

Моделирование КОГНИТИВНОЙ ЭВОЛЮЦИИ: ВЗГЛЯД ИЗ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ВЛАДИМИР РЕДЬКО

Владимир Редько. Доктор физико-математических наук, заместитель руководителя Центра оптико-нейронных технологий Научно-исследовательского института РАН.
Адрес: 119333, Москва, ул. Вавилова, 44, к. 2.
E-mail: vgrecko@gmail.com.

Ключевые слова: моделирование когнитивной эволюции, познавательные способности животных, автономные агенты.

В статье обсуждается новое направление исследований — моделирование когнитивной эволюции, то есть эволюции познавательных способностей биологических организмов, в результате которой произошли логика, мышление, интеллект человека. Характеризуются заделы исследований когнитивной эволюции и предлагаются контуры программы будущих исследований когнитивной эволюции. Рассматриваются биологические эксперименты по когнитивным способностям животных. Обсуждаются перспективы моделирования когнитивной эволюции.

MODELING OF COGNITIVE EVOLUTION: VIEW FROM ARTIFICIAL INTELLIGENCE

VLADIMIR RED'KO. PhD in Physics and Mathematics, Deputy Director for Research of the Center of Optical Neural Technologies, Scientific Research Institute for System Analysis of the Russian Academy of Sciences.
Address: 44/2 Vavilova Street, 119333 Moscow, Russia.
E-mail: vgrecko@gmail.com.

Keywords: modeling of cognitive evolution, animal cognitive abilities, autonomous agents.

This paper discusses the modeling of cognitive evolution. Modeling of cognitive evolution is the study of evolution of animal cognitive abilities by means of mathematical and computer models. The goal of these investigations is to analyze evolutionary roots of human thinking. In this paper, the background of models of cognitive evolution are characterized and a proposal for a future program of cognitive evolution investigations is proposed. Biological experiments on cognitive features of animals are also considered. Finally, the paper discusses perspectives for cognitive evolution modeling.

ВВЕДЕНИЕ



КАК ПРОИЗОШЛО человеческое мышление? Почему наше мышление таково, каково оно есть? Почему оно применимо в научном познании? Можно ли с помощью математического и компьютерного моделирования представить эволюционные процессы происхождения человеческого мышления?

В настоящей статье обсуждается новое научное направление, которое связано с этими интригующими вопросами. Речь пойдет о моделировании когнитивной эволюции, эволюции познавательных способностей биологических организмов, той эволюции, в процессе которой произошло наше логическое мышление.

Моделирование когнитивной эволюции — это междисциплинарное направление исследований, связанное с основами математики, теорией познания, когнитивной наукой, биологическими исследованиями, работами в области вычислительных наук.

И хотя речь пойдет о серьезных научных исследованиях, эта статья научно-популярная. В ней будут приведены ссылки на близкие по теме научные работы, но для понимания статьи читать эти работы не обязательно. Однако ссылки на работы могут быть полезны тем читателям, которые захотят поглубже разобраться в обсуждаемых здесь проблемах, а возможно, и подключиться к конкретным научным исследованиям.

Структура статьи следующая. Статья начинается с краткого воспоминания автора о том, как возникли идеи моделирования когнитивной эволюции, затем будут изложены точки зрения философов, которые думали над близкими проблемами. Далее будут кратко охарактеризованы направления исследований, в которых уже имеются заделы (подходы, концепции и модели) моделирования когнитивной эволюции. Потом мы рас-

смотрим интересные биологические эксперименты последних лет по элементарному мышлению животных. И наконец, остановимся на начальных моделях когнитивной эволюции, на связанных с этими моделями прикладных направлениях и на перспективах дальнейших исследований.

УРОК ЭЙНШТЕЙНА

Начну с краткого воспоминания. Когда я учился на втором курсе МФТИ, мне попала в руки хорошо написанная биография Альберта Эйнштейна. Из этой книги я неожиданно для себя узнал, что Эйнштейн не только занимался физикой и математикой, но и серьезно интересовался философией. Более того, по словам А. Эйнштейна, изучение философии способствовало его научной работе. Особенно ему помогли работы Д. Юма и Э. Маха. Этот интерес Эйнштейна к философии заинтриговал меня, и я решил почитать трактаты тех философов, которых изучал Эйнштейн.

Это было во второй половине 1960-х годов, в самый «расцвет застоя». К философии у нас, студентов-физиков, было скептическое отношение как к неестественной науке, которую нас заставляли изучать в принудительном порядке. Высшее образование подразумевало обязательное изучение довольно ограниченного курса марксистско-ленинской философии. Тем не менее в те «застойные» времена в книжных магазинах наряду с серьезной и обширной физико-математической литературой можно было купить собрания сочинений Платона, Аристотеля, Б. Спинозы, Д. Юма, И. Канта — философская классика тогда издавалась достаточно хорошо. Так что свою любознательность — что же полезного мог вынести А. Эйнштейн из трудов философов — я вполне мог попытаться удовлетворить.

Начал я с трудов нидерландского философа XVII века Бенедикта Спинозы — этого философа тоже читал Эйнштейн, и я, подражая ему, решил познакомиться с трактатами Спинозы. Сначала меня удивило, что серьезная работа мысли возможна не только в естественных науках, но и в философии — при изучении принудительного курса общественных наук этого как-то не чувствовалось, а при прикосновении к классике философии это сразу бросилось в глаза. Но самое интересное произошло дальше — однажды при чтении «Краткого трактата о боге, человеке и его счастье» Б. Спинозы я задумался над вопросом: «А почему, собственно говоря, математика применима к физике?»

Действительно, физик, изучая природные явления, использует доказанные математиками результаты (теоремы, мето-

ды решения уравнений и т. п.). Математики же обосновывают эти результаты чисто логическим путем, совершенно независимым от реальной природы. Так почему же результаты математиков применимы к природным явлениям? В общем виде проблему можно поставить так: *почему наше человеческое мышление применимо к познанию природы?* В особенности это важно для научного познания: создавая величественную научную картину мира, ученые неизбежно используют свою логику, свое мышление.

Студентом я даже попытался построить свою собственную философскую концепцию, пытаюсь чисто философски обосновать применимость нашего мышления, нашей логики к научному познанию. Ключевая идея моей философской концепции была «Человек воспринимает природу, видит, какова она есть, и принимает, что она логична». Позднее я узнал, что близкие идеи высказывал известный философ Э. Гуссерль. Но тогда я просто отступил от философии и пошел по более надежному и более традиционному научному пути: стал работать в области физики магнитных явлений под руководством ученого с хорошим педагогическим стилем — А. К. Звездина. И уже после защиты кандидатской диссертации, набравшись научного опыта и почувствовав способность к самостоятельной научной работе, я стал вновь задумываться над проблемой применимости нашего мышления к научному познанию. И как сейчас понимаю, есть очень интересный подход к решению этой проблемы, причем не только на чисто философской основе, а и на более надежной, естественно-научной основе.

Итак, в чем проблема?

Существует наука. Наука — это познание природы. Но способен ли человек познавать законы природы?

Еще раз обратимся к физике, одной из наиболее фундаментальных естественно-научных дисциплин. Мощь физики связана с эффективным применением математики. Но математик строит свои теории совсем независимо от внешнего мира, используя свое мышление (в тиши кабинета, лежа на диване, в изолированной камере...). Почему же результаты, получаемые математиком, применимы к реальной природе?

Для того чтобы как следует разобраться в этом вопросе, было бы очень интересно попытаться строить модели эволюционного происхождения мышления. По-видимому, наиболее четкий путь такого исследования — построение математических и компьютерных моделей «интеллектуальных изобретений» биологической эволюции. Целесообразно с помощью моделей представить общую картину эволюции когнитивных способностей животных и эволюционного происхождения ин-

теллекта человека. Но, прежде чем обсуждать моделирование когнитивной эволюции, обратимся к истории теории познания и обсудим точки зрения философов, думавших над близкими проблемами. Проследим цепочку: Д. Юм → И. Кант → К. Лоренц.

ИЗ ИСТОРИИ ТЕОРИИ ПОЗНАНИЯ

В «Исследовании о человеческом познании» (1748) Давид Юм подверг сомнению понятие причинной связи¹. А именно он задался вопросом: почему когда мы видим, что за одним явлением (*A*) постоянно следует другое (*B*), то мы приходим к выводу, что *A* является причиной *B*? Например, когда мы наблюдаем, что Солнце освещает камень и камень нагревается, то мы говорим, что солнечный свет есть причина нагревания камня.

Фактически Юм задался вопросом: что заставляет нас делать выводы о происходящих в природе явлениях? Что лежит в основе этих выводов? Юм попытался понять, откуда мы берем основание заключать, что *A* есть причина *B*. Он посмотрел на этот вопрос, как он пишет, со всех сторон и не нашел никакого другого основания, кроме некоторого внутреннего чувства привычки. То есть имеется какое-то наше внутреннее свойство, которое заставляет нас утверждать, что если за *A* постоянно следует *B*, то *A* есть причина *B*. И это внутреннее чувство заставляет нас после того, когда мы сделали такое умозаключение и снова видим событие *A*, ожидать, что за *A* вновь последует и событие *B*.

