Современные методы познания

Философские проблемы моделирования мультиагентных систем*

И.Ф. Михайлов
Институт философии РАН, Москва, Россия

DOI: 10.30727/0235-1188-2018-12-56-74
Оригинальная исследовательская статья

Аннотация

Мультиагентные системы (МАС) – это технология, основанная на распределенных вычислениях, предназначенная для эмуляции сложного эмерджентного поведения множества агентов на основе различных механизмов обратной связи. Философский интерес к ним связан, во-первых, с когнитивными свойствами отдельных агентов и системы в целом; во-вторых, с возможностями моделирования сложного социального поведения и, возможно, более глубокого понимания социальности как таковой. Главная проблема – в какой степени мультиагентные модели соответствуют моделируемой реальности. Поскольку они представляют собой вычислительные системы, необходимо рассмотреть общую теорию вычислений и их онтологический статус; определить, каким образом вычислительные процессы формируют правилосообразные системы, состоящие из множества когнитивно-нагруженных агентов; сделать правдоподобные предположения относительно закономерностей эволюции таких систем в естественных и искусственных условиях. Особое внимание уделяется в статье гипотезе ограниченной рациональности агентов или, что то же самое, их когнитивной ограниченности, что делает необходимым использование аппарата эпистемической логики для формализации процессов в МАС. Полученные результаты дают дополнительные аргументы в пользу известного философского тезиса: высшие когнитивные способности определяются и объясняются сложной социальной организацией их носителей.

Ключевые слова: вычисление, репрезентация, когнитивная наука, мультиагентные системы, роевой интеллект, рациональный агент,

* Исследование выполнено в рамках проекта Российского фонда фундаментальных исследований (РФФИ) «Когнитивные основания социальности», грант № 18-011-00316А.

когнитивные системы, искусственный интеллект, эпистемическая логика, социальная гиперсеть.

Михайлов Игорь Феликсович – кандидат философских наук, старший научный сотрудник сектора методологии междисциплинарных исследований человека, Институт философии РАН.

ifmikhailov@gmail.com

http://orcid.org/0000-0001-8511-8849

Цитирование: *Михайлов И.Ф.* (2018) Философские проблемы моделирования мультиагентных систем // Философские науки. 2018. № 12. С. 56–74.

DOI: 10.30727/0235-1188-2018-12-56-74

Philosophical Problems of Multi-Agent Systems Modeling*

I.F. Mikhailov

Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

DOI: 10.30727/0235-1188-2018-12-56-74

Original research paper

Summary

Multi-agent systems (MAS) is a technology based on distributed computing, designed to emulate the complex emergent behavior of multiple agents on the basis of various feedback mechanisms. Philosophical interest therein is caused, firstly, by the cognitive properties of individual agents and of the system as a whole; secondly, with the possible modeling of complex social behavior and, possibly, by the perspective of a progress in social understanding. The main problem is the extent to which multi-agent models correspond to the simulated reality. Since they are computing systems, it is necessary to consider the general theory of computations and their ontological status; to determine how the computational processes form rule-like systems consisting of multiple cognitively enabled agents; to make plausible assumptions about the regularities of the evolution of such systems under natural and artificial conditions. Particular attention is paid to the hypothesis of the limited rationality of agents or, what is the same, to their cognitive limitations, which demands the application of epistemic logic formalism to MAS functioning. The obtained results give additional arguments in favor of the well-known philosophical thesis: the higher cognitive

* The research is completed within the framework of the project of the Russian Foundation for Basic Research (RFBR) "Cognitive grounds of social being," grant no. 18-011-00316A.

capabilities are determined and explained by the complex social organization of their bearers.
Keywords: computation, representation, cognitive science, multi-agent systems, swarm intelligence, rational agent, cognitive systems, artificial intelligence, epistemic logic, social hypernet.

Igor Mikhailov – Ph.D. in Philosophy, Senior Research Fellow at the Department of Methodology of the Interdisciplinary Study of Man, Institute of Philosophy, Russian Academy of Sciences.
ifmikhailov@gmail.com
http://orcid.org/0000-0001-8511-8849

Citation: Mikhailov I.F. (2018) Philosophical Problems of Multi-Agent Systems Modeling. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. 2018. No. 12, pp. 56–74.
DOI: 10.30727/0235-1188-2018-12-56-74

1. Введение. Вычисления и вычислительные системы

После научной революции XVII в., когда дескриптивные теории мира постепенно ушли из-под сени спекулятивной философии, за нею остались лишь сфера сознания и его познавательных способностей, а также сфера человеческой социальности с ее надприродными правилами, такими как мораль. У классической науки не было концептуальных средств, чтобы справиться с явлениями подобного рода. Поэтому для одних они были и во многом остаются свидетельствами присутствия в мире чего-то «нефизического», для других – основанием для отделения «наук о духе» от «наук о природе» в той или иной методологической форме.

