы покажем, что машины способны обладать репрезентационным несущим внутренним состоянием (аналогом убеждения), которое может характеризоваться истинностью и эпистемической надежностью. Этого достаточно, чтобы заключить, что при определенной трактовке этих аспектов пропозиционального знания машины им уже обладают.

Существует два базовых сценария в развитии искусственного интеллекта [6]. Первый из них предполагает возможность создания очень мощного и сложноорганизованного компьютера, который будет обладать сознанием. Второй предположительно возможный способ исходит из того, что наука сможет достичь впечатляющего прогресса в усилении способностей человеческого мозга. Во втором сценарии мы сразу переходим к тому, что искусственный интеллект будет обладать знанием, которое превосходит обычное человеческое. Однако в рамках данной статьи сосредоточимся на исследовании перспектив первого сценария. При этом выводы, к которым мы приходим, применимы также и ко второму.

Исходя из первого сценария, вполне естественно предположить, что разум искусственного интеллекта будет радикально отличаться от разума человека, поэтому описание его внутренних состояний будет устроено по иным принципам, и человеческая эпистемология будет к нему неприменима. Это очень важное замечание справедливо только по отношению к внутренним состояниям, однако то, что отвечает за надежность (reliability) машинной эпистемологии, мы сможем оценивать так же, как оцениваем надежность знаний людей. Например, в терминах вероятностей и истины.

Знание может принимать разные формы. Так, философы различают пропозициональное знание, которое сообщает о положении дел в мире, и знание-как, которое обозначает умение или навык создавать желаемые положения дел. Существуют дискуссии, в которых философы пытаются решить, являются ли умения и навыки разновидностью пропозиционального знания [6]. Оставим эти споры, хотя позиции в них нетривиальным образом влияют на возможные выводы по исследуемой нами проблеме. Так, слабый ИИ обладает знанием того, как водить автомобиль, и в обозримом будущем он будет водить его лучше человека. Однако в рамках данной статьи сосредоточимся на рассмотрении именно пропозиционального знания. Во всех популярных теориях оно характеризуется наличием информационно-несущего состояния, которое способно быть истинным и надежным (safety). Принципиальная разница между имеющимися подходами состоит лишь в том, как именно эти характеристики оказываются включены в понятие знания. В редуктивных теориях они представляют собой отдельные независимые элементы, тогда как в нередуктивных подходах они включены в информационное состояние, которое выходит за пределы тела [7]. Рассмотрим эти аспекты знания по порядку.

Первый важнейший элемент знания – это наличие информационно-несущего состояния. Когда человек видит кота на ковре, в его голове формируется визуальный образ. Когда слабый искусственный интеллект распознает картинку, загруженную в поисковое окно, он приобретает способность описать картинку в виде кода и привести в соответствие изображение и описание картинки с самой картинкой.

Ортодоксальные теории знания многокомпонентны. Это означает, что наличие информационного состояния (убеждения) представляет собой только один из аспектов понятия знания (хотя в редуктивных теориях необходимость убеждения как такового оказывается вообще под вопросом [8]).

Наиболее универсальное определение понятия убеждения таково: S убежден, что X, если и только если S думает об X с принятием [9]. Кажется, машина может обладать такими убеждениями, если процесс мышления и принятия мы будем понимать достаточно широко. Например, если мышление представляет собой любой функционально достаточно сложный внутренний процесс обработки информации, с которым может справиться из биологических организмов только такой, который имеет развитый мозг уровня людей или высших млекопитающих. Под такое определение с *ad hoc* критерием «уровня людей или высших млекопитающих» подпадает машина, которая способна выполнять функциональные операции высокого уровня сложности – например, водить машину. В таком случае в функциональном смысле иногда машины способны к осуществлению мыслительной деятельности. Если же под принятием X мы будем понимать обращение с X так, как если бы X было истинно, то машине также по силам удовлетворить такому критерию. С этой задачей справляется, например, голосовой ассистент банка, когда клиент спрашивает, как купить билет с кэшбэ-ком, а ассистент дает лаконичный и точный ответ-инструкцию. Ассистент дает ответ так, как если бы он *знал*, что этот ответ верен.