Юм взглянул на наш познавательный процесс со стороны, извне. Он как бы вышел на некий метауровень рассмотрения наших собственных познавательных процессов и задался вопросом о том, откуда взялись эти познавательные процессы и почему они работают.

Острота сомнений Юма была в том, что он задался вопросом о принципиальной способности человека познавать мир.

Остроту сомнений Юма очень хорошо почувствовал Иммануил Кант. Но Кант также видел мощь и силу современной ему науки. Тогда уже была глубокая, серьезная и развитая математика, мощная ньютоновская физика, дающая *картину мира*, позволяющая объяснить множество явлений на основе немногих четких предположений, использующая многозвенные и сильные математические дедуктивные выводы. И что было делать Канту? Подвергнуть сомнению все эти познавательные процессы?

1. См.: Юм Д. Исследование о человеческом познании // Юм Д. Соч.: В 2 т. М.: Мысль, 1966. Т. 2. С. 5–169.

И, развивая сомнения Юма дальше, отвергнуть всю науку? Ведь на самом деле — драма!

Конечно же, Кант, как научно образованный человек, не стал отвергать современную ему науку, а постарался разобраться, как же работают познавательные процессы. В результате появились знаменитая «Критика чистого разума»² и популярное изложение «Критики...» — «Пролегомены ко всякой будущей метафизике, могущей появиться, как наука»³. Кант провел исследование познавательных процессов в определенном приближении — приближении фиксированного мышления взрослого человека. Он не задавался вопросом, *откуда* берутся познавательные способности, он констатировал факт, что они существуют, и исследовал, *как* они работают. В результате этого анализа Кант пришел к выводу, что существует система категорий, концепций, логических правил и методов вывода (таких как заключения относительно причинных связей между событиями), которые используются в познании природы. Эта система «чистого разума» имеет априорный характер — она существует в нашем сознании прежде всякого опыта — и является основой научного познания природы.

Естественно, что приближение фиксированного мышления человека наложило свой отпечаток: Кант утверждает — и в рамках этого приближения вполне логично(!), — что так как «чистый разум» априорен, то наш рассудок в познавательном процессе предписывает свои законы природе:

...хотя вначале это звучит странно, но тем не менее верно, если я скажу: *рассудок не черпает свои законы (a priori) из природы, а предписывает их ей*⁴.

Наверное, во времена Канта было разумно ограничиться приближением фиксированного мышления взрослого человека — все сразу не охватишь. Кроме того, не было еще теории Чарльза Дарвина. По-видимому, если бы Кант знал теорию происхождения видов, то он явно бы задумался об эволюционном происхождении «чистого разума». Тем более что эволюционные идеи ему были не чужды — можно вспомнить его теорию происхождения Солнечной системы. Но приближение фиксированного мышления взрослого человека накладывает свои ограничения — оно не позволяет ответить на вопросы: откуда же взялись

2. См.: Кант И. Критика чистого разума // Кант И. Соч.: В 6 т. М.: Мысль, 1964. Т. 3. С. 69–695.

3. См.: Он же. Пролегомены ко всякой будущей метафизике, могущей появиться как наука // Кант И. Соч.: В 6 т. М.: Мысль, 1965. Т. 4. Ч. 1. С. 67–210.

4. См.: Там же.

познавательные способности? познаем ли мы истинные законы природы или наш рассудок «предписывает их ей»? Фактически Кант ушел от наиболее острой части вопроса, поставленного Юмом, — он не задавался вопросом, откуда взялся «чистый разум», а только тщательно и детально исследовал свойства «чистого разума» и применение его в научном познании.

Естественно, что после появления теории происхождения видов Дарвина должна была произойти ревизия концепции априорного «чистого разума». И она произошла. Очень четко ее выразил Конрад Лоренц в знаменитой статье «Кантовская доктрина априорного в свете современной биологии»⁵. Согласно Лоренцу, кантовские априорные категории и другие формы «чистого разума» произошли в результате естественного отбора:

...наши категории и формы восприятия, зафиксированные до индивидуального опыта, адаптированы к внешнему миру в точности по тем же причинам, по которым копыто лошади адаптировано к степному грунту еще до того, как лошадь рождается, а плавник рыбы — к воде до ее появления из икринки.

Составляющие «чистого разума» возникали в процессе эволюции, в результате многочисленных взаимодействий с внешним миром. В эволюционном контексте «чистый разум» совсем не априорен, а имеет явные эволюционные эмпирические корни.

Фактически И. Кант и К. Лоренц показали, что если не рассматривать эволюционное происхождение методов познания, то нет ответа на ключевой вопрос о применимости логического мышления человека к познанию природы.

МОЖНО ЛИ В ПРИНЦИПЕ ВЕСТИ ИССЛЕДОВАНИЯ КОГНИТИВНОЙ ЭВОЛЮЦИИ?

Как же разобраться в происхождении логических форм мышления? Можно ли промоделировать это происхождение? Какие методы использовать? Как такие исследования связаны с основами науки, основаниями математики? Есть ли задел работ по моделированию эволюции познавательных способностей животных? Далее обсуждаются перечисленные вопросы. Но прежде всего кратко остановимся на вопросе: можно ли вообще в принципе вести такие исследования?

5. См.: Лоренц К. Кантовская концепция *a priori* в свете современной биологии // Эволюция. Язык. Познание. М.: Языки русской культуры, 2000. С. 15–41.

Итак, можно ли проследить эволюционные корни логических правил, используемых в дедуктивных доказательствах? По мнению автора настоящей статьи, да, можно.

Приведем следующую аналогию. Одно из элементарных правил, которое использует математик в логических заключениях, — правило *modus ponens*: «Если имеет место A и из A следует B , то имеет место B », или:

$$\frac{A, A \rightarrow B}{B}.$$

Перейдем от математика к собаке, у которой вырабатывают классический условный рефлекс. При выработке рефлекса в памяти собаки формируется связь: «За УС должен последовать БС» (УС – условный стимул, БС — безусловный стимул). Когда после выработки рефлекса собаке предъявляют УС, то она, помня о хранящейся в ее памяти «записи» УС → БС, делает элементарный «вывод»:

$$\frac{\text{УС, УС} \rightarrow \text{БС}}{\text{БС}}.$$

И собака ожидает БС. Например, в опытах И. П. Павлова после выработки условного рефлекса собака после сигнала звонка (УС) ожидает появления пищи (БС), и у собаки происходит слюноотделение.

Конечно, чисто дедуктивное применение формального правила *modus ponens* математиком и основанный на обобщении опыта индуктивный «вывод», который делает собака, явно различаются. Тем не менее и в первом, и во втором случаях речь идет о следственной связи между математическими утверждениями либо событиями: из A следует B , за УС следует БС.

Итак, применение правила *modus ponens* при дедуктивном выводе аналогично «выводу» на основе классического условного рефлекса. Указанная аналогия позволяет задуматься об эволюционных корнях логических правил, используемых в математике.

Таким образом, можно анализировать эволюционные корни логического мышления и строить модели эволюционного происхождения логических правил, используемых в научном познании. Естественно, что при моделировании когнитивной эволюции целесообразно рассматривать использование познавательных способностей животных при их приспособлении к внешней среде. Познавательные свойства были полезны для животных и закреплялись в процессе естественного отбора.

При этом результат эволюции — правила логического вывода, используемые при математических доказательствах, — известны

и достаточно хорошо формализованы⁶. В основе этих выводов — элементарные правила, такие как *modus ponens*.

Как же конкретно вести моделирование когнитивной эволюции? Как уже отмечалось, было бы целесообразно с помощью математических и компьютерных моделей представить общую картину эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека.

Отметим, что, действуя упрощенно, мы, конечно, можем представить когнитивную эволюцию следующим образом. В процессе биологической эволюции животные приобретали способности познавать внешний мир. Эти познавательные способности помогали им приспосабливаться к окружающей среде, и приспособленности тех животных, которые приобретали «хорошие» познавательные способности, увеличивались. Животные с повышенными приспособленностями распространялись в популяциях, следовательно, «хорошие» познавательные способности фиксировались эволюцией. Шаг за шагом, эволюционно возникали все более сложные и более интеллектуальные способности, что в конечном итоге привело к появлению человеческого мышления, логики человеческого мышления. Но этих простых рассуждений явно недостаточно. Проблема принципиальной способности человека познавать природу — фундаментальная проблема, связанная с основаниями науки, и она должна быть проанализирована настолько глубоко, насколько это возможно.

Есть ли задел моделирования когнитивной эволюции? Оказывается, что да, такой задел есть. В настоящее время в нескольких научных направлениях ведутся исследования, которые могут быть использованы при моделировании когнитивной эволюции. Эти направления таковы: адаптивное поведение, искусственная жизнь, когнитивные архитектуры, научные основы искусственного интеллекта. В основном это работы по основанным на биологических аналогиях математическим и компьютерным моделям автономных агентов с когнитивными свойствами. Кратко охарактеризуем эти направления.