Однако начиная со второй половины XX в. – практически на наших глазах – зарождается и формируется нечто, что может стать (а может и не стать) основой следующей научной революции. Я имею в виду общую теорию вычислений и научные направления, возникшие и возникающие в той или иной мере на ее основе – комплекс когнитивных наук, вычислительную биологию, вычислительную нейрофизиологию, такую же экономику, термодинамику и даже астрономию. Дело не только в том, что компьютеры заполонили офисы, стали средством моделирования практически любых процессов и полезной метафорой для объяснения когнитивных способностей. Дело еще и в том, что сама теория вычислений (компьютации), которая способствовала в своем первозданном – тьюринговском [Turing 1936] – виде этой

компьютерной революции, претерпевает важные изменения, которые, возможно, позволят ей претендовать на статус новой общенаучной парадигмы.

Алан Тьюринг заложил основы общей теории вычислений и вычислимости в ее наиболее строгой – символьной – форме, и многими этот подход принимается за теорию вычислений как таковую, вплоть до того, что словосочетание «нетьюринговы вычисления» воспринимается ими как оксюморон. Однако появление различных архитектур параллельных вычислений, включая всем теперь известные нейросети, развитие концепций нано- и квантовых вычислений, а также так называемых суперкомпьютеров, поставило под вопрос представление о применимости тьюринговой модели для всего множества вычислительных процессов. Вместе с этой проблемой встал вопрос и о явном определении вычислений как таковых.

Но, что более важно для науки, вычисления, как бы они ни понимались, все больше перемещаются из сферы инструментов познания, моделей и полезных метафор в онтологическую сферу предметов познания: биологические, экономические, астрономические и др. объекты многими понимаются теперь как вычислительные системы, что позволяет применить к ним нетрадиционные теоретические средства.

Так, представляет интерес и кажется перспективной концепция натуральных вычислений как способа сокращения термодинамических затрат в природных процессах [Kempes et al. 2017]. В статье утверждается, в частности, что социальные системы наименее термодинамически эффективны по сравнению с биологическими и даже астрономическими. Но тем не менее они все же являются вычислительными системами. Однако что отличает вычислительные процессы от невычислительных в природе? Большинство авторов сходятся в признании двух критериев: формально-алгоритмического характера процесса и его независимости от материального субстрата, или, как вариант, его переносимости на любой другой материальный субстрат с тем же или большим количеством степеней свободы [MacLennan 2004, 121; Piccinini Bahar 2013, 458, 464]. С этой точки зрения, такие процессы, как приготовление пищи или пищеварение, несомненно, не являются вычислительными, поскольку существенно зависят от свойств участвующих материальных субстратов. Напротив, такие процессы, как трансляция РНК или, тем более, нервные процессы, эко-

номические процессы, в основном признаются вычислительными. А вот по поводу астрономических систем согласия нет. Такое разнообразие мнений, на мой взгляд, отражает пока не преодоленную концептуальную неопределенность самого понятия вычисления, вышедшего за тьюринговы рамки. Очевидно, что для достижения применимости этого понятия к природным процессам необходимо избавиться от антропоморфной презумпции, что вычисления – это когда кто-то вычисляет (считает) что-то с какой-то целью. Попробуем, опираясь в значительной степени на интуицию, сформулировать наше понимание следующим образом: *вычисления – это такие процессы, каузальная роль которых определяется не их материальным субстратом, а их формально-алгоритмической структурой.* То есть *система является вычислительной, если при замене ее на другую, но функционально идентичную систему причинно-следственная цепочка, звеном которой она является, остается без изменений.*

Так, если для ориентации в пространстве одни живые организмы используют анализ отраженных электромагнитных волн, а другие – ультразвуковой эхолот, то и тот и другой инструменты представляют собой вычислительные системы, чья каузальная роль в значительной степени не зависима от физической реализации, а определяется исполняемыми алгоритмами.

Такое определение, впрочем, порождает две концептуальные проблемы. Первая связана с философской неопределенностью таких понятий, как материальный субстрат. Очевидно, что мы не можем полагаться на феноменальные определения вещей и веществ природы, участвующих в интересующих нас процессах, – их цвет, теплоту, вкус и т.п. Но их объективная определенность, в свою очередь, сводится к таким формально-структурным определениям, как атомное или молекулярное строение и т.п., т.е., к структурам и процессам, относительно которых также можно поставить вопрос об их вычислительной природе. И тогда возникает своего рода «искушение панкомпьютационализма», основанное на возможности представить весь мир как иерархию вычислительных систем и процессов, в которой результаты вычислений на более «низких» уровнях участвуют в вычислениях на более «высоких» в качестве репрезентаций. Однако весь спектр философских импликаций такого подхода, учитывая их крайнюю дискуссионность, вынужденно остается за пределами данной статьи.

Вторая проблема связана с вынужденной «метафоричностью» определяемого понятия: кто-то может возразить, что коль скоро мы отказываемся от антропоморфного смысла слова «вычисление» (1), то стоит ли его вообще оставлять в качестве научного термина. Да, можно попытаться подыскать более подходящее слово. Но, с другой стороны, столь же метафоричными были слова «сила» или «поле» на заре становления классической физики. Однако определенный релевантный аспект их смысла, вкупе со строгими математическими определениями, позволил им надолго задержаться в корпусе науки таким образом, что никто уже не испытывает никакого дискомфорта при обращении с ними. Кроме того, в некотором существенном отношении *человек вычисляющий* со всеми его целями и смыслами «тоже является частью Вселенной» (2). И это оставляет надежду на то, что когда-нибудь сбудется мечта молодого Маркса о единой науке о природе и человеке.

