

Интеррогативная эпистемология Я. Хинтикки и перспективы ее применения в исследовании искусственного интеллекта

А.Ю. Моисеева

*Институт философии и права Сибирского отделения РАН,
Новосибирск, Россия*

Аннотация

В статье изложены перспективы применения интеррогативной эпистемологии Я. Хинтикки для моделирования когнитивных операций, осуществляемых познающим агентом, с целью создания машины, способной к полноценному познанию. Установлено, что задача моделирования распадается на две задачи: моделирование самих познавательных операций и моделирование стратегических рассуждений. Первую задачу интеррогативная эпистемология успешно решает, поскольку она опирается на теоретико-игровой формальный аппарат, позволяющий корректно описывать все виды возможных ходов в рамках отдельной познавательной игры. Вторая задача не может быть решена в общей эпистемологической теории, т.к. стратегические рассуждения об организации исследования и даже понятие знания оказываются неизбежно зависимыми от более широкого контекста практической деятельности агента. Поэтому интеррогативная эпистемология отказывается от претензий на универсальное и полное описание процесса познания, концентрируясь на иных задачах, в первую очередь на выявлении скрытых посылок, которые используются познающими агентами в своих рассуждениях. По этой же причине возможность существования познающей машины с точки зрения интеррогативной эпистемологии не может быть формально обоснована и представляется сомнительной. Однако приведены аргументы в пользу того, что интеррогативная эпистемология служит приемлемым концептуальным каркасом для прояснения характера трудностей, возникающих при попытке решения этой второй задачи, и в целом для улучшения качества философского дискурса в области исследований искусственного интеллекта. Автор статьи аргументирует данную позицию тем, что способность к познанию может предполагать обладание всеми основными функциями, которые обычно называются психическими.

Ключевые слова: модели познания, теория игр, вопросы и ответы, стратегия, формализация, алгоритмизация, философия искусственного интеллекта.

Моисеева Анна Юрьевна – кандидат философских наук, научный сотрудник Института философии и права Сибирского отделения РАН.

ajumo@yandex.ru

<https://orcid.org/0000-0003-1440-9598>

Для цитирования: Моисеева А.Ю. Интеррогативная эпистемология Я. Хинтикки и перспективы ее применения в исследовании искусственного интеллекта // Философские науки. 2021. Т. 64. № 7. С. 46–67. DOI: 10.30727/0235-1188-2021-64-7-46-67

J. Hintikka's Interrogative Model of Inquiry and Prospects for Its Application in the Study of Artificial Intelligence

A. Yu. Moiseeva

Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science, Novosibirsk, Russia

Abstract

The article outlines the prospects of using J. Hintikka's interrogative epistemology for modelling cognitive operations carried out by a cognizing agent to create a machine capable of full cognition. It was established that modeling is divided into two objectives: modeling the cognitive operations and modeling the strategic reasoning. Interrogative epistemology presents a solution to the first objective. It relies on a game-theoretic formal apparatus that allows one to correctly describe all types of possible moves within the framework of a particular cognitive game. The second objective cannot be accomplished in a general epistemological theory framework since strategic considerations about the organization of research and even the concept of knowledge inevitably depend on the broader context of an agent's practical activity. Therefore, interrogative epistemology does not claim a universal and complete description of the process of cognition, focusing on other tasks: primarily on identifying hidden premises that cognitive agents use in their reasoning. For the same reason, interrogative epistemology deems the possibility of the existence of a cognizing machine as formally unsubstantiated and questionable. However, arguments are given in favor of the fact that interrogative epistemology serves as an acceptable conceptual framework for clarifying the nature of the difficulties that arise with attempts to accomplish the second objective and improving the quality of philosophical discourse in the field of artificial intelligence research. The author substantiates this position by arguing that the ability to cognize presupposes the ability to perform all the essential functions that are usually called mental.

Keywords: cognitive models, game theory, questions and answers, strategy, formalization, algorithmization, philosophy of artificial intelligence.

Anna Yu. Moiseeva – Ph.D. in Philosophy, Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Science
ajumo@yandex.ru
<https://orcid.org/0000-0003-1440-9598>

For citation: Moiseeva A.Yu. (2021) J. Hintikka's Interrogative Model of Inquiry and Prospects for Its Application in the Study of Artificial Intelligence. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 64, no. 7, pp. 46–67. DOI: 10.30727/0235-1188-2021-64-7-46-67

Введение

Интеррогативная эпистемология или, как ее еще называют, интеррогативная модель поиска знания [Knuuttila 2018] – это теория познания, которую разработал Я. Хинтикка в книгах «Исследование как исследование: логика научного открытия» (*Inquiry as Inquiry: A Logic of Scientific Discovery*) [Hintikka 1999], «Сократическая эпистемология: исследование поиска знания посредством вопросов» (*Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning*) [Hintikka 2007], а также в ряде статей, не вошедших в книги. На мой взгляд, данная теория представляет собой удобный и многообещающий концептуальный каркас для разработки вопросов, связанных с тем, что мы имеем в виду, когда говорим о возможности мыслящей, точнее, познающей машины¹. Попытаюсь аргументировать свою точку зрения.

Проблемы эпистемологии Я. Хинтикка рассматривает еще в работах 1960-х годов, прежде всего в книге «Знание и вера: введение в логику двух понятий» (*Knowledge and Belief: An Introduction to the Logic of the Two Notions*) [Hintikka 1962], «самой цитируемой книге в современной эпистемической логике и классике философии прошедшего века» [Knuuttila 2018, 414]. Он развивает теоретико-

¹ Термин «познающая машина» не используется у Я. Хинтикки, он говорит только о мыслящей машине, но именно в контексте способности к познанию. Я буду использовать данный термин нестрого, исключительно в целях пояснения позиции Я. Хинтикки относительно темы искусственного интеллекта, а не для создания на его базе теоретических построений. Поэтому я не даю ему определения, ограничиваясь тем, что термин «говорит сам за себя». Вместе с тем далее в статье имеются некоторые рассуждения о том, что именно он может означать в контексте исследований искусственного интеллекта и какие следствия он может повлечь в данном контексте.