Таким образом, наиболее распространенный взгляд на убеждение состоит в том, что это ментальная репрезентация, которая имеет своим содержанием некоторое (возможное или действительное) положение дел. Существует отдельная дискуссия о том, какого уровня

должен быть интеллект для того, чтобы обладать убеждениями. Так, в приматологии ведется дискуссия о том, какого рода ментальными репрезентациями оперируют высшие приматы [10]. С другой стороны, существует известная линия мысли Д. Дэвидсона [11], которая направлена на защиту идеи о том, что животные и дети не обладают убеждениями, поскольку они не владеют языковыми способностями. При этом аргументы Дэвидсона и известные возражения на них оказываются вполне применимы к дискуссии об искусственном интеллекте.

Чего же не хватает машинам для обладания убеждениями? Здесь можно выделить три линии рассуждения.

Первая линия, родственная Ф. Дредске или Д. Дэвидсону, предполагает, что искусственному интеллекту не хватает обширной сети из репрезентаций, а сами репрезентации имеют недостаточный уровень сложности. Эта линия не содержит онтологической разницы между убеждениями машин и людей. Вторая линия рассуждений, характерная для некоторых функционалистов, например Дэниела Деннета [12], предполагает, что машины уже обладают убеждениями. Третья линия исходит из того, что компьютерные процессы в принципе не могут порождать убеждения.

Второй элемент понятия знания – это способность информационно-несущих состояний характеризоваться в качестве истинных. При этом теории знания, как правило, опираются на корреспондентную теорию истины [13]. Соответственно, в рамках современных версий этого подхода выделяют, во-первых, *носитель истинности* (truth bearer), т. е. то, что, собственно, может быть корректно охарактеризовано в качестве истинного или ложного в рамках нашей модели. В зависимости от конкретной теории эту роль могут выполнять высказывания, пропозиции или информационно-несущие ментальные состояния. Вторым элементом концепции соответствия выступает *фактор истинности* (truth maker), который представляет собой внешнюю по отношению к информационному состоянию знания сущность, в силу наличия которой в актуальном мире тот или иной потенциальный носитель истинности действительно обладает свойством истинности. На роль таких факторов истинности претендуют разные сущности в зависимости от теории. Например, это могут быть объекты, их мереологические суммы, факты или положения дел. И, наконец, имеется само *отношение соответствия*, т. е. отношение между фактором и носителем истинности. Отдельная проблема внутри корреспондентных теорий заключается в выявлении природы этого соответствия. В качестве такового может выступать структурная изоморфность, подобие или особого рода различимость.

Дискуссию о природе соответствия, носителя и фактора истинности в данном случае не рассматриваем. Однако остается открытым вопрос, насколько сложной должна быть функциональная структура системы обработки информации, чтобы быть способной выполнять интересующее нас соответствие в том случае, если на роль носителя истинности мы выбираем информационно-несущее состояние системы. Возможно, что для этого вполне достаточно возможностей сравнительно простой когнитивной системы, которая при этом, с одной стороны, сохраняет подобие, а с другой – различимость. Примером такого рода может быть машина, которая способна определять, к какому фактору истинности относится то или иное ее информационно-несущее состояние, а также отличать этот фактор от других, к которым ее

информационное состояние не относится. Если это – минимальное требование, то, например, нейронные сети Яндекса по определению картинок в отдельных ситуациях обладают истинными информационно-несущими состояниями.

Третий элемент понятия знания – это выполнение условия надежности [14]. Общая идея условия надежности знания такова: S знает, что P, только если P не может ошибаться легко [15, p. 12].

Разумно предположить, что машина вне зависимости от своего устройства, чтобы что-то знать, должна соответствовать условию надежности. Например, если она определяет объекты визуально и фиксирует на картинке кота, то она не должна ошибаться в схожих ситуациях, если ей показать еще десять картинок с котами.

У условия надежности есть определенные рамки, в пределах которых оно должно выполняться. Так, чтобы знать, что на картинке изображен кот, не нужно справляться с опознанием кота в кромешной тьме, открытом космосе, взрывах, землетрясениях и пр. При этом существенно то, что ситуации должны быть схожими. Но до какой степени? Вполне возможно, что этот вопрос всякий раз будет решаться на прагматической основе. В таком случае машина должна не ошибаться в схожих ситуациях в достаточной мере, чтобы она справлялась с решением тех задач, для которых была разработана и сконструирована. Однако это решение не отвечает на вопрос о знании.