Существуют две вычислительные модели, которые, как мне кажется, наиболее продуктивны для исследования когнитивных и социальных функций человека. Это коннекционистская модель распределенных вычислений, на которой так или иначе, с теми или иными вариациями, построены современные искусственные нейронные сети (ИНС) и методики глубокого (машинного) обучения (deep learning). И это мультиагентные системы (МАС), которые также достаточно «реалистично» воспроизводят когнитивно-нагруженные социальные взаимодействия.

Обе модели относятся к классу так называемых «биологически вдохновленных» (bio-inspired) ИИ-парадигм – имеются в виду подходы в когнитивной науке, отказывающиеся от классической символьной вычислительной модели по типу машины Тьюринга в пользу архитектур, почерпнутых из эмпирических наук о живых существах (коннекционизм, «воплощенное познание») или навеянных математическими теориями («динамические системы», «предсказывающий ум»). В целом у нетьюринговых концепций вычислений есть преимущества и недостатки. Первые сводятся к тому, что такие концепции преодолевают концептуальные ограничения классической модели, закрепляющие ее антропоморфизм, поскольку она предполагает только символы (требующие интерпретатора), только правилосообразные манипуляции. Отказ от этих ограничений позволяет «вывести» вычислительные процессы за пределы человеческого мира, в которых они непременно предполагают того, кто вычисляет, осознаваемый предмет

и осознаваемую цель вычислений. Вместе с тем – и это можно отнести к их недостаткам – как правило, нетьюринговы модели вычислений оказываются привязанными к определенной предметной области и связанными определенными эпистемологическими презумпциями, что, впрочем, справедливо и для классической модели [MacLennan 2004]. Неклассические, нетьюринговы модели вычислений противопоставляют классической, обычно исходя из их предполагаемой континуальности в противовес дискретности (близкий, но не тождественный вариант – аналоговый характер в противовес цифровому), или исходя из концепции *функционального механизма* [Bechtel 2008; Craver 2007; Glennan 2002; Wimsatt 2002] в противовес концепции машины Тьюринга (МТ), или противопоставляя распределенные (параллельные) вычисления линейным или серийным. В рамках последней антитезы ИНС и МАС представляются примерами реализации неклассической модели вычислений, хотя и этот пункт остается дискуссионным.

2. Теория и технологии мультиагентных систем

Теперь подробнее о мультиагентных системах. Мультиагентные системы представляют собой класс биологически-вдохновленных (bio-inspired) вычислительных систем, в которых взаимодействуют автономные рациональные агенты, наделенные обратной связью и способностью принимать решения. Они запрограммированы на преследование определенных целей, имеют ограниченный набор возможных состояний и/или действий, которые принимают или предпринимают, реагируя на свойства среды и действия других агентов по определенным правилам. Интеллектуальность системы обнаруживается как результирующая характеристика поведения всего «роя», поскольку все последующие состояния системы эмерджентны и непредсказуемы, но в высокой степени адаптивны. Как и в случае с нейросетями, разработчики управляют МАС с помощью настроек и локального программирования, гипотетически предполагая будущее поведение системы.

Согласно нестрогому определению Ниази и Хуссейна, «агент был бы чем-то таким в системе, что может быть рассмотрено и идентифицировано экспертами в любой дисциплине и интересующем субдомене как играющее в этой системе важную, индивидуальную, интерактивную и интересную роль, приводящую к поведению, которое называют интеллектуальным, рациональным, когнитивным или эмерждентным» [Niazi, Hussain 2014, 2–3].

Не претендуя на профессиональное суждение в этой области, я готов предположить, что необходимым условием саморегуляции в обоих случаях выступает та или иная форма обратной связи – алгоритм обратного распространения ошибки в случае с ИНС и зависимость действий агента от положения или поведения других агентов в случае с МАС.

Основы данной технологии формировались еще в 1980-х гг. в виде объектно-ориентированного программирования, распределенных вычислений, первых экспериментов в области «искусственной жизни» и «искусственных обществ», а также искусственного интеллекта и когнитивных систем. Значительную роль сыграли также данные бурно развивающихся биологии, этологии и т.п. Особенности этого класса ИИ-систем часто обозначают терминами «эмерджентный интеллект» или «интеллект роя», имея в виду рациональное, но в целом непредсказуемое их поведение.

Сегодня область практического применения МАС оказалась значительно шире той, что планировалась изначально. К ней относятся: экономика [Городецкий 2014; Редько Сохова 2013], приборостроение [Афанасьев 2012], техническая стандартизация [Аронов и др. 2015] и даже управление беспроводными сетями [Thomas 2007].

3. Предшественники мультиагентных систем

В.Л. Макаров описывает первые попытки создания «искусственной жизни» и «искусственных обществ», в ходе которых вырабатывались основные принципы мультиагентных моделей [Макаров 2009].