модельный подход к формализации содержания эпистемических установок и показывает, как вывести логические свойства знания и веры из семантики утверждений о знании и вере соответственно. Модальности знания и веры формализуются у него в виде сентенциальных операторов *K* и *B*, которые индексируются по эпистемическим агентам. Эпистемическая логика описана как аналог (*counterpart*) логики для алетических модальностей; в этом смысле утверждение «*a* знает, что *p*» является эпистемическим аналогом утверждения «необходимо, что *p*», если не учитывать множественность агентов.

В работах указанного периода Я. Хинтикка исследует исключительно статические свойства систем знаний и не касается вопросов, связанных с условиями и способами модификации этих систем. Логика, которую он развивает, – эпистемическая логика «первого поколения», как позднее он стал ее называть, – предназначена не для формализации собственно процесса познания, а только для описания результата этого процесса, т.е. уже имеющегося знания. Более того, понятие знания, которое формализуется с помощью эпистемической логики «первого поколения», настолько идеализировано, что обладать на практике эпистемической установкой с такими свойствами оказывается невозможно ни для одного реально существующего агента или, во всяком случае, ни для одного человека. В частности, Я. Хинтикка признает логическое всеведение агента и его способность всегда знать о знаниях, которые у него имеются (т.н. КК-тезис). Хотя причины, по которым принимаются эти допущения, носят сугубо логический характер, сами допущения имеют очень сильные психологические следствия: принятие логического всеведения находится в противоречии с предположением о конечности вычислительных ресурсов, а КК-тезис, понятый в психологическом смысле, видимо, означает, что агент пребывает одновременно в бесконечном количестве эпистемических состояний, что само по себе нуждается в объяснении. Иными словами, идеализация Я. Хинтикки в некоторых аспектах контринтуитивна. Но это не мешает ей быть полезной идеализацией, поскольку понятие знания Я. Хинтикки, как он многократно подчеркивает, – это не то понятие знания, которым мы оперируем в обыденной жизни, а то, которое лежит в основе принимаемой нами логики эпистемических высказываний. Мы не можем обладать таким знанием, но можем к нему стремиться, и некоторые свойства обыденного понятия знания

показывают, что нам действительно свойственно стремиться к нему².

Все эти особенности вписывают понятие знания, используемое Я. Хинтиккой в работах указанных лет, скорее в контекст классической эпистемологии, чем в контекст современных тенденций в теории познания, где все большую популярность завоевывают различные натурализирующие и прагматизирующие подходы. Однако в дальнейшем автор в значительной степени отходит от классических идеалов. Так, он предпринимает попытки избежать логического всеведения, но не дает достаточно разработанного формального аппарата для работы с этой новой семантикой знания [Хинтикка 1980]. Размышляет он и о сфере действия КК-тезиса, анализируя вопросы, связанные с ним. Единственное, чего он долгое время не касается, — это статический характер его семантики, не позволяющий формализовать когнитивные операции, которые мы совершаем, когда ищем, находим и принимаем новое знание в качестве такового. Между тем и методологический, и эпистемологический интерес Я. Хинтикки, очевидно, распространяется на такие операции, поскольку именно в них реализуется наше стремление к идеалу, который задает эпистемическая логика «первого поколения». Решительным прорывом в данном направлении становится радикально новая концепция знания, которая формулируется Я. Хинтиккой в рамках интеррогативной эпистемологии. Эта концепция будет занимать существенное место в аргументации настоящей статьи. Но прежде чем адекватно изложить ее, нам потребуется сначала дать общую характеристику интеррогативной эпистемологии как подхода и обозначить идеи, которые Я. Хинтикка заложил в его основу.

Концептуальные основы интеррогативной эпистемологии

Главная отличительная черта интеррогативной эпистемологии, определившая ее название, состоит в том, что познание представляется при этом подходе в виде процесса генерации вопросов и получения ответа на них от «оракула», под которым можно понимать природу, человека или, например, электронный носитель

² Утверждается, что этот подход отражает методологические взгляды Я. Хинтикки, которые он называет семантическим нео-кантианством и согласно которым «истины логики не являются законами мышления. Они не приходят к нам автоматически, но требуют от нас работы, будучи конструируемыми в смысле трансцендентального анализа» [Knuuttila 2018, 417].

информации³. В общем виде цель такого процесса состоит в том, чтобы получить объяснение конкретного положения дел, обладающее заданными свойствами. Исследование достигает цели, когда агент получает ответ на вопрос, ради которого оно начато; но в начале исследования этот «самый большой» вопрос может быть неизвестен, и даже если он известен, получить на него ответ сразу невозможно – иначе все исследование выродилось бы до единственного шага. Результат получается при помощи последовательности, состоящей из «малых» вопросов и дедуктивных шагов, выводящих следствия из множества посылок, которые либо приняты изначально, либо получены в качестве ответов на вопросы предыдущих шагов. Например, в детективном расследовании обычно требуется установить личность преступника и то, какое именно преступление он совершил (если заранее известно, что совершено преступление). Но, чтобы установить это, как правило, нужно узнать, каким способом преступление совершено, зачем оно совершено и т.д. В науке исследователь отправляется от стремления подвести некоторое частное явление под общий естественный закон. Однако он не может этого сделать, пока не квалифицирует частное явление как явление определенного *вида*, не сформулирует условия, при которых наблюдаются явления данного вида (причем иногда требуется формулировка отдельно необходимых и достаточных условий), и факторы, принимающие участие в обеспечении этих условий. В психологическом объяснении человеческих действий требуется знание мотивов, но это знание во многих случаях невозможно получить, если не знать той части убеждений человека, которая имеет значение для формирования соответствующего мотива, не знать особенностей его когнитивных и эмоциональных процессов и т.д.

На каждой стадии агенту приходится решать, какие именно посылки следует в данный момент использовать в качестве основания для дедукции или вопроса. Задача, которая стоит здесь перед агентом, является задачей *стратегической*, если понимать под стратегией способ применения в процессе исследования формальных правил, управляющих собственно дедукцией и интеррогацией. Успех стратегии зависит от того, способен ли агент предугадать, куда может привести рассуждение, если руководствоваться ею, а такая способность, в свою очередь, зависит от исследовательского

³ В ряде случаев, о чем пойдет речь далее, в качестве оракула выступает сам агент, т.е. часть его психики, обладающая информацией, недоступной прямому осознанию и вербализации.