ЗАДЕЛЫ ИССЛЕДОВАНИЙ КОГНИТИВНОЙ ЭВОЛЮЦИИ

Адаптивное поведение. Направление исследований «Адаптивное поведение» сформировалось в начале 1990-х годов⁷. Основ-

6. См.: Математическая теория логического вывода / Под ред. А. В. Идельсона, Г. Е. Минца. М.: Наука, 1967.

7. См.: From Animals to Animats. Proceedings of the First International Conference

ной подход этих исследований — конструирование и изучение искусственных «организмов» (в виде компьютерной программы или робота), способных приспосабливаться к переменной внешней среде. Исследователи адаптивного поведения разрабатывают такие модели, которые применимы к описанию поведения как реального животного, так и искусственного модельного организма. Дальняя цель этих работ — анализ эволюции когнитивных способностей животных в контексте происхождения интеллекта человека — близка к задаче моделирования когнитивной эволюции. Работы отечественных исследователей адаптивного поведения представлены в сборнике⁸.

Хотя «официально» направление «Адаптивное поведение» сформировалось в начале 1990-х годов, были явные провозвестники этого направления. Приведем примеры из истории отечественной науки. В 1960-х годах блестящий кибернетик и математик М. Л. Цетлин предложил и исследовал модели автоматов, способных адаптивно приспосабливаться к окружающей среде. Работы М. Л. Цетлина инициировали целое научное направление, получившее название «Коллективное поведение автоматов»⁹. А в 1960–1970-х годах под руководством талантливого кибернетика М. М. Бонгарда был предложен весьма нетривиальный проект «Животное», характеризующий целенаправленное адаптивное поведение искусственных организмов¹⁰.

Искусственная жизнь. Близким направлением к «Адаптивному поведению» является интересное направление исследований «Искусственная жизнь», которое сформировалось в конце 1980-х годов¹¹. Основной мотивацией исследований искусственной жизни служит желание понять и промоделировать формальные принципы организации биологической жизни. Как сказал руководитель первой международной конференции по искусственной жизни К. Лангтон, «основное предположение искусственной жизни состоит в том, что „логическая фор-

on Simulation of Adaptive Behavior / J.-A. Meyer, S. W. Wilson (eds). L.; Cambridge, MA: MIT Press, 1990.

8. См.: От моделей поведения к искусственному интеллекту / Под ред. В. Г. Редько. М.: УРСС, 2006.

9. См.: Цетлин М. Л. Исследования по теории автоматов и моделирование биологических систем. М.: Наука, 1969.

10. См.: Бонгард М. М., Лосев И. С., Смирнов М. С. Проект модели организации поведения — «Животное» // Моделирование обучения и поведения. М.: Наука, 1975. С. 152–171. Хороший обзор ранних работ по адаптивному поведению представлен в книге: Гаазе-Рапопорт М. Г., Поспелов Д. А. От амебы до робота: модели поведения. М.: УРСС, 2011.

11. См.: Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems / C. G. Langton (ed.). Redwood City, CA: Addison-Wesley, 1989.

ма“ организма может быть отделена от материальной основы его конструкции».

Сторонники направления «Искусственная жизнь» часто считают, что они исследуют более общие формы жизни, чем те, которые существуют на Земле. То есть изучается жизнь, какой она могла бы в принципе быть (*life-as-it-could-be*), а не обязательно та жизнь, какой мы ее знаем (*life-as-we-know-it*). Искусственная жизнь — это синтетическая биология, которая по аналогии с синтетической химией пытается воспроизвести биологическое поведение в различных средах. Это жизнь, созданная человеком, а не природой (*life made by Man rather than by Nature*). Исследования искусственной жизни направлены не только на теоретические исследования свойств жизни, но и (аналогично синтетической химии) на практические приложения, такие как подвижные роботы, медицина, нанотехнология, исследования «жизни» социальных и экономических систем. Обзор ранних работ в этом направлении содержится в книге В. Г. Редько¹².

Когнитивные архитектуры. Под когнитивными архитектурами понимаются структуры и принципы функционирования познающих систем, которые можно использовать в искусственном интеллекте¹³. В последние годы большое внимание уделяется биологически инспирированным когнитивным архитектурам¹⁴.

Автономные агенты. Это близкое к когнитивным архитектурам направление исследований, в котором большое внимание уделяется основанным на биологических аналогиях автономным агентам и компьютерным моделям агентов, обладающих когнитивными и некоторыми интеллектуальными свойствами¹⁵.

Необходимо подчеркнуть, что автономные агенты вполне могут рассматриваться как объединяющее понятие для указанных направлений. Работы по автономным агентам ведутся как со стороны биологических наук (агенты вполне естественно могут моделировать биологические организмы), так и со стороны вычис-

12. Редько В. Г. Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики. М.: УРСС, 2005.

13. Обзор исследований в этом направлении см. в: Langley P., Laird J. E., Rogers S. Cognitive architectures: Research issues and challenges // Cognitive Systems Research. 2009. Vol. 10. № 2. P. 141–160.

14. См.: Biologically Inspired Cognitive Architectures 2011. Proceedings of Second Annual Meeting of the BICA Society / A. V. Samsonovich, K. R. Johansdottir (eds). Amsterdam; Berlin; Tokyo; Washington, DC: IOS Press, 2011.

15. Обзор исследований по автономным агентам см. в: Vernon D., Metta G., Sandini G. A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents // IEEE Transactions on Evolutionary Computation. Special issue on Autonomous Mental Development. 2007. Vol. 11. № 2. P. 151–180.

лительных наук (можно изучать модели и искусственные аналоги организмов, например, роботы). Вполне естественно изучать познавательные свойства и знания автономных когнитивных агентов. Также разумно исследовать эволюцию популяций агентов и коммуникации, обмен информацией между агентами.

ПРИМЕРЫ МОДЕЛЕЙ АВТОНОМНЫХ КОГНИТИВНЫХ АГЕНТОВ

Приведем два примера моделей автономных агентов, обладающих когнитивными свойствами.

Нейробиологическая модель адаптивного поведения. Разработана в Институте нейронаук¹⁶, руководимом лауреатом Нобелевской премии Дж. Эдельманом, где более 25 лет ведутся разработки нескольких поколений моделей работы мозга (*Darwin I*, *Darwin II*...). Основное внимание при этом уделяется попыткам понимания работы мозга путем построения его моделей.

Иллюстрируя поход, развиваемый в Институте нейронаук, отметим интересный доклад на семинаре по нейронным сетям, который несколько лет назад делал Е. М. Ижикевич, работавший в этом институте. Он рассказывал, что над пониманием мозга работает группа исследователей, которые стараются как можно точнее представить, как именно функционируют структуры мозга. А если исследователи чувствуют, что они «зацикливаются», перестают конструктивно обдумывать нервную систему, то они идут на корт, играют в теннис, а потом снова возвращаются и продолжают свою работу.

Приведем пример конкретной нейробиологической модели, разработанной в Институте нейронаук. Речь пойдет о роботоподобном устройстве «Дарвин X», система управления которого построена на базе моделей мозга. Иначе это устройство называется *NOMAD* (*Neurally Organized Mobile Adaptive Device*), причем, хотя это устройство подобно роботу, авторы называют его не роботом, а прибором, устройством (*device*). Часто также используется термин *Brain Based Device* (устройство, основанное на понимании работы мозга).

Были хорошо продуманы принципы моделирования *NOMAD*. Эти принципы состоят в следующем:

- 1) устройство помещается в реальную физическую среду;
- 2) имеется некоторая поведенческая задача, которую должно решать устройство;

16. См. URL: <http://www.nsi.edu>.

- 3) поведение устройства контролируется модельной нервной системой, которая отражает архитектуру мозга и динамику процессов в мозге;
- 4) поведение устройства и процессы в модельной нервной системе должны допускать сравнение с экспериментальными биологическими данными.

В одной из работ по *NOMAD*¹⁷ было промоделировано поведение мыши в лабиринте Морриса.

Исследования поведения мыши или крысы в лабиринте Морриса — один из канонических биологических экспериментов, который состоит в следующем. Имеется бассейн с непрозрачной жидкостью (например, это может быть вода, подкрашенная молоком), на бортах бассейна есть разные рисунки, которые мышь видит и может использовать для ориентировки. В определенном месте бассейна есть скрытая от зрения платформа (находящаяся неглубоко под водой), которую мышь может найти и тем самым спастись — не утонуть. Мышь бросают в бассейн, она плавает некоторое время и либо находит платформу и спасается, либо начинает тонуть (тогда ее спасает экспериментатор). После ряда экспериментов мышь начинает использовать рисунки на бортах бассейна и, ориентируясь по рисункам, находит платформу за достаточно короткое время.

