

УДК 51:101.8

DOI 10.15372/PS20230403

EDN WMGKKW

В.М. Резников**О КРИТЕРИЯХ ПОНИМАНИЯ В ЕСТЕСТВОЗНАНИИ,
ИССЛЕДОВАНИЯХ В ОБЛАСТИ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА И ФИЛОСОФИИ НАУКИ**

В статье сформулированы аргументы в пользу самостоятельной значимости проблемы понимания в современной философии науки. Очевидно, что невозможно создать серьезную концепцию понимания без исследования индикаторов понимания, используемых учеными. Изучив публикации в области искусственного интеллекта, естествознания и философии науки, автор описал критерии понимания, применяемые исследователями в этих областях знания. Основной результат состоит в предварительной классификации критериев понимания с учетом степеней универсальности этих критериев и степеней понимания изучаемых проблем, которые возникают у исследователей в процессе обдумывания и решения этих проблем при использовании формальных методов.

Ключевые слова: понимание без объяснения; интеллигибельная теория; интеллигибельная модель; понятный феномен

V.M. Reznikov**ON THE CRITERIA OF UNDERSTANDING IN NATURAL
SCIENCE, RESEARCH IN THE FIELD OF ARTIFICIAL
INTELLIGENCE AND PHILOSOPHY OF SCIENCE**

The article formulates arguments for the independent significance of the problem of understanding in contemporary philosophy of science. Obviously, it is impossible to develop a serious concept of understanding without examining the indicators of understanding used by scientists. Having studied publications in the field of artificial intelligence, natural science and philosophy of science, the author describes the criteria of understanding used by researchers in these fields of knowledge. The main result is a preliminary classification of the criteria of understanding in view of their degrees of universality and the degrees of understanding of the problems being studied, which researchers get in the course of thinking about and solving these problems when using formal methods.

Keywords: understanding without explanation; intelligible theory; intelligible model; understandable phenomenon

Для постпозитивистской философии, в отличие от позитивистской, характерно нейтральное отношение к метафизике, однако, как и в предшествующей философии, приоритет отдается естественным и формальным наукам. Другая особенность постпозитивизма, также доставшаяся в наследство от позитивизма, состоит в методологическом монизме. Согласно методологическому, монизму все научные дисциплины, в том числе социальные и гуманитарные, должны следовать нормам и принципам, принятым в естествознании, и прежде всего в физике как фламане естествознания [1].

В 40-е годы прошлого столетия логик К. Гемпель в духе постпозитивистской философии предложил так называемую дедуктивно-номологическую схему объяснений. Эта схема предназначена для объяснения сингулярных феноменов, событий, положений дел на базе общих законов и ранее обоснованных единичных утверждений. В простейшем случае схема Гемпеля является схемой покрывающего закона. Она представляет собой триаду утверждений следующей структуры, а именно в ней две посылки и заключение [2]:

$$\frac{(\forall x)(H(x) \rightarrow E(x)) \\ H(b)}{E(b)}$$

Первая посылка схемы является общим суждением, вторая – частным суждением, последнее суждение – это заключение. Здесь $H(b)$ является посылкой, а $E(b)$ – заключением.

Схема Гемпеля была весьма популярной и до сих пор не лишена некоторой популярности, однако с самого начала работа Гемпеля не была свободна от критики. Сформулируем кратко недостатки схемы Гемпеля, так как ее критика в конечном счете привела к выделению проблемы «понимание» в самостоятельное направление исследований.

1. Материальная импликация несовершенна. Неадекватность материальной импликации состоит в том, что при ложной посылке утверждение оказывается истинным, или, как говорят логики, «из

лжи следует все что угодно». В качестве примера приведем известный парадокс красного яблока. Орнитолог изучает черных ворон, пытаясь обосновать гипотезу, что все вороны черные, или фальсифицировать эту гипотезу с помощью контрпримера. Он рассуждает следующим образом. Быть вороной – значит обладать черным цветом. Отсюда по принципу контрапозиции получаем: если не обладать черным цветом, то не быть вороной. После долгих наблюдений за воронами орнитолог решает подкрепиться. Он открывает холодильник, видит красное яблоко и рассуждает: «Оно не обладает черным цветом и не является вороной, значит, красное яблоко подтверждает гипотезу о том, что все вороны черные», – что является абсурдным.