опыта агента, от того, насколько успешно он применял те или иные стратегии в прошлом. Поэтому выбор стратегии исследования, как правило, не полностью обусловлен информацией, эксплицитно представленной в посылках, а зависит и от фоновой информации, которую (зачастую не сознательно) агент может привлечь для обоснования своего решения, в том числе от информации, имплицитно содержащейся в его жизненной практике.

Как полагает Я. Хинтиikka, основной недостаток «чисто» выводной (дедуктивной или индуктивной) модели поиска знаний состоит в том, что вся информация, которая может быть использована в процессе вывода, должна быть представлена с самого начала в виде явно сформулированных предложений языка [Хинтиikka, Хинтиikka 1987, 272–273]. В интеррогативной же модели мы не обязаны формулировать все фоновое знание, на основе которого даются ответы на вопросы, что лучше соответствует реальному положению дел, поскольку основная задача исследователя зачастую состоит в том, чтобы понять, *какого вида* информация ему требуется. В сущности, вопрос – это и есть запрос на экспликацию фоновой информации, выбранной по определенному формальному критерию. Форма вопроса определяет форму возможного ответа на него, а значит, чтобы задать правильный (т.е. продвигающий исследование к цели) вопрос, агенту нужно предугадать, какая информация содержится в фоне и в какой форме она может быть эксплицирована адекватно цели исследования. Но агент не обязан заранее обладать *самой* этой информацией во всей ее полноте. Фактически агент обладает чаще только частью фоновой информации, позволяющей ему успешно предугадывать (т.е. исследовательским опытом). Остальная информация извлекается им в процессе взаимодействия с какой-то внешней инстанцией, которое может быть выражено в форме диалога, наблюдения или (в естественных науках) контролируемого эксперимента.

Точнее, механизмом предугадывания служит моделирование на каждом этапе различных ситуаций или, как называет их Я. Хинтиikka, сценариев. Каждый из сценариев соответствует определенному ходу развития исследования, причем в каждом из них агенту нужно представить и фактическое состояние, в котором при данном сценарии находится познаваемая им реальность, и эпистемическое состояние, в котором находится он сам. Иными словами, он должен иметь представление о собственных потенциальных знаниях, которые могут стать реальными знаниями, если он задаст определенный вопрос и получит на него

определенный ответ. При таком условии он сможет предугадать, какие дальнейшие вопросы и дедукции станут доступны ему в том эпистемическом состоянии. В итоге все это необходимо агенту, чтобы понять, какой стратегии ему следует придерживаться, чтобы «попасть» в сценарий, где он получает в конце желаемое им знание. Таким образом, формальные правила, управляющие дедукцией и интеррогацией, должны быть дополнены формальными или неформальными правилами, управляющими выбором стратегии. Как подчеркивает Я. Хинтиikka, «никакая эпистемологическая теория не может рассказать историю, которая была бы полной и имела бы дело лишь с правилами для отдельных шагов или с эпистемической оценкой отдельной когнитивной ситуации. Такая теория может дать нам правду и ничего кроме правды, но это не будет вся правда» [Hintikka 2007, 8].

В качестве примера того, какую роль играют стратегические соображения в интеррогативной модели поиска знания, Я. Хинтиikka предлагает рассмотреть метод, с помощью которого раскрывал преступления великий (хотя и вымышленный) сыщик Шерлок Холмс. Например, в расследовании Холмсом похищения знаменитого рысака Серебряного ключевым шагом становится вопрос о «странном поведении собаки в ночь преступления», точнее, два вопроса: «Лаяла ли собака тренера, сторожившая конюшню, когда похититель выводил лошадь?» (ответ, как известно, отрицательный) и «Кто тот человек, на которого обученная сторожевая собака может не залаять?» [Хинтиikka, Хинтиikka 1987, 271]. Ответ на второй вопрос приводит детектива к осознанию роли тренера в преступлении. При этом вся информация, необходимая для ответа на него после того, как получен ответ на первый вопрос, содержится в фоновых знаниях каждого образованного человека. Тем не менее, очевидно, многие читатели не могут самостоятельно разобраться в обстоятельствах дела. Суть метода Холмса, дающего ему преимущество, заключается не в специфическом знании, а в том, как он реконструирует в своем воображении ситуацию преступления. Он спрашивает себя о том, чего недостает в известных ему данных, чтобы «странное» поведение собаки стало выглядеть вполне естественным, и приходит к окончательному выводу: недостающий элемент – некоторая характеристика личности похитителя. Поэтому он задает именно такой вопрос.

О формализации

Для формализации собственной модели поиска знания Я. Хинтиikka использует аппарат теории игр. В виде игры он представляет

отдельное исследование, в виде игроков – агента и «оракула», а ходами в игре являются дедуктивные и интеррогативные шаги, предпринимаемые агентом в процессе исследования («оракул» пассивен – он может только отвечать на вопросы). Последовательность ходов и их результаты фиксируются в специальных таблицах, напоминающих таблицы Бета. Для каждого из типов ходов существуют правила о том, при каких условиях данный ход возможен. Так, дедуктивные ходы описываются вариантом обычных правил построения табличного доказательства, записанных в стиле Генцена, где антецедент соответствует правому столбцу таблицы, а консеквент – ее левому столбцу [Hintikka 1999, 48]. Например, аналог правила удаления дизъюнкции представлентак:

$$(L.\vee) \frac{\Gamma, (S_i \vee S_j) \rightarrow \Delta}{\Gamma, S_i \rightarrow \Delta \quad \Gamma, S_j \rightarrow \Delta} . \quad (1)$$

Данная запись означает, что всегда при появлении дизъюнкции в левом столбце таблицы такая таблица должна быть разделена на две подтаблицы, в одной из которых в левом столбце необходимо указать первый дизъюнкт, а в другой – второй.

Правила, регулирующие удаление кванторов, также представляют собой аналоги обычных правил табличного вывода. В частности, удаление экзистенциального квантора предполагает возможность расширения языка посредством введения пустого символа (функциональной константы), означающего неопределенный индивид, который удовлетворяет заданным условиям.