Поведение *NOMAD* в лабиринте Морриса моделировалась следующим образом. *NOMAD* представлял собой подвижное устройство на колесах, управляемое нейронной сетью, состоящей из 90 тысяч нейронов, в которой было выделено 50 различных нейронных областей, в частности были выделены несколько областей гиппокампа. В сети было $1,4 \cdot 10^6$ синаптических контактов между нейронами. Программно нейронная сеть была реализована на основе компьютерного кластера. При моделировании детально исследовались процессы, происходящие в разных нейронных областях.

Сенсорная система *NOMAD* включала зрение, обонятельную систему, позволяющую отслеживать свои собственные следы, систему инфракрасных приемников-излучателей, обеспечивающую избегание столкновений, и специальный детектор скрытой от зрения платформы, позволяющий обнаруживать эту платформу только тогда, когда *NOMAD* находился непосредственно над ней.

NOMAD помещался в комнату, в которой была скрытая платформа; на стенах комнаты были разноцветные полосы — ори-

17. См.: *Krichmar J. L., Seth A. K., Nitz D. A., et al. Spatial navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions // Neuroinformatics. 2005. Vol. 3. № 3. P. 197–221.*

ентиры. В начале каждого из компьютерных экспериментов *NOMAD* помещался в разные участки комнаты, задача *NOMAD* была найти скрытую платформу. Обучение нейронных сетей *NOMAD* осуществлялось по модифицированному правилу Хебба (увеличение или уменьшение веса синаптической связи между активными нейронами) на основе поощрений (получаемых при нахождении скрытой платформы) и наказаний (получаемых при приближении к стенам комнаты).

Было продемонстрировано, что:

- 1) *NOMAD* самостоятельно обучается находить платформу, ориентируясь по полосам на стенах комнаты, причем обучается достаточно быстро (за 10–20 попыток);
- 2) в модельном гиппокампе формируются «нейроны места», активные только тогда, когда *NOMAD* находится в определенных участках комнаты;
- 3) в модельном гиппокампе формируются связи между отдельными нейронными областями, отражающие причинно-следственные зависимости.

Итак, изложенная модель представляет собой интересное компьютерное исследование самообучающегося адаптивного устройства, хорошо продуманное с биологической точки зрения и дающее понимание функционирования определенных структур мозга.

Бионическая модель поискового адаптивного поведения. Одно из актуальных направлений исследований в рамках моделирования адаптивного поведения — имитация поискового поведения животных. В нашей работе¹⁸ исследовано поисковое поведение на примере личинок ручейников *Chaetopteryx villosa*, обитающих на дне водоемов. Личинки носят на себе «домик» — трубку из песчинок и других частиц, собираемых со дна водоемов. Частицы скрепляются между собой по краям с помощью клейкой белковой нити. Строительство домика требует меньше времени, усилий и белка, если личинки используют относительно крупные и плоские частицы. Однако поиск крупных частиц на дне водоема требует затрат времени и энергии, неизвестных личинке заранее. Задача осложняется еще и тем, что личинки при поиске частиц не пользуются зрением и могут обнаружить частицу и определить ее размер только на ощупь, что требует дополнительных затрат времени.

18. См.: Непомнящих В. А., Попов Е. Е., Редько В. Г. Бионическая модель адаптивного поискового поведения // Известия РАН. Теория и системы управления. 2008. № 1. С. 85–93.

В нашей работе была построена компьютерная модель поискового поведения личинок ручейников, строящих чехол-домик из частиц разного размера и ведущих поиск скоплений подходящих частиц. Модель сопоставлялась с результатами биологического эксперимента, в котором личинки строили чехол-домик из крупных и мелких частиц.

При этом удачная модель была построена не сразу. Сначала мы попробовали построить модель так, чтобы модельные личинки максимизировали скорость роста площади домика. Такая модель работала, но она не соответствовала биологическому эксперименту. В биологическом эксперименте личинки строили домик в основном из крупных частиц (экономя усилия и склеивающую частицы нить), а в модели домик строился преимущественно из мелких частиц.

После этого мы учли, что поисковое поведение личинок можно представить как чередование двух тактик: 1) сбор и прикрепление частиц на одном и том же участке; 2) поиск нового участка с подходящими частицами. Переключение между поведенческими тактиками инерционно: личинка не сразу покидает участок, на котором она нашла крупные частицы, а затем встречала только мелкие, так как на этом участке крупные частицы могут еще встретиться, и не всегда сразу начинает сбор частиц, если встретилась только одна крупная частица, так как эта частица могла встретиться случайно.

Основываясь на понимании такого инерционного переключения между тактиками поведения, мы ввели и использовали понятие мотивации M к прикреплению частиц. И считали, что если мотивация M была достаточно велика, то происходил сбор и прикрепление частиц к домику; если M была мала, то модельная личинка искала новое место с подходящими размерами частиц. Динамика регулирующей поведение мотивации M учитывала инерцию переключения между тактиками поведения, случайные вариации и направленное изменение мотивации. Направленное изменение M состояло в следующем: если модельная личинка тестировала крупную частицу после мелкой, то мотивация M к прикреплению следующих частиц повышалась, если, наоборот, тестировалась мелкая частица после крупной, то M понижалась. Важно, что в модель вводились знания личинки о размерах последних протестированных и прикрепленных частиц и эти знания использовались личинками.

Новая модель, основанная на понятии мотивации M к прикреплению частиц и на знаниях о размерах последних протестированных и прикрепленных частиц, оказалась удачной. Она хорошо согласовывалась с биологическими экспериментальными

ми данными: как в эксперименте, так и в модели к домику преимущественно прикреплялись крупные частицы, среднее число прикрепленных частиц практически совпадало у реальных и модельных личинок.

Когнитивные свойства личинок в нашей модели были простыми: личинки использовали знания о размерах протестированных и прикрепленных частиц. Но тем не менее эти когнитивные свойства, хоть и простые, были весьма существенными.

В этом разделе были очерчены два примера моделей, близких к моделированию когнитивной эволюции. В следующих двух разделах мы немного отвлечемся от конкретных моделей. Для исследований когнитивной эволюции полезны не только конкретные модели, но и концептуальные подходы, а также интересные биологические эксперименты, демонстрирующие нетривиальные познавательные свойства живых организмов. Начнем с примера концептуального подхода, который был предложен талантливым физиком и кибернетиком В. Ф. Турчиным.

ТЕОРИЯ МЕТАСИСТЕМНЫХ ПЕРЕХОДОВ В. Ф. ТУРЧИНА

Яркий пример, иллюстрирующий разработку эволюционных концепций, — книга В. Ф. Турчина «Феномен науки. Кибернетический подход к эволюции»¹⁹. Книга была написана в нашей стране в 1970 году, однако из-за политической деятельности ее автора впервые была издана только в США в 1977 году.

Турчин рассматривает биологическую эволюцию с кибернетической точки зрения, а эволюцию научного познания — как продолжение биокибернетической эволюции. В книге последовательно проанализированы ступени биологической эволюции, а также этапы возникновения и развития математического знания.

В качестве основы исследования Турчин использует предложенную им концептуальную теорию метасистемных переходов.

Кратко суть *теории метасистемных переходов* сводится к следующему: переход от нижних уровней системной иерархии к верхним происходит путем метасистемных переходов. Каждый метасистемный переход можно рассматривать как объединение ряда подсистем S_i нижнего уровня и появление дополнительного механизма *управления* S объединенными под-

19. См.: Турчин В. Ф. Феномен науки. Кибернетический подход к эволюции. М.: ЭТС, 2000. URL: <http://www.refal.ru/turchin/phenomenon/>.



РИС. 1. Схема метасистемного перехода.

системами. В результате метасистемного перехода формируется система S' нового уровня, которая может быть включена как подсистема в следующий метасистемный переход (рис. 1).

Турчин уделяет особое внимание количественному накоплению «потенциала развития» в подсистемах S_i перед метасистемным переходом на качественно новый уровень иерархии, а также процессу размножения и развития подсистем предпоследнего уровня иерархии после метасистемного перехода.

Турчин характеризует биологическую эволюцию следующими метасистемными переходами:

- управление положением = движение;
- управление движением = раздражимость (простой рефлекс);
- управление раздражимостью = (сложный) рефлекс;
- управление рефлексамии = ассоциации (условный рефлекс);
- управление ассоциациями = человеческое мышление;
- управление человеческим мышлением = культура.

Подчеркнем, что теория метасистемных переходов — это именно концептуальная теория. Ее главное достоинство состоит в том, что она позволяет взглянуть на эволюционные процессы сверху и проинтерпретировать на качественном уровне ряд ступеней кибернетической эволюции.

Закljučая обсуждение этой теории, отметим два важных метасистемных перехода, которые существенны для анализа когнитивной эволюции, анализа происхождения интеллекта: 1) переход от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов; 2) переход от первобытного мышления к критическому.

1. Переход от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов можно рассматривать как появление в «сознании» животного свойства «понятие». Обобщенные об-

разы можно представить как мысленные аналоги наших слов, не произносимых животными, но реально используемых ими.