2. Схема Гемпеля, опирающаяся на общие законы, неадекватна для исторических исследований. Критика схемы Гемпеля известным методологом в области истории У. Дреем заключалась в том, что историкам для понимания поступков государственных деятелей важны мотивы их поступков, а не общие законы истории [3]. Отметим еще критику Г.Х. фон Вригтом некоторых обобщений в истории на примере с Людовиком XV [1].

3. Схема не представляет интереса для формальных наук, так как в математике все выводится, но не все оказывается понятным.

4. Схема не всегда подходит для естественных наук. Например, в случае двух и более законов непонятно, какой закон является их обобщением [7].

5. Не для всех изучаемых феноменов известны теоретические законы. Например, для изменения облаков известны экспериментальные законы [7].

6. Схема не защищена от нерелевантных объяснений. Пусть дано общее утверждение такого вида: «Все мужчины, принимавшие противозачаточные таблетки, не забеременели. Джон Джойс принимал противозачаточные таблетки. Следовательно, он не забеременел» [12].

7. Схема Гемпеля имеет минимальные возможности для исследования проблемы понимания. Начиная с М. Фридмана, концепцию Гемпеля критиковали за минимальные ресурсы для исследования проблемы понимания [9]. Согласно Гемпелю, в дедуктивно-номологической схеме понимание возникает в двух случаях. Во-первых, понимание оказывается побочным продуктом основательных объяснений. Во-вторых, понимание возникает в случае реали-

зации ожидаемых событий. Второе условие не является особо серьезным. Приведем известный пример ожидаемых событий: после грома появляется молния, но отсюда не следует, что после раската грома и вспышки молнии эти атмосферные явления становятся понятными. Гемпеля интересовали логические аспекты проблемы объяснения, прагматика же была вне сферы его интересов, однако в случае понятия «понимание» невозможно ограничиться логикой. Действительно, вполне допустимо, что одному исследователю теория представляется очень понятной, другому – не особо понятной, а третьему – совершенно непонятной. Не вызывает сомнений, что серьезные научные результаты могут быть получены с использованием моделей и теорий, понятных исследователям, поэтому невозможно избежать прагматики при изучении проблемы «понимание».

Со временем интерес к проблеме понимания в философии науки усиливался, и приблизительно два десятилетия назад в ней выделилось новое направление под названием «философия понимания», со специальным интересом к пониманию без объяснений. Дело в том, что для науки не является характерной апелляция к объяснениям, исследователи заинтересованы в обретении понимания, например, в результате осуществления мысленных или реальных экспериментов.

Рассмотрим кратко базовые идеи концепции Х. де Регта, используемые при исследовании проблемы понимания в науке [8]. В этой концепции основными понятиями являются понятия интеллигибельной, или понятной, теории и интеллигибельной модели. Отметим, что свойство интеллигибельности не является внутренним свойством теории, а этой характеристикой она наделяется учеными, которые ее эффективно применяют. Изучаемые явления и положения дел становятся понятными на основе применения интеллигибельных теорий и моделей. Приведем данное де Регтом определение интеллигибельной теории: «Научная теория T является интеллигибельной для ученых, если они могут качественно понять характерные особенности заключений на основе T без осуществления точных вычислений» [8]. Фактически де Регт предложил критерий понимания для естественно-научных теорий. Этот критерий близок к требованиям Р. Фейнмана о понимании уравнений, описывающих физические явления: уравнения являются понятными, если исследователь способен получить их качественное, практически правильное решение, не производя вычислений.