Что касается «искусственной жизни» (AL – artificial life), то начало этим исследованиям было положено в докомпьютерных исследованиях клеточных автоматов. Наиболее известным воплощением этой простой технологии стала «сахарная модель» (SUGARSCAPE) (3), на которую ссылаются и зарубежные исследователи, далее упомянутые в этой статье. По мере развития искусственных обществ в направлении большего социального реализма все более релевантной становилась *гипотеза ограниченной рациональности агентов*, которая отражала факт ограниченности их когнитивных ресурсов, таких как объем памяти [Макаров 2009, 18]. Эта гипотеза имеет важное значение и для современных социально-когнитивных теорий, поскольку ясно,

что конфигурации реальных социальных сетей определяются в том числе и тем фактом, что их участники не могут воспринять и удержать информацию о состоянии системы в целом. Кроме того, это добавляет оснований для использования когнитивных моделей в социальных симуляциях, что сегодня является предметом активного обсуждения.

По мнению автора, элементами социальной реальности должны считаться действия, а не люди [Макаров 2009, 19]. Мне близок этот подход, так как, согласно моей собственной концепции [Михайлов 2017], в рамках конкретнонаучных социальных онтологий действия могут быть интерпретированы как отношения. Если мы предлагаем онтологию, в которой объекты не только вступают в отношения, но и наделены свойствами, то мы обрекаем себя на поиски некой метаонтологии, призванной объяснить природу этих свойств. Если же нам удается свести свойства к отношениям, то необходимость в метаонтологии отпадает, сама онтология упрощается, поскольку объекты примитивизируются, а свойства оказываются эмерджентными эффектами межобъектных взаимодействий. Система объектов и отношений между ними обретает форму сети, которая как упрощенное описание реальности релевантна как когнитивным, так и социальным наукам.

Философское значение реальных МАС и научных теорий, за ними стоящих, состоит в осознании эмерджентности социальной жизни, возникающей на основе правил, регулирующих поведение рациональных, но когнитивно ограниченных агентов.

4. Когнитивные распределенные вычисления

Концептуальный анализ мультиагентных систем позволяет пересмотреть распространенный философский предрассудок, согласно которому в основе возникновения высших когнитивных – собственно интеллектуальных – способностей лежит сравнительно более развитый мозг. Если перевести это убеждение на язык вычислений, мозг оказывается главным и единственным процессором, возрастающая сложность которого в какой-то момент позволяет появиться социальным взаимодействиям. Но, опираясь на опыт МАС, мы можем теперь говорить о *распределенных социальных вычислениях*, процессором которых оказывается мультиагентная система в целом. Репрезентации этих социальных вычислений в когнитивных аппаратах агентов обнаруживаются как «рациональное» мышление.

Теория рефлексивного управления [Новиков, Чхартишвили 2013] как один из вариантов теории МАС показывает, что интеллектуальные эффекты в распределенной сети рациональных агентов могут быть поняты как следствие сложного правилосо-образного взаимодействия на основе автономных решений и обратной связи. И выглядит правдоподобным предположение, что когнитивный аппарат агента в системе – будь это естественный мозг или искусственный процессор – должен иметь или эволюционно развить в себе блок функций, обеспечивающих социальные взаимодействия и способность манипулировать ими в собственных интересах. Если же взаимодействующие агенты научаются пользоваться символическими системами, то когнитивная «мощность» системы в целом и каждого конкретного агента возрастает многократно.

Концептуализируя ту же мысль об эмерджентности социального интеллекта, Сингх [Singh 1994, 11] говорит о *коммуникативных интенциях*. Мне больше импонирует понятие *коммуникативных модальностей*, поскольку оно, во-первых, указывает на специфические логические структуры коммуникативных ситуаций, а во-вторых позволяет и оправдывает использование эпистемических логик по отношению к известным интенциональным актам – таким как знание, полагание, сомнение, опасение и т.п.

5. Мультиагентная природа социальности

Социальная система не может состоять из некогнитивных агентов, иначе трудно себе представить, как было бы возможно ее функционирование. Агенты, объединенные в социум должны обладать рядом способностей, среди которых – самоидентификация и идентификация себе подобных, осмысленная коммуникация, запоминание и реализация социальных сценариев. Для этого им нужен блок ввода и обработки данных, информационный процессор, память, модуль принятия решений. Таким образом, физические условия когнитивности одновременно являются необходимыми физическими условиями социальности. Можно ли тогда указать на достаточные условия бытия агентов в качестве общественных единиц? МАС приоткрывают эту тайну: социальность обеспечивается способностью действовать по правилу – отвечать определенным действием на определенные действия соседей или на состояние среды. В основе этой способности, конечно же, лежит целый ряд когнитивных функций.