2. Переход от первобытного мышления к критическому (от мышления первобытных племен к возникновению формального логического мышления). Критическое мышление отличается от первобытного тем, что возникает оценка мыслительного процесса самим мыслящим субъектом. «Критическое мышление рассматривает каждое объяснение (языковую модель действительности) наряду с другими, конкурирующими объяснениями (моделями), и оно не удовлетворится, пока не будет показано, чем данное объяснение лучше, чем конкурирующее». Здесь уместно упомянуть «Феномен человека» П. Тейяра де Шардена²⁰, где этот переход называется возникновением рефлексии.

Первый переход можно рассматривать как возникновение простых когнитивных способностей, а второй — как возникновение высших форм когнитивной деятельности.

Далее перейдем к биологическим экспериментам, демонстрирующим нетривиальные познавательные способности живых организмов, иногда эти способности называют элементарным мышлением животных.

БИОЛОГИЧЕСКИЕ ЭКСПЕРИМЕНТЫ ПО ЭЛЕМЕНТАРНОМУ МЫШЛЕНИЮ ЖИВОТНЫХ

В последние годы проведены интересные биологические исследования, показывающие, что элементарные формы мышления присущи не только высшим животным, но и достаточно простым, например насекомым (пчелам, муравьям), а также врановым птицам, которые способны самостоятельно создавать и целесообразно использовать «орудия труда». В данном разделе кратко характеризуются биологические экспериментальные данные и некоторое связанное с ними моделирование.

Пчелы сообщают другим пчелам, где искать медоносные цветы. Еще в 1920-е годы Карл фон Фриш начал исследования способа передачи пчелами информации друг другу. К. фон Фриш обратил внимание на то, что возвращающиеся в улей пчелы-сборщицы совершают танец, привлекающий внимание других

20. Тейяр де Шарден П. Феномен человека. М.: Устойчивый мир, 2001.

пчел. Угол, составленный между осью танца и вертикалью, соответствует углу между направлением на пищу и направлением на солнце. Причем этот угол для танцующей пчелы меняется в соответствии с движением солнца. Скорость танца соответствует расстоянию между пищей и ульем. Танцующая пчела также пахнет, ее запах характеризует медоносный цветок. Другие пчелы наблюдают этот танец и затем находят те цветы, с которых прилетела танцующая пчела (на расстоянии до нескольких километров). Долгое время результаты К. фон Фриша вызывали споры, которые продолжались и после того, как в 1973 году К. фон Фриш получил за свое открытие Нобелевскую премию²¹.

Интересно, что в 1990-е годы Б. Андерсен и А. Михельсен создали искусственную пчелу-робота, которая также танцевала и с помощью танца передавала информацию живым пчелам. Размер робота был близок к размеру пчелы. И живые пчелы воспринимали этого робота вполне естественно. Живые пчелы летели из улья на поляну, руководствуясь указаниями пчелы-робота²².

Муравьи могут передавать информацию со скоростью около 1 бита в минуту и запоминать небольшие целые числа. В семье рыжих лесных муравьев можно выделить рабочие группы, состоящие из одного муравья-разведчика и 3–8 муравьев-фуражиров. Каждый разведчик, найдя пищу, вступает в контакт со своей группой и передает фуражирам информацию, где нужно искать пищу. В экспериментах и расчетах Ж. И. Резниковой и Б. Я. Рябко исследовалось поведение муравьев с помощью специально изготовленных ветвящихся лабиринтов небольшой глубины: каждый путь разветвлялся на два, было до 5–6 развилок в лабиринте. Было установлено, что муравьи-разведчики могли запомнить путь к кормушке, а затем путем контактов с фуражирами передавать им информацию о том, в каком месте лабиринта находится пища²³. Зная глубину лабиринта

21. Недавно на русском языке опубликована книга «Анатомия научного противостояния. Есть ли язык у пчел?» (М.: Языки славянских культур, 2011) А. Веннера и П. Уэллса, опровергших гипотезу К. Фриша и подробнейшим образом проанализировавших эту полемику в более широком контексте проблем истории и методологии науки. — *Прим. ред.*

22. Подробнее об этой передаче символической информации см. обзор: Резникова Ж. И. Современные подходы к изучению языкового поведения животных // Разумное поведение и язык. Коммуникативные системы животных и язык человека. М.: Языки славянских культур, 2008. С. 293–337. URL: <http://www.reznikova.net/Lango9.pdf>.

23. См.: Там же; Резникова Ж. И., Рябко Б. Я. Теоретико-информационный анализ «языка» муравьев // Журнал общей биологии. 1990. Т. 51. № 5. С. 601–609.

и время контакта, Ж. И. Резникова и Б. Я. Рябко определяли количество информации, передаваемой разведчиками фуражирам, и скорость передачи этой информации. В результате оказалось, что скорость передачи информации составляла около 1 бита в минуту.

Причем если путь был достаточно простой, например в каждой развилке лабиринта надо было поворачиваться только в левую ветку, то разведчики могли использовать «сжатую» информацию и передавать ее фуражирам с большей скоростью. То есть муравьи-разведчики не только передавали информацию муравьям-фуражирам, но и обладали способностью сжимать информацию и передавать ее фуражирам в сжатом виде.

Кроме этого, была продемонстрирована способность муравьев к простому счету чисел. Исследовалась передача информации между муравьями-разведчиками и фуражирами для более простых лабиринтов-гребенок, состоящих из одного основного канала, от которого ответвляются в одну сторону одинаковые каналы (до 40 веток). Пища находилась в одном из ответвлений. Муравьи оказались способными запоминать и передавать друг другу сведения о номере ветки, то есть они умели определенным образом считать.

Новокаледонские вороны могут самостоятельно изготавливать орудия. Обычно в природе новокаледонские вороны (Новая Каледония — остров к востоку от Австралии) могут обкусывать веточки так, что получаются простые орудия: заостренные палочки или крючки. На заостренную палочку ворона может насаживать личинки насекомых, а крючком вытаскивать личинок из-под коры.

В Университете Оксфорда проводили исследования с воронами, находившимися долгое время в неволе²⁴. Двум воронам (молодой самке и самцу постарше) предлагали добывать ведрко с пищей со дна прозрачного вертикального цилиндра. Рядом с цилиндром были прямая проволока и проволока, согнутая крючком. Ведрко можно было вытащить крючком, но не прямой проволокой. Раньше с проволокой вороны дела не имели. Тем не менее они сразу поняли, что ведрко можно вытащить с помощью крючка.

Неожиданность произошла, когда самец утащил крючок. Тогда самка сначала немного попыталась подцепить ручку ведерка прямым куском проволоки (что было безуспешно), а потом быстро научилась делать из прямой проволоки крючок, зажимая один конец проволоки в одной из щелей экспериментальной

24. См.: Weir A. A. S., Chappell J., Kacelnik A. Shaping of hooks in New Caledonian crows // Science. 2002. Vol. 297. № 5583. P. 981–983.

установки и загибая проволоку. А затем с помощью изготовленного ей крючка доставала ведро с пищей. В дальнейшем, если крючок убирала, а оставляли прямую проволоку, ворона сразу делала крючок и доставала пищу.

А что делал самец? Казалось бы, он должен был перенять опыт самки. Однако он не перенял этот опыт, а наблюдал за ней и примерно в трети случаев отнимал у нее пищу.

Таким образом, ворона сама, без какого-либо обучения, без каких-либо инструкций, изобрела способ изготовления орудия труда. Можно вспомнить Ф.Энгельса, который подчеркивал роль труда и изобретений орудий труда в процессе превращения обезьяны в человека. На примере новокаледонских ворон мы видим, что орудия способны изобретать и сравнительно простые животные²⁵.

Новокаледонские вороны могут мысленно составлять планы цепочек целенаправленных действий. Еще одно интересное исследование с новокаледонскими воронами провели исследователи из Новой Зеландии²⁶. Задание для ворон состояло из трех следующих частей. (1) Сначала надо было подтянуть к себе шнуром и освободить от шнура маленькую палочку, которая висела на шнуре. (2) Затем с помощью маленькой палочки достать из одного зарешеченного контейнера длинную палочку. (3) Наконец, с помощью длинной палочки надо было достать пищу из второго контейнера. Причем без короткой палочки нельзя было достать длинную, а короткой палочкой нельзя было дотянуться до пищи во втором контейнере. То есть весь процесс добывания пищи должен был состоять из трех четко определенных последовательных шагов, на которых надо было использовать все три «инструмента»: шнур, маленькую палочку, длинную палочку.

Предварительно вороны тренировались в более простых условиях: они могли использовать часть или все эти инструменты по отдельности. Когда им нужно было выполнить полное задание из трех шагов, то более тренированные вороны, которые имели опыт использования всех трех инструментов по отдельности, выполняли задание с первого раза. А менее тренированные вороны, которые предварительно освоили инструменты частично, выполняли задание не всегда с первого

25. Подробней об очерченном эксперименте см.: Там же, и сайт исследователей новокаледонских ворон Университета Оксфорда, URL: <http://users.ox.ac.uk/~kgroup/index.html>.

26. См.: Taylor A. H., Elliffe D., Hunt G. R., et al. Complex cognition and behavioural innovation in New Caledonian crows // Proceedings of the Royal Society B: Biological Sciences. 2010. Vol. 277. P. 2637–2643.