Правила для вопросов определяют понятие пресуппозиции вопроса и устанавливают, что вопрос может быть задан, т.е. записан в правый столбец таблицы, лишь в том случае, если его пресуппозиция уже дана, т.е. присутствует в левом столбце. Например, пресуппозиции для закрытых вопросов имеют форму дизъюнкции:

$$S_1 \vee S_2 \vee \dots S_n . \quad (2)$$

Ответом на них, соответственно, может быть только один из дизъюнктов S_i . Пресуппозицией открытого вопроса, где агент находится в поисках конкретного индивида, для которого будет выполняться определенное условие, может быть либо дизъюнкция, либо предложение, включающее в себя экзистенциальный квантор, вида

$$S_0 [(\exists x) S_1 [x]]^4 . \quad (3)$$

⁴ Сохранена оригинальная нотация из [Hintikka 1999, 51].

Существуют специальные правила закрытия «веток» (под-таблиц) таблицы. Согласно правилам, если в таблице закрыты все «ветки», это означает, что игра окончена, т.е. искомое знание получено агентом.

Каждому ходу агента, модифицирующему таблицу, сопоставляется определенная «стоимость». Так, «стоимость» вопроса зависит от сложности ответа на него. Сложность ответа на любой закрытый вопрос в рамках данной теории получает количественную оценку, связанную с квантификационной сложностью пропозиции, истинность которой устанавливается посредством этого вопроса. Сложность ответа на открытые вопросы тоже можно оценить путем переформулировки их в некоторое множество закрытых вопросов (то, что переформулировка всегда возможна, доказано в [Hintikka 2007, 55]). «Стоимость» дедуктивного хода равна количеству пустых символов, которые он вводит. Если допускаются дефиниционные ходы (об этой возможности говорится в [Хинтикка, Хинтикка 1987, 279]), то их «стоимость» соответствует квантификационной сложности дефиниенса. В целом, как пишет Я. Хинтикка, интуитивный принцип, лежащий в основе оценки различных ходов, сводится к тому, что «ход тем дороже, чем больше он осложняет конфигурации индивидов, которые рассматриваются в дедукции (в смысле введения новых индивидов в доказательство). Эта “стоимость” добавления новых индивидов к нашему дедуктивному доказательству (независимо от того, являются ли они подлинными индивидами или “индивидами произвольными”, которые репрезентируются пустыми именами) отражает важность таких ходов для успешной дедукции. Выбор нового индивида для введения – это важнейшее стратегическое соображение в наших играх» [Хинтикка, Хинтикка 1987, 279]. Как он утверждает, нечто похожее несложно обнаружить в настоящих доказательствах типа приведенных Холмсом. Рассмотрим эпизод с поведением собаки в ночь, когда произошло преступление. Холмс впервые связывает друг с другом трех индивидов: неизвестного вора, собаку и тренера (согласно Я. Хинтикке, фигура собаки вводится в доказательство, и с ее помощью устанавливается связь между двумя другими индивидами). Неожданность «дедукции» Холмса состоит не в установлении связи между тремя индивидами (двое из них оказываются тождественными), а в том, что «впервые задается вопрос о связывающих их отношениях. А это стало возможным благодаря введению третьего индивида» [Хинтикка, Хинтикка 1987, 279].

Формализуема ли стратегия?

С учетом изложенного можно допустить, что выбор стратегии представляется вполне рациональным и, более того, формализуемым процессом. Суммируя «стоимость» ходов, легко рассчитать общую «стоимость» для различных стратегий при различных возможных ответах «оракула». Общая стоимость далее может быть сопоставлена с долей закрытых «веток» в таблице при разных вариантах развития исследования. Тем самым формируется сводная оценка ожидаемой результативности стратегий. Остается лишь выбрать стратегию, у которой оценка будет максимальной. Поскольку собственно познавательные ходы, дедуктивные и интеррогативные, полностью заданы правилами игры, формально описанными выше, формализуемость процесса выбора стратегии означает, что можно описать машину, осуществляющую все операции, которые конституируют познавательную деятельность в самом «человеческом» смысле этого слова, т.е. абстрактную (логическую) возможность существования познающей машины можно строго доказать. Если бы это было так, получение соответствующего доказательства стало бы большим успехом, далеко продвигающим вперед дискуссии об искусственном интеллекте⁵.

Вместе с тем на основании ряда высказываний Я. Хинтикки создается впечатление, что он полагает задачу выбора стратегии не разрешимой полностью формальным способом. В частности, в одной из его статей последних лет к общему замечанию о том, что «в любой игре (в теоретическом смысле этого слова) можно различать определяющие правила, специфицирующие, какие

⁵ Конечно, логическая возможность существования такой машины еще не означает, что возможна ее реализация в физическом устройстве. Однако обоснование последней – это, скорее, дело технических наук, а не философии. Вместе с тем существует отдельный философский вопрос о соотношении терминов «познающая машина» и «искусственный интеллект». В настоящей статье познание представлено в качестве одной из функций, присущих человеку как сознательному существу, которые, следовательно, должны быть присущи тому, что мы согласны назвать искусственным интеллектом в т.н. сильном смысле этого слова. Но действительно ли познание является необходимой функцией сознательного существа? Если познавать означает целенаправленно получать знание, то нельзя ли представить себе такое существо, которое было бы сознательным, т.е. имело бы ментальные состояния, не стремясь при этом к знанию о чем бы то ни было? В настоящее время получение однозначного ответа на этот вопрос не представляется возможным. Хотя, думается, на него могут пролить свет соображения Я. Хинтикки, изложенные в статье далее.