На наш взгляд, подход де Регта заслуживает внимания. В частности, представляется перспективным ознакомление с критериями понятности, используемыми учеными, на основе изучения публикаций и контактов с работающими учеными, с тем чтобы создать некоторый банк определений понимания в различных областях знания. Наша рабочая гипотеза состоит в том, что анализ этих определений позволит выявить некоторые общие подходы и специфические идеи ученых относительно критериев понимания. В данной статье на основе анализа публикаций в таких областях знания, как естествознание и искусственный интеллект, будут описаны требования ученых к возникновению понимания решаемых проблем с учетом особенностей полученных ими решений, а также специфики формальных методов, используемых при решении этих проблем. Итогом станет классификация этих критериев с учетом их степени общности и качества полученных результатов.

Сначала опишем критерии, претендующие на универсальный статус. Будем полагать, что утверждение о понимании решаемой задачи эквивалентно утверждению о понимании событий, феноменов и положений дел, изучаемых в рамках этой задачи.

Сведение решаемой задачи к ранее решенной понятной задаче. По-видимому, среди известных критериев понимания это наиболее универсальный и часто используемый критерий.

Конструирование предсказаний, которые оказались оправданными. Степень уверенности возрастает и, возможно, у исследователя имеет место понимание в том случае, когда используемые модели обеспечили построение предсказаний и их правильность была подтверждена экспериментально. Критерий на основе предсказаний не является в полной мере универсальным. Так, в области квантовой физики исследователи умеют корректно строить модели, с помощью которых получают предсказания, последние с поразительной вероятностью реализуются, однако до сих пор физики не знают механизмы квантово-механических явлений природы.

Приведем несколько критериев, в которых используются характеристики успешности решенных задач и степени их понимания. Начнем с универсального критерия, который свидетельствует о предпонимании, или минимальном понимании, решаемой задачи.

Критерий минимального понимания. Исследователь обладает минимальным пониманием решаемой задачи, если он безошибочно определяет, что задача решена корректно, когда она действительно

решена корректно. При этом не принципиально, кто решил задачу – сам исследователь или кто-то другой, и даже не принципиально, решена ли эта задача. Такого плана критерий был сформулирован в совместных публикациях известного математика академика Ю.Л. Ершова и доктора философских наук К.Ф. Самохвалова при работе над концепцией задачного подхода. Эта концепция была разработана в связи с исследованием влияния ограничительных теорем Геделя на работающую математику. Ю.Л. Ершов и К.Ф. Самохвалов показали, что корректность каждой отдельной математической дисциплины обосновывается независимо от этих теорем, поэтому они не имеют особого значения для обоснования работающей математики. Приведем критерий минимального понимания задачи, сформулированный этими авторами: «мы *понимаем* задачу только тогда, когда ей сопоставили обоснованное чувство уверенности в том, что всякое состояние нашего сознания мы сумеем убедительным и безошибочным образом распознать как такое, когда решение найдено, или как такое, когда решение не найдено» [4].

Теперь сформулируем несколько критериев понимания основанных на характеристиках качества полученных решений. Приведем два критерия максимального понимания.

Критерий качественного решения задачи Фейнмана. Знаменитый физик Р. Фейнман утверждал, что на основе понятной теории он понимает изучаемые явления или понимает задачу, в которой они исследуются, если ему удастся качественно решить уравнения, описывающие эти явления, до того, как эти уравнения были фактически решены [5].

Критерий аналитического решения задачи Эбботта. Известный специалист в области электроники Д. Эбботт отмечает, что когда физика занималась изучением макроэлектронных явлений, они были глубоко поняты и для уравнений, описывающих эти явления, были получены аналитические решения, при этом практически не было необходимости в использовании компьютеров. А сейчас физики изучают очень сложные микроэлектронные явления, и они до сих пор не столь глубоко изучены, при этом уравнения, описывающие явления в микроэлектронике, не удается решать аналитически, они допускают только численные решения [6].

Критерий понимания уравнений на основе их итерационных решений. Данный критерий относится к исследованиям в области техники, точного естествознания и т.д. Предположим, что изучае-

мые положения дел описываются уравнениями, которые допускают итеративное решение. Если исследователь на основе нескольких итераций приходит к качественно правильному решению до того, как уравнения фактически решены, то естественно считать, что получено решение хорошего качества, и это свидетельствует о весьма достойном, не хуже среднего уровня понимании изучаемой проблемы [10].