И здесь возникает трудный методологический вопрос: если МАС – это вычислительные системы, то как они могут реализовывать какой-либо алгоритм, если неизвестны их конечные состояния? Можно предположить, что понимание вычисления как реализации одного заранее данного алгоритма – еще один предрассудок. Так же как и в нейросети, алгоритмы здесь если и реализуются, то на уровне агентов (нейронов, в случае НС). А собственно вычисление как процесс получения выходных данных на основе входных реализуется на уровне системы в целом через смену ее состояний. Именно поэтому мы говорим о параллельности таких вычислений и эмерджентности их результатов. Вместе с тем между этими распределенными вычислительными архитектурами есть одно заметное различие: между социальными агентами нет физических – например, синаптических – связей. Они часто свободны в движениях и перемещениях, обеспечивая гибкость и динамичность конфигурации сети, улучшая ее адаптивные свойства. Более того, физический состав агентов может меняться: какие-то из них могут выбывать, другие – вливаться в процесс, но система остается тождественной себе, пока правила участия в ней остаются существенно неизменными.

Если нейросеть – например, мозг – имеет пространственную структуру, то мультиагентная система – например, общество – существует во времени через динамику социальных взаимодействий когнитивных агентов. Поэтому иногда приходится слышать, что социальные связи существуют в сознании их участников. Но нам очевидно, что сознание – набор когнитивных компетенций – выступает необходимой предпосылкой социальной системы, которая как таковая существует в темпорально-динамическом измерении. Она представляет собой как бы производную от удачного сочетания когнитивных способностей составляющих ее агентов.

Если говорить о естественных социальных системах, то они, конечно, – продукт эволюции, а значит, благодаря некоторому механизму, преимущества, случайно полученные на уровне индивидов, закрепляются. Если это генный механизм, то отбор идет на уровне вида: индивиды с удачными мутациями выживают, с неудачными – гибнут. Если это самообучаемая нейросеть, то отбор становится гораздо эффективнее в смысле расхода биоматериала – случайно найденные «лайфхаки» закрепляются и существенно повышают выживаемость индивида. Естественно предположить, что эволюция, нащупав значительно более эффек-

тивный механизм отбора, идет по линии увеличения объема мозга и повышения удельной доли коры, ответственной за прижизненное обучение, по отношению к мезенцефалону, обеспечивающему инстинктивное поведение [Knoll 2005, 7].

Но генный и нейросетевой механизм не связаны друг с другом на уровне наследования – результаты обучения невозможно передать следующему поколению напрямую. Однако если бы это было возможно, эффективность эволюции – как количество полезных изменений по отношению к расходу биоматериала в единицу времени – возросла бы экспоненциально. Рано или поздно эволюция должна нащупать такую возможность – и она появляется в результате того, что особи научаются идентифицировать друг друга как себе подобных и получают некий механизм обратной связи, позволяющий реагировать на действия друг друга. Эту гипотезу можно рассматривать как эволюционный аргумент в пользу когнитивной основы животной социальности.

Возросшая вычислительная эффективность биологических единиц в результате первого этапа их социализации в качестве одного из своих следствий должна иметь увеличение продолжительности жизни отдельных особей, соединяя в одном временном интервале уже не два, а три поколения. И здесь приведем пример эвристической роли МАС в науке: мультиагентная модель сообщества обезьян продемонстрировала роль «бабушки» в укреплении социальных связей внутри сообщества и в дальнейшем увеличении срока жизни особей [Kim et al. 2012]. Другие мультиагентные модели поддерживают этот вывод: голландские исследователи обнаружили, что «эволюционный подход [к обучению агентов] способен поддерживать более крупные и более стабильные популяции агентов, а также более высокий уровень индивидуального успеха, по сравнению с обучением на протяжении жизни [одного агента]» [Eiben et al. 2006, 155]. Как мы понимаем, дальнейшим шагом на пути повышения вычислительной эффективности должно стать появление системы социальной аккумуляции опыта – знаковых систем и письменности.

Замечу попутно: возможно, *натуральные вычисления* (вычислительные процессы в природных системах) суть способ противостоять энтропии. Если бы появилась удовлетворительная теория, позволяющая утверждать это номотетически, возникла бы возможность увязать биологию, психологию и термодинамику в единый эволюционный проект.

6. Логические основы мультиагентности

Философия играет двоякую роль в теории МАС: с одной стороны, она получает некоторые инсайты в результате успешного развития этой технологии, а с другой – поставляет логические и семантические теории, которые успешно адаптируются теоретиками МАС.

Так, в пользу социально-коммуникативной природы знания говорит, в частности, широкое использование эпистемических модальных логик в теоретическом обосновании агент-ориентированных систем (4). Относительно простое решение предлагает Майкл Вулдридж в своем руководстве по мультиагентным системам. Он дополняет логику высказываний первого порядка одноместным модальным оператором K_i – «агент *i* знает, что…» [Wooldridge 2002, 279]. Далее он вскрывает семантические проблемы, с которыми его логика сталкивается при интерпретации ее на МАС. Например, при отсутствии абсолютно надежного обмена сообщениями внутри МАС идеал «общего знания» (*common knowledge*) может оказаться недостижимым, а процесс его достижения может превратиться в бесконечные, никогда не завершающиеся итерации. Распределенное знание – ситуация, когда, например, знания, имеющиеся у разных агентов, оказываются посылками силлогизма, следствие которого, таким образом, содержится в системе в неявном виде – также оказывается проблематичным. Для его формализации Вулдридж предлагает формализацию пересечения эпистемических миров агентов. Тогда «ограничения, накладываемые на возможные миры, в общем случае означают увеличение знания» [Wooldridge 2002, 283].