Другая проблема состоит в том, насколько широко нам следует определить способность или метод, по которому машина обеспечивает надежность знания? Эта проблема называется проблемой генеральности (*generality problem*) [16]. Можно определить эту способность довольно широко, например, как способность визуального распознавания объектов. Однако эта способность все-таки не предполагает необходимой для знания степени надежности. Можно также задать способность в узком смысле как способность опознать кота на картинке K1 во время T1 в ситуации C1. Однако в таком случае надежность способности определяется тем, является ли сформулированная в результате ее применения пропозиция истинной. Если да, то способность является надежной. Однако такой ответ также не может нас устроить, поскольку такая способность может давать нам истинный результат просто в силу совпадения. Фелдман называет такое последствие «беспокойством единичного случая». Именно с этим типом беспокойства сталкиваются нейронные сети. У нас есть способность слабого искусственного интеллекта как исходная данность, и нам надо оценить, является ли она достаточно общей.

Интересно при этом, что способности ИИ являются более конкретными, чем способности людей. Нейронная сеть может гораздо лучше человека опознавать картинки, но неожиданное отклонение в сценарии, которое для человека не было бы проблемой, для машины окажется непреодолимой ситуацией. Однако нам самим трудно оценить зависимость собственных способностей от среды. То, что человек более гибок в применении методов и в адаптации к изменению среды, представляет собой его несомненное эпистемическое преимущество. Однако для того, чтобы это преимущество стало метафизически значимым отличием мышления человека от возможных версий компьютерного мышления, необходимо более сильное доказательство.

Возможно, что на роль такого метафизически значимого преимущества может претендовать способность понимания смысла. Ведь если бы машина могла понимать смыслы, то

ее внутренние информационно-несущие состояния было бы гораздо легче признать убеждениями, которые могут быть истинными. Однако в рамках аналитической эпистемологии нет консенсуса по поводу ответа на вопрос, что именно представляет собой способность понимания.

Один из возможных ответов состоит в том, что понимание – это разновидность знания. А именно, понимание X – это знание причин X . В этом подходе мы имеем дело с причинным пониманием, которое следует отличать от общего понимания. Пример причинного понимания таков: Игорь понимает, почему у электрона отрицательный заряд. В свою очередь, пример общего понимания таков: Игорь понимает квантовую механику. В этом подходе нам следует через знание объяснять понимание, а не наоборот. Непривлекательность этого подхода, помимо прочего, для нашей темы заключается в том, что формулировка «Машине не хватает понимания, чтобы обладать знанием» оказывается некорректной. Другой ответ на вопрос о понимании предполагает, что понимание представляет собой вид когнитивного достижения [17], которым мы способны обладать благодаря наличию у нас надежного навыка.

Наконец, третья идея состоит в том, что понимание – это компонент знания. Чтобы что-то знать, нужно понимать смысл репрезентации. Если не понимать, что такое кот, то нельзя знать, что «кот на ковре». Нельзя быть убежденным, что P и не понимать смысл P . В пропозициональном знании пропозиция – это смысл, который мы понимаем по определению, так как иначе пропозиция была бы нам недоступна.

Можно заметить, что понимание смысла и понимание причин – это разные типы понимания. Когда я говорю, что понимаю, что значит «кот на ковре», я не имею в виду, что я понимаю, что кот на ковре (во втором случае «понимаю» взаимозаменяемо со «знаю»). Понимание смысла не означает понимание причин, по которым этот смысл возник, или причин, по которым я понимаю нечто так, а не иначе. Однако в нашем случае нас интересует именно понимание смысла в качестве такой когнитивной способности, которой не хватает машинам для обладания знанием.

При этом данный вопрос представляет собой не только философский интерес, его можно также адресовать и когнитивным наукам. Что происходит на психологическом уровне, когда человек понимает значение?

Предложим только один из ответов: человек соотносит высказывание с понятиями базового уровня, которыми он владеет с помощью опыта. Таким образом, любое высказывание, имеющее смысл, может быть соотнесено с опытными данными. Кот – с визуальным опытом других котов, ковер – с парадигмальными примерами опыта созерцания ковров, отношение «на» – через пространственное отношение других объектов [1, с. 380–381]. Эти базовые единицы опыта при помощи воображения переносятся для структурирования и понимания другого опыта. После переноса становится возможным понимание. Такой ответ предполагает, во-первых, принятие эмпиризма – тезиса, что познание возможно только *a posteriori*, и, во-вторых, холизма – тезиса, что содержание убеждения зависит от множества других убеждений.