“ходы” можно делать в данной игре, и стратегические правила, или принципы, облегчающие достижение целей “игроков”), сделано специальное уточнение относительно того, что «правила, управляющие мысленным экспериментом логика, неизбежно стратегические, и не могут быть сведены к определяющим правилам никакой другой игры» [Хинтикка 2011]. Далее идет фрагмент, относящийся непосредственно к теме настоящего исследования: «Философы и нефилософы задавались вопросом, является ли человеческий разум (или, возможно, скорее, может ли он быть смоделирован) цифровым компьютером. В популярной форме этот вопрос звучит как “может ли компьютер мыслить?” Если мыслить значит следовать определяющим правилам, ответ тривиально утвердительный. Но совершенно другой вопрос, в какой мере можно считать, что компьютеры способны овладеть стратегическими правилами, например, способностью самим формировать стратегии и изменять их в свете опыта» [Хинтикка 2011]. В приведенном фрагменте не говорится о том, что ответ на этот вопрос обязательно отрицательный. Однако в контексте предыдущего предложения можно понять сказанное так, что несводимость стратегических правил к определяющим является вероятным следствием их формальной невыразимости вообще.

Наряду с этим следует отметить, что Я. Хинтикка ранее, насколько известно, не был замечен в иррационализме или ментализме. Наоборот. Его вклад в дело формализации содержания пропозициональных установок трудно переоценить. Было бы странно, если б он стал безосновательно настаивать на том, что какие-то когнитивные операции, доступные человеку, принципиально неформализуемы, а значит, не могут быть воспроизведены компьютером. И было бы вдвойне странно, если б, имея основания для этого, он не привел их в явном виде. Более правдоподобным видится, что он просто не знает, каким образом можно было бы учесть в компьютерной программе все релевантные факторы (не будем забывать о роли неявных посылок в его модели познания), но если бы они были учтены, то компьютер мог бы осуществлять познавательную деятельность с тем же успехом, что и человек. Известно, по крайней мере, что в вопросах о научном познании Я. Хинтикка придерживается рационалистических взглядов. По его утверждению, присутствие в рассуждениях ученого неявных посылок не обязывает нас к тому, чтобы считать выбор стратегии научного исследования иррациональным [Hintikka 2007, 86]. Таким образом, если предположить, что полностью рациональная методология исследования существует, вопрос будет сводиться

к тому, возможно ли эту рациональную методологию выразить языком компьютерной программы.

В другом фрагменте, вновь в связи с различием между определяющими и стратегическими правилами, речь идет об энтимемах. Я. Хинтиikka пишет о том, что, поскольку в интеррогативной модели познания вывод частично опирается на посылки, взятые из фонового знания, рассуждение исследователя является энтимемическим. Он высказывает мысль о современных логиках, занимающихся амплиативными (т.е. расширяющими знание) рассуждениями, точнее, о том, что они пытаются формализовать процесс обоснования выбора посылок для принятия или, наоборот, отбрасывания. Он утверждает, что это – ошибка, т.к. выбор посылок – вопрос стратегический, и ответ на него не может быть выражен в каких бы то ни было определяющих правилах. По мнению Я. Хинтиikka, «амплиативные логики могут представлять значительный практический интерес и ценность, но с окончательной эпистемологической точки зрения они представляют собой лишь типы энтимемических рассуждений, полагающихся на скрытые предпосылки в такой же степени, как и на ограничительные (*circumscriptive*) рассуждения. Основная задача эпистемолога здесь состоит не в том, чтобы изучать технические детали таких способов рассуждения, хотя они и увлекательны сами по себе. Она состоит в том, чтобы раскрыть неявные предпосылки, на которых на самом деле основано такое энтимемическое рассуждение» [Hintikka 2007, 21]. Сказанное можно понять так, что, по мысли философа, исследование логических аспектов энтимемических рассуждений, предпринимаемое создателями современных амплиативных логик, характеризуется как малополезное для эпистемологии начинание, поскольку в конечном счете такая «логика» должна формализовать, помимо всего прочего, стратегические соображения, а сделать это она не способна. Следовательно, для интеррогативной эпистемологии сознательно избирается иная цель – не создание эксплицитной и полностью проработанной логики поиска знаний, а выявление неявных посылок, задействованных в поиске. Как пишет Я. Хинтиikka, с изобретением способа формализации всех видов вопросов формальный аппарат интеррогативной эпистемологии получил свою завершенность. В нем появилась возможность выразить все определяющие правила интеррогативной игры, т.е. единственное, что нужно. Никакая амплиативная логика не требуется. Стратегические соображения являются «внешними» по отношению к игре, и они могут использоваться неформально.

По поводу последнего приведенного фрагмента также можно спорить, насколько он репрезентирует то, что Я. Хинтикка думает об алгоритмизации стратегического выбора вообще. Возможно, фрагмент относится исключительно к конкретным логикам, с которыми был знаком автор в период написания книги. Возможно, он имел в виду не то, что никакая амплиативная логика не согласуется с теорией познания, которую он отстаивает, а лишь то, что «технические детали таких способов рассуждения» не обогатят эту теорию ничем существенным. Представляется нецелесообразным ставить перед собой вопрос о том, какова на самом деле его точка зрения относительно рассматриваемой нами проблемы. Разрешение данного вопроса по праву относится к компетенции историков философии. Для нас более важно понять суть его утверждения о том, что стратегические правила «не могут быть сведены к определяющим правилам никакой другой игры», и установить, какие следствия оно имеет для темы настоящего исследования. Но поскольку «сходу» разобраться в этом не удастся, в качестве промежуточной стадии я предлагаю использовать анализ Я. Хинтиккой природы знания и места, которое знание занимает в интеррогативном исследовании.

Знание как цель

В своих работах Я. Хинтикка много внимания уделяет вопросу о том, что такое знание и чем оно не является. Согласно идеям философа, знание, наряду с верой (*belief*), представляет собой в первую очередь основание (*entitlement*) для принятия решения, как действовать или рассуждать в некоторой ситуации. Постоянно повторяется следующий пример Я. Хинтикки: если я знаю, что сегодня не будет дождя, я имею основания для того, чтобы, уходя из дома, оставить там свой зонт [Hintikka 2007, 11–12]. Технически это описано с помощью сечения множества возможных сценариев, включающих в себя данную ситуацию, на два подмножества: те, которые согласуются со знанием агента, и те, которые не согласуются с ним. Первые агент учитывает, принимая решение, вторые – игнорирует. В отношении приведенного примера можно утверждать следующее: мое знание о том, что сегодня не будет дождя, означает, что я игнорирую все сценарии, в которых истинно, что сегодня идет дождь, и принимаю во внимание только те, в которых дождя сегодня нет. Это дает мне основания заключить, что зонт мне не понадобится, и оставить его дома.