С семантикой возможных миров связано еще одно обстоятельство: она предполагает, что действующий агент должен быть «идеальным логиком», который должен совершить все возможные логические операции над имеющимися знаниями, получить все возможные выводы и увидеть все скрытые противоречия. Но реальные агенты, как искусственные, так и живые, идеальными логиками обычно не являются и потому толерантны к неявным противоречиям. Поэтому Вулдридж считает, что для МАС, в отличие, например, от научной теории, достаточно требования слабой непротиворечивости [Wooldridge 2002, 276].

Все это можно рассматривать в качестве экспликации тезиса ограниченной рациональности агентов, предложенного Макаровым и обсуждавшегося ранее в этой статье.

Похожая эпистемическая теория содержится в другом комплексном руководстве по МАС [Vlassis 2007]. Здесь предлагается определение, согласно которому, *некоторый агент знает некоторое событие, если множество всех состояний, которые из его перспективы представляются возможными, содержат это событие*. В эпистемической логике, к которой апеллирует Влассис, считается, что субъект знает некоторый факт, если этот факт содержится во всех возможных мирах, которые достижимы из перспективы этого агента. Сам же он предлагает событийный подход, согласно которому агент знает некоторое событие, если все состояния, которые агент считает возможными, содержат это событие. Поддержку своему подходу Влассис находит в популярной у других авторов (5) работе [Fagin et al. 1995], где утверждается эквивалентность логического и событийного подходов [Vlassis 2007, 39].

Поскольку принцип ограниченной рациональности агентов предполагает не только ограниченность доступного им знания, но и несовершенство, или, по крайней мере, «неклассичность» используемых ими логических средств, теоретики распределенного искусственного интеллекта, разновидностью которого являются МАС, интегрируют не только модальные, но и «нечеткие» логики, различные версии математической статистики и т.п.

7. Заключение. Система самообучаемых агентов

Мультиагентные системы, как и искусственные нейросети и как до них привычные нам компьютеры фон-неймановской архитектуры, воссоздают в готовом виде то, на что природе понадобились миллионы лет эволюции. В МАС мы имеем когнитивно ограниченных агентов, поскольку изначально носителем интеллекта является система в целом. В природных же мультиагентных системах – стаях и обществах – когнитивная ограниченность агентов объясняется эволюционной целесообразностью. Мы и другие животные – по сути биологические мобильные устройства, которые не подключены к постоянным источникам энергии. Поэтому наши процессоры и дисплеи должны быть максимально энергоэффективными, ограничиваясь самым необходимым набором когнитивных компетенций: минимально необходимый чувственный интерфейс, рассудок, стремящийся к возможно более простым объяснениям, социальная сеть, дающая возможность загружать необходимые знания и ресурсы из «облака» и беру-

щая на себя часть наиболее сложных когнитивных вычислений. В человеческом обществе существенным подспорьем являются символические системы и различные устройства аккумуляции коллективного опыта того или иного уровня технологической сложности. Я полагаю, что именно беспрецедентная вычислительная мощь нашего распределенного социального процессора, а не достоинства наших личных бортовых компьютеров, вознесла нас так высоко в царстве живой природы.

Возвращаясь к двум основным архитектурам распределенных вычислений (распределенного искусственного интеллекта) – нейросетям и мультиагентным системам, нужно сказать следующее. Главное ограничение искусственной нейронной сети состоит в ее собственной «пассивности» – она честно отрабатывает входные сигналы, осуществляя на них когнитивные функции, которым она обучена, но она сама не ищет никакие данные и не знает для чего ей все это нужно. МАС, напротив, состоят из активных агентов, преследующих определенные цели, но и эти цели, и правила взаимодействия на пути к их достижению, и сами репрезентации предметной области, на которой задаются цели и правила – все это заложено программистом, и обучение, если и происходит, то на уровне системы в целом, а не на уровне отдельных агентов. Таким образом, каждая из рассматриваемых архитектур имеет свои ограничения: одна лишена собственной активности, другая нуждается в программисте.

Если мы – пока на абстрактном уровне – представим себе мультиагентную систему, состоящую из агентов, каждый из которых имеет «на борту» искусственную нейросеть, то мы значительно продвинемся в реалистичности воспроизводства человеческой реальности. Конечно, потребуется некоторая встроенная (*embedded*) часть программного обеспечения агентов, которая будет как бы замещать простые биологические потребности – последние выступают в роли изначальных драйверов активности в живой природе, поэтому в мире искусственных устройств им нужна адекватная замена. Но правила взаимодействия будут вырабатываться самими агентами в процессе обучения своих «бортовых» ИНС. Потребуется также и некоторая система социальной обратной связи и наследования социального опыта: так, если некий агент гибнет в результате неправильно сформированных правил взаимодействия, его отрицательный опыт должен аккумулироваться и становиться обучающим фактором для других.

Моя гипотеза состоит в том, что через какое-то количество итераций такая МАС, помимо символьного языка общения, выработает две группы правил: одни будут регулировать текущие взаимодействия с целью повышения их эффективности, другие будут делать то же, но с целью сохранения и воспроизводства функциональной целостности системы.