Обратимся к области искусственного интеллекта. В этой области знания универсальным представляется следующий критерий понимания.

Общий критерий понимания в области искусственного интеллекта. Исследователь понимает некоторый класс задач, если он способен научить компьютерную программу понятному для исследователя решению этих задач. Этот критерий был предложен Дж. Перлом, известным специалистом в области искусственного интеллекта [11].

Теперь обратимся к сформулированной Дж. Перлом концепции искусственного интеллекта, так как для исследователей этого направления адекватным представляется комбинированный критерий, базирующийся на общем критерии и философском критерии понимания на основе причинных отношений. Почему в критерии понимания Перла имеется философская составляющая? По мнению Перла, отличительной особенностью естественного интеллекта является овладение причинными отношениями на уровне обыденного, практического и теоретического мышления. Поэтому овладение причинными отношениями оказывается необходимым, но недостаточным условием совершенствования искусственного интеллекта для приближения его к естественному интеллекту. Перл полагает, что многочисленные недостатки искусственного интеллекта будут в той или иной степени преодолены на пути познания причинных отношений и их дистинкции от отношения корреляции. Философской составляющей концепции является причинная интерпретация тезиса Тьюринга.

Каузальная концепция Перла представлена на трех уровнях, и только на втором действительно описаны причинные отношения. На первом уровне рассмотрены не причинные отношения, а эмпирические корреляции. Так, менеджер в магазине знает, что существует положительная связь между такими событиями, как покупка зубной пасты и покупка зубной щетки. На втором уровне исследователи создают условия, которые обеспечивают воспроизводство изучаемых феноменов. Эти условия выполняют функцию дейст-

вующей причины. На третьем уровне рассматриваются контрфактические суждения. Это условные суждения, в которых изменяется их условная часть и предполагается реализация адекватной модификации заключения. Предположим, что задано утверждение: «Если работник закончил экономический факультет университета и занимает должность старшего менеджера, то он зарабатывает N денежных единиц». Теперь изменим условную часть этого суждения и соответственно модифицируем заключение: «Если работник является студентом третьего курса экономического факультета университета и занимает должность менеджера, то он зарабатывает M денежных единиц».

Отметим, что в отличие от ранее рассмотренных критериев понимания, основанных на характере достигнутых результатов, критерии понимания в концепции Перла могут быть основаны не только на характере достижений, но дополнительно и на использовании причинных объяснений. Если более подробно, то критерии понимания в концепции Перла могут быть основаны на овладении исследователем методологии причинных отношений на всех трех уровнях, а еще более подробно – на успешном обучении причинным отношениям компьютерных программ в том случае, если они достигают понятные исследователю решения задач в области причинных отношений, и наконец, на понимании на базе причинных объяснений.

На основе изучения публикаций в области философии науки, естествознания и искусственного интеллекта нами построена предварительная классификация критериев понимания, применяемых исследователями, работающими в этих областях знания, с учетом степени общности критериев и некоторых характеристик успешности полученных результатов. В случае продолжения исследований и построения подробной классификации такая детальная классификация может быть полезной для оценивания результатов исследований в некоторых областях знания, где используются формальные методы, для изучения критериев понимания, для исследования критериев понимания, реально применяемых учеными.

Литература

1. Вригт Г.Х., фон. Объяснение и понимание // Вригт Г.Х., фон. Логико-философские исследования: Избранные труды / Под ред. Г.А. Рузавина и В.А. Смирнова. М.: Прогресс, 1986.