раза, но тем не менее тоже достаточно быстро научились с ним справляться.

Таким образом, вороны научились продумывать план решения новой задачи, мысленно связывая в плане отдельные элементы ранее освоенного опыта. Причем менее тренированные сумели не только мысленно объединить в плане элементы решения, но и догадаться до того, как выполнить недостающий элемент задания.

Автор настоящей статьи попытался продумать модель, отражающую мысленное составление планов воронами. Но наиболее близкое, что пришло в голову, — это аналогия со школьниками, решающими математические задачи. Когда решается математическая задача, то часто основная идея решения приходит как догадка. При этом чем больший опыт имеет школьник в решении задач, тем быстрее он решает сложную задачу. Это аналогично тому, как вороны, имеющий больший опыт в использовании инструментов, более быстро выполняли задание по мысленному сопоставлению плана цепочки целенаправленных действий.

«*Пространственный интеллект*», *запоминание и использование когнитивных карт*. Многие животные (птицы, грызуны, насекомые), ориентируясь в пространстве или запасая корм, формируют в своей памяти когнитивные карты местности. Например, колумбийские сойки, обитающие на юго-западе США, в конце лета начинают собирать семена сосны. Наполнив подъязычный мешок, сойка улетает на расстояние до нескольких километров, чтобы припрятать семена на южных склонах холмов, где зимой бывает мало снега. Птица может запастись до 33 тысяч семян, по 4–5 штук в одном месте, то есть получается несколько тысяч тайников. Зимой и весной птица наведывается в те места, в которых она прятала пищу в конце лета, и выкапывает пищу из своих складов. Итак, сойка запоминает и использует весьма нетривиальную когнитивную карту местности²⁷.

В заключение раздела подчеркнем, что экспериментальные биологические работы и моделирование когнитивных свойств биологических организмов по смыслу близки друг к другу, и возможно налаживание серьезных и интересных перспективных междисциплинарных связей между этими исследованиями. Можно двигаться навстречу друг другу представителям разных дисциплин: биологам и специалистам точных наук (математике и компьютерного моделирования). Более того, целесообразно

27. Подробнее о различных формах пространственного интеллекта животных см.: Резникова Ж. И. Указ. соч.

сочетание между экспериментальными (биологические эксперименты) и теоретическими (моделирование когнитивных способностей) исследованиями, подобное эффективному сочетанию работ между экспериментальной и теоретической физикой.

КОНТУРЫ ПРОГРАММЫ БУДУЩИХ ИССЛЕДОВАНИЙ КОГНИТИВНОЙ ЭВОЛЮЦИИ

Возвращаемся к основному обсуждаемому направлению исследований — моделированию когнитивной эволюции.

Отталкиваясь от анализа биологических экспериментов по когнитивным свойствам живых организмов, от уже разработанных моделей познавательных способностей организмов, можно предложить следующие этапы моделирования когнитивной эволюции:

- A. *Моделирование адаптивного поведения агентов с несколькими естественными потребностями: питание, размножение, безопасность.* Это могло бы быть моделирование достаточно естественного и полноценного поведения простых модельных организмов. Моделирование в этом направлении уже начато (см. ниже).
- B. *Исследование перехода от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов.* Такой переход можно рассматривать как появление в «сознании» животного свойства «понятие». Обобщенные образы можно представить как мысленные аналоги наших слов, не произносимые животными, но реально используемые ими. Использование понятий приводит к существенному сокращению и требуемой памяти, и времени обработки информации, поэтому оно должно быть эволюционно выгодным.
- B. *Исследование процессов формирования причинных связей в памяти животных.* Запоминание причинно-следственных связей между событиями во внешней среде и адекватное использование этих связей в поведении — одно из ключевых свойств активного познания животным закономерностей внешнего мира. Такая связь формируется, например, при выработке условного рефлекса: животное запоминает связь между условным стимулом (УС) и следующим за ним безусловным стимулом (БС), что позволяет ему предвидеть события в окружающем мире и адекватно использовать это предвидение. Возможно, что,

анализируя эволюционные корни процессов формирования связи между причиной и следствием, можно будет прояснить природу чувства привычки к нахождению причинных связей, о котором говорил Давид Юм.

Естественный следующий шаг — переход от отдельных причинных связей к логическим выводам на основе уже сформировавшихся знаний:

- Г. *Исследование процессов формирования логических выводов в «сознании» животных.* Фактически уже на базе классического условного рефлекса животные способны делать «логический вывод» вида: $\{УС, УС \rightarrow БС\} \Rightarrow БС$ или «Если имеет место условный стимул и за условным стимулом следует безусловный, то нужно ожидать появления безусловного стимула». В определенной степени такие выводы подобны выводам математика, доказывающего теоремы. И целесообразно разобраться в системах подобных выводов, понять, насколько адаптивна логика поведения животных и насколько она подобна нашей, человеческой логике.
- Д. *Исследование коммуникаций, возникновения языка.* Наше мышление тесно связано с языком, с языковым общением между людьми. Поэтому целесообразно проанализировать, как в процессе биологической эволюции возникал язык общения животных, как развитие коммуникаций привело к современному языку человека, как развитие коммуникаций и языка способствовало развитию логики, мышления, интеллекта человека.

Перечисленные пункты очерчивают круг исследований от моделирования простейших форм поведения к логическим правилам, используемым в математике. Опираясь на эти пункты, мы начали соответствующее моделирование.

НАЧАЛЬНЫЕ ШАГИ МОДЕЛИРОВАНИЯ

Модель автономных агентов с естественными потребностями. В работе²⁸ построена и исследована компьютерная модель автономных агентов, обладающих естественными для живых орга-

28. См.: Коваль А. Г., Редько В. Г. Поведение модельных организмов, обладающих естественными потребностями и мотивациями // Математическая биология и биоинформатика. 2012. Т. 7. № 1. С. 266–273. URL: [http://www.matbio.org/2012/Koval2012\(7_266\).pdf](http://www.matbio.org/2012/Koval2012(7_266).pdf).

низмов потребностями: питание, безопасность и размножение. Эта модель соответствует пункту А предложенных выше контуров программы.

В модели предполагалось, что в каждый такт времени одна из потребностей агента была наиболее приоритетной. Система управления агента была основана на наборе правил вида: $S_k \rightarrow A_k$ (в ситуации S_k нужно выполнить действие A_k). Ситуация S_k определялась активностью хищника рядом с агентом, предыдущим действием агента и приоритетной потребностью. Каждый такт времени агент мог выполнять одно из следующих действий A_k : 1) поиск пищи; 2) питание; 3) подготовку к размножению; 4) размножение; 5) оборону от хищника; 6) покой. Каждое правило имело свой вес W_k . Веса правил W_k настраивались методом обучения с подкреплением²⁹ в соответствии с результатами действий, направленными на удовлетворение приоритетных потребностей. Отметим, что метод обучения с подкреплением — хорошо известный метод обучения, он основан на получаемых агентом положительных и отрицательных наградах, получаемых агентом при поощрениях и наказаниях. Причем в этом методе предусмотрена процедура переоценки последовательно выполняемых действий, направленная на максимизацию суммарной награды, которую можно получить в будущем.

В результате компьютерного моделирования было показано, что формируется цикличность поведения агента. В цикле агент сначала накапливает свой энергетический ресурс путем питания, затем его действия направлены на максимизацию безопасности, то есть защиты от хищников, а когда обе потребности (пищевая и потребность безопасности) удовлетворены, агент размножается. После размножения действия агента снова направлены на питание, и цикл «питание → оборона → размножение» повторяется.

Таким образом, было показано, что происходит формирование циклов поведения агентов, в которых последовательно удовлетворяются потребности питания, безопасности и размножения³⁰.

Модель формирования обобщенных эвристик и простых обобщенных образов. Формирование обобщенных эвристик и простых обобщенных образов (пункт Б контуров программы) в процессе обучения агентов при поиске агентами пищи в двумерной клеточной среде было продемонстрировано в модели³¹.

29. См.: Саттон Р. С., Барто Э. Г. Обучение с подкреплением. М.: Бином, 2011.

30. Подробнее об этой модели см. в: Коваль А. Г., Редько В. Г. Указ. соч.

31. См.: Бесхлебнова Г. А., Редько В. Г. Модель формирования обобщенных понятий автономными агентами // Четвертая международная конференция

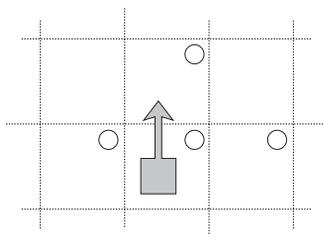


РИС. 2. Схема агента в двумерной клеточной среде. Агент обозначен квадратом, стрелкой — его направление «вперед». Окружностями отмечены клетки, из которых агент получает информацию о наличии или отсутствии пищи в клетке.