Однако, как пишет Я. Хинтикка, основание в данном случае следует понимать как связь *de jure*, которая *de facto* может и не

повлиять на поведение [Hintikka 2007, 12]. Во-первых, основанием для принятия решения служит не знание какого-то отдельного факта, а знание как целое, т.е. все, что знает агент в ситуации, о которой идет речь. Следовательно, знание одного и того же факта в разных эпистемических контекстах может приводить к различным способам поведения, вплоть до диаметрально противоположных. Скажем, если зонт для меня является не только укрытием от дождя, но и средством самозащиты, я могу взять его с собой, даже если верю, что сегодня не будет дождя. Во-вторых, речь идет именно об обосновании в юридическом смысле, буквально о делании какого-то акта правомочным (*entitled*), а не о причине этого акта. В конечном счете выбор остается за агентом, т.к. именно агент принимает на себя все риски и получает все выгоды, которые связаны с реализацией избранного им способа поведения. Знание же выступает лишь в качестве руководства в том, как сделать наименее рискованный и наиболее выгодный выбор. Другими словами, связь между знанием и действием относится к сфере практической рациональности.

Познавательная деятельность – один из видов деятельности, поэтому сказанное выше распространяется, в частности, на нее. Действия, выполняемые агентом в процессе познания, обоснованы уже имеющимся у него знанием. Особенность же исследуемого вида деятельности, по сравнению с остальными видами, состоит в том, что знание, которым руководствуется агент в своем познании, в процессе самого познания прирастает. Поскольку основанием для принятия решения, *как* познавать, является целостное знание агента, прирастая, данное знание изменяет и обоснованность различных возможных решений. Более того, каждый новый факт, который узнает агент, изменяет вес всех остальных известных ему фактов в достижении главной познавательной цели, а значит, и в обосновании выбора стратегии дальнейшего исследования. На каждом этапе исследования агент должен в свете нового знания пересмотреть все основания своего предыдущего выбора и при необходимости скорректировать его. Некоторые скрытые послылки, которые ранее считались надежными, могут быть в процессе этого пересмотра отброшены⁶ (например, в истории с похищением рысак могла быть отброшена посылка о том, что обученная и здоровая сторожевая собака всегда лает на похитителя). Поэтому применительно к тому,

⁶ Я. Хинтика обобщает отбрасывание техническим термином «взятие в скобки» (*bracketing*).

с чем оперирует агент, пока он находится «внутри» интеррогативного исследования, точнее говорить не о знании в обычном – предполагающем непогрешимость – философском смысле этого слова, а об информации. Знание как таковое появляется только в качестве цели и последнего результата исследования, т.е. в точке, где оно должно завершиться.

Итак, интеррогативное исследование является в первую очередь процессом приобретения информации и лишь затем, опосредованно, процессом приобретения знания. Соответственно, и логика исследования является логикой информации, а не логикой знания. Согласно Я. Хинтикке, эпистемическая логика «первого поколения», которую он развивал в 1960-х годах, непригодна для описания того, что происходит с информацией в ходе исследования, поскольку она не дает возможностей для отбрасывания посылок и не способна адекватно отразить все варианты вывода, встречающиеся в реальном познавательном процессе⁷. Все это означает, как считает Я. Хинтикка, что знание не играет во «внутренней кухне» интеррогативного исследования существенной роли. Он пишет: «Понятие знания связано, таким образом, с интеррогативным исследованием посредством вопроса. Когда исследование достаточно далеко продвинулось, чтобы обосновать действия исследователя на базе тех заключений, которые оно на тот момент дало?» [Hintikka 2007, 27]. Но способность или неспособность достаточно обосновывать действия – критерий очевидно прагматический, «внешний», он относится к контексту исследования⁸, а не к исследованию как таковому. Любая попытка определить знание внеконтекстным способом, как заключает Я. Хинтикка, приведет к тому, что оно будет выражено в терминах определяющих правил исследования, т.е. как момент, когда исследование останавливается. Становится понятным, что подобное определение не может считаться адекватной философской концепцией знания.

По этой же причине проблематичным видится вопрос о возможности существования познающей машины. Ведь познавательную

⁷ Помимо этого, эпистемическая логика «первого поколения» имеет проблемы, связанные с квантификацией «внутри» эпистемических контекстов. Все эти недостатки, как пишет Я. Хинтикка, преодолеваются в эпистемической логике «второго поколения» или IF-логике, которую он и предлагает в качестве формального аппарата, дополняющего интеррогативную эпистемологию [Hintikka 2007, 18–19].

⁸ Тема исследования (*subject matter*) в данном случае включена в его контекст. Обратим внимание на обсуждение этого у Я. Хинтикки [Hintikka 2007, 29].

игру невозможно полностью определить, пока не задана ее цель — знание, а сама эта цель зависима от практического контекста исследования, т.е. от совокупности целей и средств в их взаимосвязи, обосновывающих всю деятельность агента. Если таким агентом должна быть машина, это означает, что у нее должны быть собственные представления о целях и средствах, соображения об обоснованности тех или иных действий, заинтересованность в возможных выигрышах и ответственность за риски. Иными словами, познающая машина должна иметь возможность выступать в качестве субъекта практической деятельности, обладающего, по крайней мере, иллюзией свободы воли⁹. При рассмотрении вопроса в широком смысле можно предположить, что «быть познающей» для машины равнозначно тому, чтобы обладать всеми основными функциями, которые мы обычно называем психиче-