2. Гемпель К. Логика объяснения. М.: Дом интеллектуальной книги, 1988.
3. Дрей У. Еще раз к вопросу об объяснении действий людей в исторической науке // *Философия и методология истории* / Под ред. И.С. Кона. М.: Прогресс, 1977.
4. Ершов Ю.Л., Самохвалов К.Ф. Современная философия математики: недомогания и лечение. Новосибирск: Параллель, 2007.
5. Фейнман Р., Лейтон Р., Сэндс М. Фейнмановские лекции по физике. Т. 3: Электричество и магнетизм. М.: АСТ, 2020.
6. Abbott D. The reasonable ineffectiveness of mathematics // *Proceedings of the IEEE*. 2013. Vol. 101, No. 10. P. 2147–2153. DOI: 10.1109/JPROC.2013.2274907.
7. Cartwright N. *How the Laws of Physics Lie*. Oxford: Oxford University Press, 1983.
8. De Regt H. *Understanding Scientific Understanding*. Oxford: Oxford University Press, 2017.
9. Friedman M. Explanation and scientific understanding // *The Journal of Philosophy*. 1974. Vol. 71 (1). P. 5–19. DOI: 10.2307/2024924.
10. Lenhard J. Simulation modeling and scientific understanding // *Scientific Understanding: Philosophical Perspectives* / Ed. by H. de Regt, S. Lionelli. Pittsburgh University Press, 2009.
11. Pearl J., Mackenzie D. *The Book of Why: The New Book of Cause and Effect*. Basic Books, 2020.
12. Salmon W. Statistical explanation // *Statistical Explanation and Statistical Relevance* / Ed. by W. Salmon, R. Jeffrey, J. Greeno. Pittsburgh University Press, 1971.

References

1. Wright, G.H., von. (1986). Obyasnenie i ponimanie [Explanation and understanding]. In: Wright, G.H., von. *Logiko-filosofskie issledovaniya: Izbrannye trudy* [Logic and Philosophical Studies: Selected Works]. Ed. by G.A. Ruzavin and V.A. Smirnov. Moscow, Progress Publ. (In Russ.).
2. Hempel, K. (1988). Logika obyasneniya [The Logic of Explanation]. Moscow, Dom Intelktualnoy Knigi Publ. (In Russ.).
3. Dray, W. (1977). Eshche raz k voprosy ob obyasnennii deystviy lyudey v istoricheskoy nauke [The historical explanation of actions reconsidered]. In: Kon, I.S. (Ed.). *Filosofiya i metodologiya istorii* [Philosophy and Methodology of History]. Moscow, Progress Publ. (In Russ.).
4. Ershov, Yu.L. & K.F. Samokhvalov. (2007). *Sovremennaya filosofiya matematiki: nedomoganiya i lechenie* [Modern Philosophy of Mathematics: Ailments and Treatment]. Novosibirsk, Parallel Publ.
5. Feynman, R., R. Leighton & M. Sands. (2020). *Feinmanovskie lektzii po fizike. T. 3: Elektrichestvo i magnetizm* [The Feynman Lectures on Physics. Vol. 3: Electricity and Magnetism]. Moscow, AST Publ. (In Russ.).
6. Abbott, D. (2013). The reasonable ineffectiveness of mathematics. *Proceedings of the IEEE*, 101 (10), 2147–2153. DOI: 10.1109/JPROC.2013.2274907.
7. Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.
8. De Regt, H. (2017). *Understanding Scientific Understanding*. Oxford University Press.
9. Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, 71 (1), 5–19. DOI: 10.2307/2024924.

10. *Lenhard, J.* (2009). Simulation modeling and scientific understanding. In: De Regt, H. & S. Lionelli (Eds.). *Scientific Understanding: Philosophical Perspectives*. Pittsburgh University Press.

11. *Pearl, J. & D. Mackenzie.* (2020). *The Book of Why: The New Book of Cause and Effect*. Basic Books.

12. *Salmon, W.* (1971). Statistical explanation. In: Salmon, W., R. Jeffrey & J. Greco (Eds.). *Statistical Explanation and Statistical Relevance*. Pittsburgh University Press.

Информация об авторе

Резников Владимир Моисеевич – Институт философии и права СО РАН (630090, Новосибирск, ул. Николаева, 8); Новосибирский исследовательский государственный университет (630090, Новосибирск, ул. Пирогова, 2).
mathphil1976@gmail.com)

Information about the author

Reznikov, Vladimir Moiseevich. – Institute of Philosophy and Law, Siberian Branch of the Russian Academy of Sciences (8, Nikolaev st., Novosibirsk, 630090); Novosibirsk National Research State University (2, Pirogov st., Novosibirsk, 630090, Russia).

Дата поступления 06.11.2023