Схема такого агента показана на рис. 2. Модель предполагала, что в части клеток были порции пищи. При съедании пищи ресурс агента пополнялся. Система управления агента, как и в предыдущей модели, была основана на правилах вида $S \rightarrow A$ (S — ситуация, A — действие), веса правил оптимизировались методом обучения с подкреплением.

В результате обучения агент формировал обобщающие эвристики. Эти эвристики таковы: если имеется пища в той же клетке, в которой находится агент, то нужно съесть пищу; если в клетке агента нет пищи, но есть пища в клетке впереди или справа/слева от агента, то нужно выполнить действие «перемещение вперед» или «поворот направо/налево». Использование эвристик приводило к формированию последовательных цепочек действий, приводящих к пополнению ресурса агента.

Дополнительно в компьютерную программу вводилась процедура усреднения: вычислялось среднее число применений данного действия для той или иной ситуации. В результате усреднения агент формировал простые обобщенные образы или внутренние понятия: «имеется пища в моей клетке», «имеется пища в клетке впереди меня», «имеется пища в клетке справа/слева от меня».

Рассмотренные в настоящем разделе модели очень упрощенные, и их целесообразно рассматривать как начальный этап более полноценных исследований, нацеленных непосредственно на моделирование когнитивной эволюции.

ПРИМЕР ПОПУТНОГО ПРИКЛАДНОГО НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ — ИДЕЯ ПРОЕКТА НА НОБЕЛЕВСКУЮ ПРЕМИЮ МИРА

До сих пор рассматривались исследования, направленные на изучение фундаментальной проблемы — проблемы происхождения логического мышления. Но при серьезных исследованиях неизбежно попутно возникают и связанные с ними прикладные направления. Отметим одно из возможных прикладных направлений: устранение причин агрессивной конкуренции на основе моделирования эволюции конкурирующих агентов, которые могут вести борьбу между собой.

Вымирание генов агрессивности в эволюционирующей популяции конкурирующих агентов. В работах³² была построена и исследована компьютерная модель адаптивного поведения агентов в эволюционирующей популяции. В модели рассматривался двухмерный клеточный мир, в клетках могли находиться агенты и их пища. Агенты были такие же, как показано на рис. 2, однако они не только искали пищу, но и могли бороться друг с другом. В каждый такт времени агент мог выполнять одно из следующих действий: 1) питаться; 2) двигаться на одну клетку вперед; 3) повернуть налево; 4) повернуть направо; 5) отдыхать; 6) делиться (размножаться); 7) ударить агента в клетке спереди (напасть); 8) защищаться.

Каждый агент обладал внутренним энергетическим ресурсом R , который пополнялся, когда агент выполнял действие «питаться» и расходовался при выполнении других действий. При нападении одного агента на второго (выполнении действия «ударить») нападающий агент отнимал ресурс у второго агента, если второй агент не выполнял действие «защищаться»; но если второй агент защищался, то первый агент просто расходовал значительный ресурс. Также два агента теряли оба значительный ресурс, если они напали друг на друга.

Система управления агента представляла собой однослойную нейронную сеть, на вход сети подавалась сенсорная информация о ближайшем окружении агента и о его внутреннем ресурсе, выходы сети определяли действия агента. Входы ней-

32. См.: Бурцев М. С. Модель эволюционного возникновения целенаправленного адаптивного поведения. 2. Исследование развития иерархии целей // Препринт ИПМ РАН. 2002. № 69; Бурцев М. С., Редько В. Г. Влияние агрессии на эволюцию в многоагентной системе // Сборник трудов 9-й Международной конференции «Проблемы управления безопасностью сложных систем». М.: ИПУ, 2002.

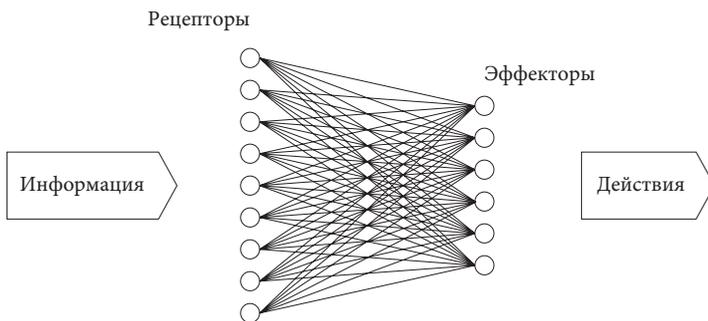


РИС. 3. Структура нейронной сети агента.

ронной сети представляли собой рецепторы или сенсоры агента, выходы сети — эффекторы агента (рис. 3).

Структура нейронной сети и веса синаптических связей между входом и выходом сети определяли геном агента. При делении потомок агента наследовал измененный мутациями геном, то есть нейронную сеть агента-родителя. При мутациях модифицировались как веса синаптических связей, так и структура нейронной сети, в частности могли исчезать и появляться отдельные рецепторы или эффекторы. Популяция агентов эволюционировала во времени за счет вариаций структуры и весов синапсов нейронных сетей и отбора тех агентов, которые достаточно быстро набирают ресурс и размножаются.

При моделировании в некоторых компьютерных экспериментах наблюдались неожиданные пики в зависимостях численности популяции агентов от времени. Анализ системы управления агентов показал, что пики соответствовали тем моментам времени, когда в нейронной сети агентов отсутствовали эффекторы, ответственные за борьбу между агентами. Для подтверждения этого полная компьютерная модель (с борьбой между агентами) сравнивалась с моделью, в которой эффекторы борьбы были полностью устранены из системы управления агентов. В случае такого «глобального пацифизма» установившаяся численность популяции была примерно в два раза выше, чем для обычных агентов, имеющих средства для борьбы между собой (рис. 4). Наличие агрессивной конкуренции приводило к тому, что агенты тратили свой ресурс R именно на борьбу друг с другом; этим агентам не хватало ресурса на «созидательную» жизнь, для активного накопления ресурса и расходования его на процессы размножения, приводящие к возрастанию численности популяции. Устранение борьбы между агентами, исчезновение

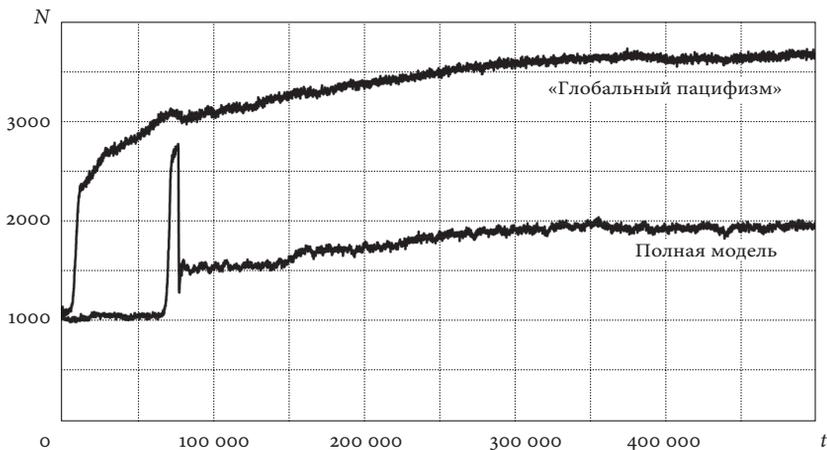


РИС. 4. Зависимость численности популяции N от времени t .

«генов агрессивности» приводило к увеличению численности популяции агентов.

Наблюдавшийся при этом эффект исчезновения «генов агрессивности» носил кратковременный характер: наблюдавшиеся пики в численности популяции были довольно узкими. Кратковременность связана с тем, что мутационное появление новых агрессивных агентов (с эффекторами борьбы) быстро приводило к тому, что агрессивные агенты вели активную борьбу с «мирными» агентами (не имевшими эффекторов борьбы) и выживали их из популяции.

На основе изложенной модели возможен анализ процессов исчезновения агрессивной конкуренции для различных социально-экономических приложений: для популяций конкурирующих фирм, регионов, государств. Например, если рассматривать эволюцию агентов как эволюцию государств, то можно думать о научном обосновании концепции глобального пацифизма и даже предложить идею проекта на Нобелевскую премию мира «Разработка научных основ всемирного разоружения».

ЗАКЛЮЧЕНИЕ

Возвращаемся к обсуждению общей постановки моделирования когнитивной эволюции. Как аргументировано выше, моделирование когнитивной эволюции — новое междисциплинарное направление исследований. Это направление связано с широким кругом дисциплин:

- 1) с основаниями математики;
- 2) с теорией познания;
- 3) с анализом познавательных способностей биологических организмов;
- 4) с когнитивными исследованиями;
- 5) с научными основами искусственного интеллекта.

Моделирование когнитивной эволюции направлено на пояснение фундаментальной научной проблемы: *почему логическое мышление человека применимо к познанию природы?*

Существенно, что в настоящее время уже имеются разносторонние заделы работ по моделированию когнитивной эволюции. Причем идейно эти заделы можно объединить концепцией моделирования когнитивных автономных агентов, что интересно как со стороны биологических наук, так и со стороны вычислительных наук. Предпосылки моделирования когнитивной эволюции связаны с интереснейшими исследованиями когнитивных способностей биологических организмов, причем эти способности вполне можно сопоставлять с элементами мышления человека.