⁹ Нецелесообразно здесь вдаваться в дискуссии о том, существует ли настоящая, неиллюзорная свобода воли. Даже если она существует, человеку не обязательно обладать ею, чтобы быть «познающим». Достаточно думать о себе как о ком-то, способном к свободным действиям, а не только к пассивному переживанию тех или иных состояний. Или даже сильнее: достаточно, чтобы другие думали о нем как о ком-то, кто сам думает о себе так. Позиция по этому вопросу, наиболее согласующаяся с совокупностью тезисов Я. Хинтикки, может быть сформулирована следующим образом. Любое суждение о сравнительных преимуществах и недостатках той или иной стратегии познавательной деятельности предполагает признание наличия у агента хотя бы принципиальной возможности свободного выбора между разными стратегиями. В этом смысле эпистемология всегда будет нормативной дисциплиной (хотя она, о чем пойдет речь в настоящей статье далее, и не способна выработать ни одного универсального стратегического правила). Тем не менее признание выбора свободным в контексте нормативно-эпистемологического рассуждения не обязывает нас к тому, чтобы отрицать детерминированность этого выбора в более широком контексте. А значит, возможна также наука о склонностях, обычаях и привычках тех, кто в более узком контексте рассматривается нами в качестве свободных познающих агентов. Данная наука тоже может быть названа эпистемологией, но с приставкой «натурализованная» во избежание путаницы. В этом смысле эпистемология как дисциплина содержит и дескриптивную часть. Собственно интеррогативная эпистемология, т.е. теория познания, которую развивает Я. Хинтикка, описательна в минимальной степени, в основном если нужно указать на причины успеха определенного агента в достижении знания в случаях, когда заранее известно о том, что он действительно преуспел (подтверждение тому содержит фрагмент настоящей статьи, где говорится о методе Холмса).

скими¹⁰. Таким образом, понятие познающей машины, изначально введенное сугубо для удобства применения теории Я. Хинтикки к нашей теме, оказывается в итоге одной из возможных конкретизаций понятия искусственного интеллекта вообще. Конечно, существует еще много неясностей, которые требуется устранить прежде, чем говорить о какой-то новой концепции искусственного интеллекта, предлагаемой интеррогативной эпистемологией. Тем не менее уже сейчас можно утверждать, что интеррогативная эпистемология способна добавить определенности в ряд вопросов, связанных с моделированием когнитивной деятельности, а значит, она служит перспективным концептуальным каркасом для обсуждения проблем разработки искусственного интеллекта.

Границы эпистемологии

Итак, Я. Хинтикка говорит нам о том, что неизбежно определение знания контекстно зависимо – по крайней мере, в том смысле, что оно зависит от области познания, к которой относится. Знание для детектива – не то же самое, что знание для ученого-естествоиспытателя. Отсюда можно заключить, что и правила успешного поиска знания будут для них разными: о стратегических принципах детективной работы следует говорить отдельно от стратегических принципов науки. Могут существовать принципы, общие для этих двух областей, например, принципы деятельности, направленной на открытие вообще. Но и такие принципы не будут универсальными: для познания посредством изобретения (в технических науках) они могут оказаться неэффективны, т.е., следуя им, мы негарантированно сможем получить искомое техническое знание. И насколько бы широкую область

¹⁰ Данное утверждение – очень грубое приближение к действительному значению термина «познающая машина», и оно не должно рассматриваться как претензия на обоснованный вывод. Чтобы обосновать такой вывод, потребовалось бы как минимум перечислить все функции, «обычно» относящиеся к психическим, указать, какие из этих функций являются «основными» и почему, а далее необходимым образом связать каждую из этих функций с концепцией познания, которая развивается в рамках интеррогативной эпистемологии. Настоящая статья этого не содержит и не предполагала изначально, и вряд ли это возможно на данном этапе. В ней представлен лишь набросок того, как мог бы выглядеть подход к проблеме искусственного интеллекта с точки зрения интеррогативной эпистемологии, если бы задача формулировать такой подход была всерьез воспринята исследователями, работающими в соответствующей области.

познания мы ни взяли, принципы, работающие для данной области, за ее пределами могут оказаться неэффективными.

Иными словами, универсально значимое стратегическое знание недостижимо. Представляется допустимым сформулировать стратегический принцип в виде определяющего правила некоторой метаигры, в которой точно определено, что считать знанием в игре нижнего уровня. Но нельзя однозначно утверждать, какую из возможных метаигр следует выбрать, чтобы получить то, что является знанием с точки зрения нашей более общей практики. Полагаться на какие-то принципы выбора возможно. Однако даже если сформулировать эти принципы в виде определяющих правил мета-метаигры, в которой точно определено, как выбирать метаигру, нельзя с уверенностью утверждать, какая из возможных мета-метаигр будет наиболее адекватной практическим потребностям. К такому заключению и приводит, по-видимому, фраза Я. Хинтикки о невозможности сведения стратегических правил к определяющим правилам не только этой, но и любой другой игры¹¹.

Однако, по дальнейшему замечанию Я. Хинтикки, понятие знания принадлежит прикладной эпистемологии, а не эпистемологии как общей дисциплине. Эпистемологию как общую дисциплину возможно выстроить без использования понятия знания, на одном понятии информации. Более того, утверждается, что в действительности именно закономерности поиска информации являются наиболее содержательными и наиболее интересными эпистемологическими закономерностями, поэтому эпистемология с одним базовым понятием информации практически ничего не потеряет по сравнению с традиционной эпистемологией. И такая эпистемология, по мнению Я. Хинтикки, не будет контекстно зависимой: «Общие эпистемологические вопросы касаются в первую очередь природы интеррогативного исследования, и они ни в коем случае не зависят от контекста или даже от предмета исследования» [Hintikka 2007, 30]. Поэтому использование интеррогативной модели в эпистемологии не предполагает какого бы то ни было релятивизма или скептицизма, как можно было

¹¹ Не могу утверждать, что мною точно отражен ход мысли Я. Хинтикки в данном случае. Однако он, по крайней мере, ссылается на позднего Л. Витгенштейна как на философа, развивающего близкое к нему понимание знания [Hintikka 2007, 12], и можно заметить, что изложенное в статье понимание знания близко по содержанию к некоторым витгенштейновским формулировкам.

бы предположить с учетом приведенной позиции о контекстной зависимости понятия знания.