Заключение. В статье проведено сопоставление структуры понятия знания, применимого к машинам, с понятием знания, которое мы применяем в отношении людей. Наше

соотнесение продемонстрировало, что позволяет в слабом смысле утверждать: нейронные сети могут обладать информационными состояниями, способными быть истинными. Более строгая теория требует понимания смысла. При этом условие надежности представляет собой единственный из критериев, который машина может выполнять в случае некоторых задач, в связи с чем особенно остро встает проблема генеральности.

СПИСОК ЛИТЕРАТУРЫ

1. Лакофф Д. Женщины, огонь и опасные вещи. Что категории языка говорят нам о мышлении / пер. с англ. И. Б. Шатуновского. М.: Языки славянской культуры, 2004.
2. Лакофф Д., Джонсон М. Метафоры, которыми мы живем / пер. с англ. А. Н. Баранова и А. В. Морозовой. М.: УРСС, 2004.
3. Searle J. R. Minds, brains, and programs // Behavioral and brain sciences. 1980. Vol. 3, iss. 3. P. 417–424. DOI: <https://doi.org/10.1017/S0140525X00005756>.
4. Chalmers D. J. The singularity: A philosophical analysis // Science fiction and philosophy: from time travel to superintelligence / S. Schneider (ed.). Chichester: Wiley-Blackwell. 2016. P. 171–224.
5. Ichikawa J. J., Steup M. The Analysis of Knowledge // The Stanford Encyclopedia of Philosophy / ed. E. N. Zalta, 2014. URL: <https://plato.stanford.edu/archives/spr2014/entries/knowledge-analysis/> (дата обращения: 20.12.2021).
6. Pavese C. Knowledge How // The Stanford Encyclopedia of Philosophy / ed. E. N. Zalta, 2021. URL: <https://plato.stanford.edu/entries/knowledge-how/> (дата обращения: 20.12.2021).
7. Williamson T. Knowledge and its Limits. Oxford: Oxford Univ. Press, 2002. DOI: 10.1093/019925656X.001.0001.
8. Shope R. K. The analysis of Knowledge. Princeton: Princeton Univ. Press, 1983.
9. Zagzebski L. On epistemology. Belmont, CA: Wadsworth Cengage Learning, 2009.
10. Martin A., Santos L. R. What cognitive representations support primate theory of mind? // Trends in cognitive sciences. 2016. Vol. 20, iss. 5. P. 375–382. DOI: 10.1016/j.tics.2016.03.005.
11. Davidson D. Rational animals // Dialectica. 1982. Vol. 36, no. 4. P. 317–327.
12. Dennett D.C. Real patterns // The J. of Philosophy. 1991. Vol. 88, no. 1. P. 27–51. DOI: <https://doi.org/10.2307/2027085>.
13. Фролов К. Г. Метафизика соответствия: некоторые подходы к корреспондентной теории истины // Эпистемология и философия науки. 2018. Т. 55, № 1. С. 83–98. DOI: 10.5840/eps 201855110.
14. Лобанов С. Д. О проблеме надежности знания и дилемме истина/знание // Вестн. Вятского гос. ун-та. 2011. № 4 (4). С. 26–27.
15. McGlynn A. Knowledge First? London: Palgrave Macmillan, 2014. DOI: <https://doi.org/10.1057/9781137026460>.
16. Feldman R. Reliability and justification // The Monist. 1985. Vol. 68, iss. 2. P. 159–174. DOI: 10.5840/monist198568226.
17. Pritchard D. Knowledge and understanding // Virtue epistemology naturalized / A. Fairweather (ed.). Cham: Springer, 2014. Vol 366. P. 315–327. DOI: https://doi.org/10.1007/978-3-319-04672-3_18.

Информация об авторах.

Дёмин Тимофей Сергеевич – ассистент кафедры философии Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В. И. Ульянова (Ленина), ул. Профессора Попова, д. 5Ф, Санкт-Петербург, 197022, Россия. Автор 11 научных публикаций. Сфера научных интересов: аналитическая эпистемология.

Фролов Константин Геннадьевич – кандидат философских наук (2017), научный сотрудник Международной лаборатории логики, лингвистики и формальной философии НИУ «Высшая школа экономики», Покровский б-р, д. 11, Москва, 109028, Россия; научный сотрудник кафедры философии Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В. И. Ульянова (Ленина), ул. Профессора Попова, 5Ф, Санкт-Петербург, 197022, Россия. Автор 24 научных публикаций. Сфера научных интересов: аналитическая метафизика.