Отталкиваясь от известных моделей когнитивных автономных агентов, можно предложить контуры программы будущих исследований когнитивной эволюции. Пункты программы очерчивают круг исследований от моделирования простейших форм поведения к логическим правилам, используемым в математике. Сопоставляя пункты контуров программы будущих исследований когнитивной эволюции с известными моделями, можно заключить, что уже имеются отдельные элементы, соответствующие каждому из пунктов. Образно говоря, у нас уже есть некоторые небольшие фрагменты картины, но мы еще не видим всей картины. Четкой последовательности серьезных, канонических моделей, которые показывали бы общую картину происхождения логического мышления, пока еще нет.

Еще раз подчеркнем, что исследования когнитивной эволюции нацелены на понимание причин применимости логического мышления в научном познании и тем самым на укрепление фундамента науки.

REFERENCES

- Beskhlebnova G. A., Red'ko V. G. Model' formirovaniia obobshchennykh poniatii avtonomnymi agentami [Model of the Autonomous Agents' General Concepts Formation]. *Chetvertaia mezhdunarodnaia konferentsiia po kognitivnoi nauke. Tezisy dokladov: v 2 t.* [Fourth International Conference on Cognitive Science. Abstracts in 2 volumes], Tomsk, TGU, 2010, vol. 1, pp. 174–175.

- Bongard M. M., Losev I. S., Smirnov M. S. Proekt modeli organizatsii povedeniia — “Zhivotnoe” [«Animal»: A Project Model of Behavior Organisation]. *Modelirovanie obuchenii i povedeniia* [Modeling of Learning and Behavior], Moscow, Nauka, 1975, pp. 152–171.
- Burtsev M. S. Model' evoliutsionnogo voznikoveniia tselenapravlennoogo adaptivnogo povedeniia. 2. Issledovanie razvitiia ierarkhii tselei [Model of Evolutionary Emergence of Purposeful Adaptive Behavior. 2. Study of the Development of the Hierarchy of Goals]. *Preprint IPM RAN* [Preprint of Keldysh Institute of Applied Mathematics (Russian Academy of Sciences)], 2002, no. 69.
- Burtsev M. S., Red'ko V. G. Vliianie agressii na evoliutsiiu v mnogoagentnoi sisteme [Influence of Aggression on the Evolution in a Multi-Agent System]. *Sbornik trudov 9-oi Mezhdunarodnoi konferentsii “Problemy upravleniia bezopasnost'iu slozhnykh sistem”* [Proceedings of the 9th International Conference “Problems of Security Management of Complex Systems”], Moscow, IPU, 2002.
- Gaaze-Rapoport M. G., Pospelov D. A. *Ot ameba do robota: modeli povedeniia* [From Ameba to the Robot: Behavior Models], Moscow, URSS, 2011.
- Hume D. Issledovanie o chelovecheskom poznanii [An Enquiry Concerning Human Understanding is]. *Sochineniia: v 2 t.* [Works in 2 vols], Moscow, Mysl', 1964, vol. 2, pp. 5–169.
- Idel'son A. V., Mints G. E., eds. *Matematicheskaia teoriia logicheskogo vyvoda* [The Mathematical Theory of Logical Derivation], Moscow, Nauka, 1967.
- Kant I. Kritika chistogo razuma [Die Kritik der reinen Vernunft]. *Sochineniia: v 6 t.* [Works in 6 vols], Moscow, Mysl', 1965, vol. 3, pp. 69–695.
- Kant I. Prolegomeny ko vsiakoi budushchei metafizike, mogushchei poiavit'sia kak nauka [Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können]. *Sochineniia: v 6 t.* [Works in 6 vols], Moscow, Mysl', 1965, vol. 4, part 1, pp. 67–210.
- Koval' A. G., Red'ko V. G. Povedenie model'nykh organizmov, obladaiushchikh estestvennymi potrebnoostiami i motivatsiiami [Behavior of Model Organisms Which Have Natural Needs and Motivations]. *Matematicheskaia biologii i bioinformatika* [Mathematical Biology and Bioinformatics], 2012, vol. 7, no. 1, pp. 266–273. Available at: [http://www.matbio.org/2012/Koval2012\(7_266\).pdf](http://www.matbio.org/2012/Koval2012(7_266).pdf).
- Krichmar J. L., Seth A. K., Nitz D. A., Fleischer J. G., Edelman G. M. Spatial Navigation and Causal Analysis in a Brain-Based Device Modeling Cortical-Hippocampal Interactions. *Neuroinformatics*, 2005, vol. 3, no 3, pp. 197–221.
- Langley P., Laird J. E., Rogers S. Cognitive Architectures: Research Issues and Challenges. *Cognitive Systems Research*, 2009, vol. 10, no. 2, pp. 141–160.
- Langton C. G., ed. *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, Redwood City, CA, Addison-Wesley, 1989.
- Lorenz K. Kantovskaia kontsepsiia a priori v svete sovremennoi biologii [Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie]. *Evoliutsiia. Iazyk. Poznanie* [Evolution. Language. Cognition], Moscow, Iazyki russkoi kul'tury [Languages of Russian Culture], 2000, pp. 15–41.
- Meyer J.-A., Wilson S. W. *From Animals to Animats. Proceedings of the First International Conference on Simulation of Adaptive Behavior*, London, Cambridge, MA, MIT Press, 1990.
- Nepomniashchikh V. A., Popov E. E., Red'ko V. G. Bionicheskaia model' adaptivnogo poiskovogo povedeniia [Bionic Model of Adaptive Searching Behavior]. *Izvestiia RAN. Teoriia i sistemy upravleniia* [Journal of Computer and Systems Sciences International], 2008, no. 1, pp. 85–93.

- Red'ko V. G. *Evolutsiia, neuronnye seti, intellekt. Modeli i kontseptsii evoliutsionnoi kibernetiki* [Evolution, Neural Networks, Intelligence. Models and Concepts of Evolutionary Cybernetics], Moscow, URSS, 2005.
- Red'ko V. G. *Ot modelei povedeniia k iskusstvennomu intellektu* [From Behavior Model to Artificial Intelligence], Moscow, URSS, 2006.
- Reznikova Zh. I. *Sovremennye podkhody k izucheniiu iazykovogo povedeniia zhiivotnykh* [Modern Approaches to the Study of Language Behavior of Animals]. *Razumnoe povedenie i iazyk. Kommunikativnye sistemy zhiivotnykh i iazyk cheloveka* [Intelligent Behavior and Language. Communication Systems of Animals and Human Language], Moscow, Iazyki slavianskikh kul'tur [Languages of Slavic Cultures], 2008, pp. 293–337.
- Reznikova Zh. I., Riabko B. Ia. *Teoretiko-informatsionnyi analiz "iazyka" murav'ev* [Information-Theoretic Analysis of the "Language" of Ants]. *Zhurnal obshchei biologii* [Biology Bulletin Reviews], 1990, vol. 51, no. 5, pp. 601–609.
- Samsonovich A. V., Johansdottir K. R., eds. *Biologically Inspired Cognitive Architectures 2011. Proceedings of Second Annual Meeting of the BICA Society*, Amsterdam, Berlin, Tokyo, Washington, DC, IOS Press, 2011.
- Sutton R. S., Barto E. G. *Obuchenie s podkrepleniem* [Reinforcement Learning], Moscow, Binom, 2011.
- Taylor A. H., Elliffe D., Hunt G. R., Gray R. D. *Complex Cognition and Behavioural Innovation in New Caledonian Crows. Proceedings of the Royal Society B: Biological Sciences*, 2010, vol. 277, pp. 2637–2643.
- Teilhard de Chardin P. *Fenomen cheloveka* [Le Phénomène humain], Moscow, Ustoychivyi mir, 2001.
- Tsetlin M. L. *Issledovaniia po teorii avtomatov i modelirovanie biologicheskikh sistem* [Studies on Automata Theory and Modeling of Biological Systems], Moscow, Nauka, 1969.
- Turchin V. F. *Fenomen nauki. Kiberneticheskii podkhod k evoliutsii* [The Phenomenon of Science. Cybernetic Approach to Evolution], Moscow, ETS, 2000. Available at: <http://www.refal.ru/turchin/phenomenon/>.
- Vernon D., Metta G., Sandini G. A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents. *IEEE Transactions on Evolutionary Computation. Special issue on Autonomous Mental Development*, 2007, vol. 11, no. 2, pp. 151–180.
- Weir A. A. S., Chappell J., Kacelnik A. *Shaping of Hooks in New Caledonian Crows. Science*, 2002, vol. 297, no. 5583, pp. 981–983.
- Wenner A., Wells P. *Anatomiia nauchnogo protivostoianiia. Est' li iazyk u pchel?* [Anatomy of a Controversy: The Question of a "Language" Among Bees], Moscow, Iazyki slavianskikh kul'tur [Languages of Slavic Cultures], 2011.