Возникает вопрос о том, способна ли такая эпистемология что-то сказать нам о стратегии при условии, что стратегические принципы контекстно зависимы. Думаю, ответ на этот вопрос: может. Даже если она не сформулирует ни одного универсального стратегического принципа, она может описывать стратегию познания в целом. Нельзя не учитывать, что познание как таковое – игра с определенными свойствами, и эти свойства будут неизбежно отражаться на том, какими могут быть стратегические соображения для данной игры. Интеррогативная эпистемология выглядит способной выявлять такие зависимости и говорить о них. Фактически Я. Хинтикка это делает во многих фрагментах своих работ, например, когда рассуждает о том, что одним из видов стратегических соображений должна быть оценка надежности источника информации¹². Думаю, что более подробное, ясное и полное описание процесса познания «изнутри» (с точки зрения определяющих правил) в любом случае будет иметь положительное влияние на понимание стратегии познания. В ходе этого описания может оказаться целесообразным ввести какие-то конвенциональные правила стратегического мышления, которые будут признаны полезными самими игроками. Если бы такие правила были приняты, они расширили бы аппарат интеррогативной эпистемологии, позволив говорить о познании как об игре на метауровне и посредством этого еще дальше продвинуться в понимании природы последней. Подобным образом мы действуем по отношению к другим играм, когда объясняем новичку основы стратегии. Например, в самоучителях по шахматам содержится информация об условной стоимости различных фигур, выраженной в пешках. Данная информация не относится к правилам самой игры. Подразумевается конвенция, принятая шахматистами для облегчения стратегических расчетов. Конечно, в случае познания дело существенно осложняется тем, что познавательная игра недоопределена, пока не задан практический контекст исследования. Но, полагаю, стратегические конвенции все-таки возможны и в этом случае.

¹² В варианте исследования, где источником является не единственный «оракул», ответы которого всегда принимаются за истину, а один из множества различных источников, сведения, полученные от которых, могут быть отчасти или полностью ложными, – вариант детектива. У последнего много свидетелей, но кто-то из них может оказаться преступником или быть в сговоре с преступником.

Заключение

Интеррогативная эпистемология предстает перед нами как проект радикальной реформы эпистемологии с целью избавить ее от задач, ей не свойственных, и дать ей единый, концептуально обоснованный подход к решению ее собственных задач. Этот подход не претендует на то, чтобы дать ответы на все вопросы относительно познания. В частности, он *не* отвечает на вопрос о том, какая стратегия познания является лучшей в каждом случае. Он *не* дает способа формального описания всех когнитивных операций, которые совершает познающий агент, и, следовательно, не выступает в качестве готового языка для записи алгоритмов действия познающей машины. Он *не* отвечает на вопрос о том, возможна ли такая машина. Однако изложенный подход дает существенное продвижение в ряде связанных вопросов.

Так, применительно к познанию он способен четко разделить стратегические задачи от задач технических, связанных с применением «внутренних» правил познания. Он может исчерпывающим образом выразить «внутренние» правила так, что они становятся легко алгоритмизируемы. Логично предположить, что он даже способен помочь уточнить критерии выбора стратегии для конкретных контекстов познания (это относится к возможностям применения данного подхода в прикладной, а не в общей эпистемологии). Это было бы, несомненно, полезным вкладом в дело создания познающей машины, если окажется, что она все-таки возможна. Наконец, исследуемый подход помогает сформулировать и обосновать ряд важных концептуальных принципов, таких как параллелизм логической структуры дедуктивного и интеррогативного компонентов исследования и невозможность оценки того, насколько эффективен отдельный познавательный ход, вне контекста стратегии в целом¹³. Поэтому настоящую статью можно воспринимать как призыв к авторам, которые занимаются вопросами, связанными с эпистемологией и исследованиями искусственного интеллекта, осваивать и использовать изложенный подход.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Хинтиikka 2011 – Хинтиikka Я. Философские исследования: проблемы и перспективы // Вопросы философии. 2011. № 7. С. 3–17.

¹³ Ввиду ограниченного объема статьи эти принципы подробно не рассмотрены.

Хинтикка 1980 – *Хинтикка Я.* В защиту невозможных возможных миров // *Хинтикка Я.* Логико-эпистемологические исследования: сб. избр. ст. / под ред. В.М. Леонтьева. – М.: Прогресс, 1980. С. 228–244.

Хинтикка, Хинтикка 1987 – *Хинтикка Я., Хинтикка М.* Шерлок Холмс против современной логики. К теории поиска информации с помощью вопросов // *Язык и моделирование социального взаимодействия.* Переводы / общ. ред. В.В. Петрова. – М.: Прогресс, 1987. С. 265–281.

Knuuttila 2018 – *Knuuttila S.* Questions of Epistemic Logic in Hintikka // *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics* / ed. by H. van Ditmarsch, G. Sandu. – Cham: Springer International Publishing, 2018. P. 413–431.

Hintikka 1962 – *Hintikka J.* Knowledge and Belief: An Introduction to the Logic of the Two Notions. – Ithaca, New York: Cornell University Press, 1962.

Hintikka 1999 – *Hintikka J.* Inquiry as Inquiry: A Logic of Scientific Discovery. – Dordrecht; Boston: Kluwer Academic Publishers, 1999.

Hintikka 2007 – *Hintikka J.* Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning. – New York: Cambridge University Press, 2007.

REFERENCES

Knuuttila S. (2018) Questions of Epistemic Logic in Hintikka. In: van Ditmarsch H. & Sandu G. (Eds.) *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics* (pp. 413–431). Cham: Springer International Publishing AG.

Hintikka J. (2011) Philosophical Research: Problems and Prospects. *Vo-prosy filosofii*. No. 7, pp. 3–17 (Russian translation).

Hintikka J. (1980) Impossible Possible Worlds Vindicated. In: Hintikka J. *Logical-Epistemological Studies: Collection of Selected Articles* (pp. 228–244). Moscow: Progress (Russian translation).

Hintikka J. & Hintikka M.B. (1987) Sherlock Holmes Confronts Modern Logic: Toward a Theory of Information-seeking Through Questioning. In: Petrov V.V. (Ed.) *Language and Modelling of Social Interactions* (pp. 265–281). Moscow: Progress (in Russian).

Hintikka J. (1999) *Inquiry as Inquiry: A Logic of Scientific Discovery*. Dordrecht: Kluwer Academic Publishers.

Hintikka J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.

Hintikka J. (2007) *Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning*. New York: Cambridge University Press.