О конфликте интересов, связанном с данной публикацией, не сообщалось.
Поступила 29.12.2021; принята после рецензирования 25.01.2022; опубликована онлайн 22.04.2022.

REFERENCES

1. Lakoff, G. (2004), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, Transl. by Shatunovskii, I.B., Yazyki slavyanskoi kul'tury, Moscow, RUSS.
2. Lakoff G. and Johnson, M. (2004), *Metaphors We Live By*, Transl. by Baranov, A.N. and Morozova, A.V., URSS, Moscow, RUSS.
3. Searle, J.R. (1980), "Minds, brains, and programs", *Behavioral and brain sciences*, vol. 3, iss. 3, pp. 417–424. DOI: <https://doi.org/10.1017/S0140525X00005756>.
4. Chalmers, D.J. (2016), "The singularity: A philosophical analysis", *Science fiction and philosophy: from time travel to superintelligence*, in Schneider, S. (ed.), Wiley-Blackwell, Chichester, UK, P. 171–224.
5. Ichikawa, J.J. and Steup, M. (2014), "The Analysis of Knowledge", *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.), available at: <https://plato.stanford.edu/archives/spr2014/entries/knowledge-analysis/> (accessed 20.12.2021).
6. Pavese, C. (2021), "Knowledge How", *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.), available at: <https://plato.stanford.edu/entries/knowledge-how/> (accessed 20.12.2021).
7. Williamson, T. (2002), *Knowledge and its Limits*, Oxford Univ. Press, Oxford, UK. DOI:10.1093/019925656X.001.0001.
8. Shope, R.K. (1983), *The analysis of Knowledge*, Princeton Univ. Press, Princeton, NJ, USA.
9. Zagzebski, L. (2009), *On epistemology*, Wadsworth Cengage Learning, Belmont, CA, USA.
10. Martin, A. and Santos, L.R. (2016), "What cognitive representations support primate theory of mind?", *Trends in cognitive sciences*, vol. 20, iss. 5, pp. 375–382. DOI: 10.1016/j.tics.2016.03.005.
11. Davidson, D. (1982), "Rational animals", *Dialectica*, vol. 36, no. 4. pp. 317–327.
12. Dennett, D. (1991), "Real patterns", *The J. of Philosophy*, vol. 88, no. 1, pp. 27–51. DOI: <https://doi.org/10.2307/2027085>.
13. Frolov, K.G. (2018), "Metaphysics of correspondence: some approaches to the classical theory of truth", *Epistemology & Philosophy of Science*, vol. 55, no. 1, pp. 83–98. DOI: 10.5840/eps201855110.
14. Lobanov, S.D. (2011), "About of problem knowledge reliability and question of truth/knowledge dilemma", *Herald of Vyatka State Humanitarian Univ.*, no. 4 (4), pp. 26–27.
15. McGlynn, A. (2014), *Knowledge First?*, Palgrave Macmillan, London, UK. DOI: <https://doi.org/10.1057/9781137026460>.
16. Feldman, R. (1985), "Reliability and justification", *The Monist*, vol. 68, iss. 2, pp. 159–174. DOI: 10.5840/monist198568226.
17. Pritchard, D. (2014), "Knowledge and understanding", *Virtue epistemology naturalized*, in Fairweather, A. (ed.), vol. 366, Springer, Cham, CHE, pp. 315–327. DOI: https://doi.org/10.1007/978-3-319-04672-3_18.

Information about the authors.

Timofei S. Demin – Assistant Lecturer at the Department of Philosophy, Saint Petersburg Electrotechnical University, 5F Professor Popov str., St Petersburg 197022, Russia. The author of 11 scientific publications. Area of expertise: analytical epistemology.

Konstantin G. Frolov – Can. Sci. (Philosophy) (2017), Research Officer at the International Laboratory for Logic, Linguistics and Formal Philosophy, Higher School of Economics, 11 Pokrovsky blvd., Moscow 109028, Russia; Research Officer at the Department of Philosophy, Saint Petersburg Electrotechnical University, 5F Professor Popov str., St Petersburg 197022, Russia. The author of 24 scientific publications. Area of expertise: analytical metaphysics.

*No conflicts of interest related to this publication were reported.
Received 29.12.2021; adopted after review 25.01.2022; published online 22.04.2022.*