Вторая группа правил будет моделью человеческой морали.

ПРИМЕЧАНИЯ

(1) Нужно оговорить, что в русском языке, действительно, смысл слова «вычисление» трудно отделить от человеческих математических расчетов по причине присутствия «числа» в его корне. В английском и других языках, заимствующих научную терминологию из латыни, расчет в арифметическом и схожих смыслах обозначается словами, генетически восходящими к латинскому «calculi» – «галька» – главному счетному инструменту пифагорейцев. Смысл же вычислений, интересующий нас, передается производными от латинских «com» – «вместе» и «putare» – «решать», «оценивать», «полагать». Поэтому компьютер – это не усложненный калькулятор, а нечто качественно другое.

(2) Цитата из некогда популярной песни.

(3) Виртуальная игра, где агенты-жуки ползают по поверхности, на которой неравномерно рассыпан сахар. При определенных настройках правил модель может демонстрировать возникновение сложных экономических отношений, например, обмена.

(4) О связи идей, лежащих в основе (некоторых) эпистемических логик, с социальной коммуникацией я писал в: [Михайлов 2015, 51–54, 70, 86].

(5) На нее же ссылается и Вулдридж.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Аронов и др. 2015 – *Аронов И.З., Максимова О.В., Зажигалкин А.В.* Исследование времени достижения консенсуса в работе технических комитетов по стандартизации на основе регулярных марковских цепей // Компьютерные исследования и моделирование. 2015. Т. 7. № 4. С. 941–950.

Афанасьев 2012 – *Афанасьев М.Я.* Разработка и исследование многоагентной системы для решения задач технологической подготовки производства. Автореф. дисс. … канд. тех. наук. – СПб., 2012.

Городецкий 2014 – *Городецкий В.И.* Многоагентная самоорганизация в B2B сетях // XII Всероссийское совещание по проблемам управления ВСПУ-214. Москва, 16–19 июня 2014 г. – М.: Институт проблем управления им. В.А. Трапезникова РАН, 2014. С. 8954–8966.

Макаров 2013 – *Макаров В.Л.* Искусственные общества и будущее общественных наук // Лекции и доклады членов Российской Академии наук в СПбГУП (1993–2013). В 3 т. / сост., науч. ред. А.С. Запесоцкого. Т. 1. – СПб.: СПбГУП, 2013. С. 536–551.

Михайлов 2015 – *Михайлов И.Ф.* Человек, сознание, сети. – М.: ИФ РАН, 2015.

Михайлов 2017 – *Михайлов И.Ф.* К общей онтологии когнитивных и социальных наук // Философия науки и техники. 2017. № 2. С. 103–119.

Новиков, Чхартишвили 2013 – *Новиков Д.А., Чхартишвили А.Г.* Рефлексия и управление (математические модели). – М.: Физматлит, 2013.

Редько, Сохова 2013 – *Редько В.Г., Сохова З.Б.* Многоагентная модель прозрачной рыночной экономической системы // Труды НИИСИ РАН. 2013. Т. 3. № 2. С. 61–65.

Bechtel 2008 – *Bechtel W.* Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience. – L.: Routledge, 2008.

Craver 2006 – *Craver C.F.* When Mechanistic Models Explain // Synthese. 2006. Vol. 153. No. 3. P. 355–376.

Eiben et al. 2006 – *Eiben A.E., Schut M.C., Vink N.* On the Dynamics of Communication and Cooperation in Artificial Societies // Complexus. 2004/2005. Vol. 2. No. 3–4. P. 152–162. DOI: 10.1159/000093687

Fagin et al. 1995 – *Fagin R., Halpern J., Moses Y., Vardi M.* Reasoning about Knowledge.– Cambridge, MA: MIT Press, 1995.

Glennan 2002 – *Glennan S.* Rethinking Mechanistic Explanation // Philosophy of Science. September 2002. Vol. 69. No. 3. P. 342–353.

Kempes et al. 2017 – *Kempes C.P., Wolpert D., Cohen Z., Pérez-Mercader J.* The Thermodynamic Efficiency of Computations Made in Cells Across the Range of Life // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2017. Vol. 375. No. 2109. DOI: 10.1098/rsta.2016.0343

Kim et al. 2012 – *Kim P.S., Coxworth J.E., Hawkes K.* Increased Longevity Evolves from Grandmothering // Proceedings of the Royal Society B: Biological Sciences. 2012. Vol. 279. No. 1749. P. 4880–4884. DOI: 10.1098/rspb.2012.1751

Knoll 2005 – *Knoll J.* The Brain and Its Self. A Neurochemical Concept of the Innate and Acquired Drives. – Berlin; Heidelberg: Springer-Verlag, 2005.

MacLennan 2004 – *MacLennan B.J.* Natural Computation and Non-Turing Models of Computation // Theoretical Computer Science. 2004. Vol. 317. No. 1–3. P. 115–145.

Niazi, Hussain 2013 – *Niazi M.A., Hussain A.* Cognitive Agent-based Computing-I. A Unified Framework for Modeling Complex Adaptive Systems Using Agent-based & Complex Network-based Methods. – N. Y.; L: Springer, 2013.

Singh 1994 – *Singh M.E.* Multiagent Systems. – Berlin; Heidelberg: Springer-Verlag, 1994.

Thomas 2007 – *Thomas R.W.* Cognitive Networks. Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Engineering. June 15, 2007. – Blacksburg, VA.

Turing 1936 – *Turing A.M.* On Computable Numbers, with an Application to the Entscheidungsproblem // Proceedings of the London Mathematical Society. 1936. Vol. s2–42. No. 1. P. 230–265.

Vlassis 2007 – *Vlassis N.* A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence // Synthesis Lectures on Artificial Intelligence and Machine Learning. 2007. Vol. 1. DOI: 10.2200/S00091ED1V01Y200705AIM002

Wimsatt 2002 – *Wimsatt W.C.* Functional organization, analogy, and inference // A. Ariew, R. Cummins, M. Perlman (Eds.). Functions: New Essays in the Philosophy of Psychology and Biology. – Oxford, UK: Oxford University Press, 2002. P. 173–221.

Wooldridge 2002 – *Wooldridge M.* An Introduction to Multiagent Systems. – Chichester: JohnWiley & Sons Ltd., 2002.

REFERENCES

Aronov I.Z., Maximova O.V., & Zazhigalkin A.V. (2015) Research on Time of Consensus Reaching in the Work of Technical Committees on Standardization on the Basis of Regular Markov Chains. *Komputernye Issledovaniya i Modelirovanie.* Vol. 7, no. 4, pp. 941–950 (in Russian).

Bechtel W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Craver C.F. (2006) When Mechanistic Models Explain. *Synthese*. Vol 153, no. 3, pp. 355–376.

Eiben A.E., Schut M.C., Vink N. (2006) On the Dynamics of Communication and Cooperation in Artificial Societies. *Complexus*. Vol. 2, no. 3–4, pp. 152–162. doi: 10.1159/000093687

Fagin R., Halpern J., Moses Y., & Vardi M. (1995) *Reasoning about Knowledge*. Cambridge, MA: The MIT Press.

Glennan S. (2002) Rethinking Mechanistic Explanation. *Philosophy of Science*. Vol. 69, no. 3, pp. 342–353.

Kempes C.P., Wolpert D., Cohen Z., & Pérez-Mercader J. (2017) The Thermodynamic Efficiency of Computations Made in Cells across the Range of Life. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Vol. 375, no. 2109. doi: 10.1098/rsta.2016.0343

Kim P.S., Coxworth J.E., & Hawkes K. (2012) Increased Longevity Evolves from Grandmothering. *Proceedings of the Royal Society B: Biological Sciences.* Vol. 279, no. 1749, pp. 4880–4884. doi: 10.1098/rspb.2012.1751

Knoll J. (2005) *The Brain and Its Self. A Neurochemical Concept of the Innate and Acquired Drives*. Berlin; Heidelberg: Springer-Verlag.

Makarov V.L. (2013) Artificial Societies and the Future of Social Sciences. In: A.S. Zapesotsky (Ed.). *Lectures and Reports of Russian Academy of Sciences Members in the St. Petersburg Humanitarian University of Trade Unions (1993–2013). In 3 vols* (Vol. 1, pp. 536–551). Saint Petersburg: St. Petersburg Humanitarian University of Trade Unions (in Russian).

MacLennan B.J. (2004) Natural Computation and Non-Turing Models of Computation. *Theoretical Computer Science*. Vol. 317, no. 1–3, pp. 115–145.

Mikhailov I.F. (2015) *Man, Mind, Networks*. Moscow: Institute of Philosophy, Russian Academy of Sciences (in Russian).

Mikhailov I.F. (2017) Toward the Shared Ontology of Cognitive and Social Sciences. *Filosofiya nauki i tekhniki*. 2017. No. 2, pp. 103–119.

Novikov D.A. & Chkhartishvili A.G. (2013) *Reflection and Management (Mathematical Models)*. Moscow: Fizmatlit.

Niazi M.A. & Hussain A. (2013) *Cognitive Agent-based Computing-I. A Unified Framework for Modeling Complex Adaptive Systems Using Agent-based & Complex Network-Based Methods*. New York; London: Springer.

Singh M.E. (1994) *Multiagent Systems*. Berlin; Heidelberg: Springer-Verlag.

Thomas R.W. (2007) *Cognitive Networks* (Dissertation Submitted to the Faculty of the Virginia Polytechnic Institute and State University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Computer Engineering). Blacksburg, VA.

Turing A.M. (1936*)* On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*. Vol. s2–42, no. 1, pp. 230–265.

Vlassis N. (2007) A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Vol. 1. doi: 10.2200/S00091ED1-V01Y200705AIM002

Wimsatt W.C. (2002) Functional Organization, Analogy, and Inference. In: Ariew A., Cummins R., & Perlman M. (Eds.). *Functions: New Essays in the Philosophy of Psychology and Biology* (pp. 173–221). Oxford, UK: Oxford University Press.

Wooldridge M. (2002) *An Introduction to Multiagent Systems*. Chichester: JohnWiley & Sons Ltd.