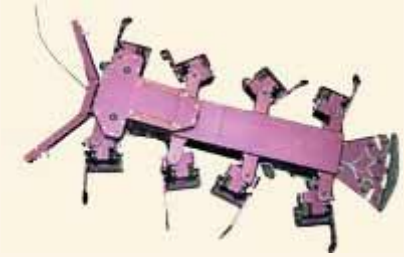


Тема этого номера - компьютерное моделирование как инструмент исследования когнитивных, интеллектуальных процессов в мозге человека или животного.
В обстоятельной беседе со мной ***Константин Анохин*** *(в свои сорок четыре - профессор, член-корреспондент Российской Академии медицинских наук, заведующий отделом системогенеза Института нормальной физиологии им. П. К. Анохина и руководитель Российско-британской лаборатории нейробиологии памяти) рассказал о месте такого моделирования в общей стратегии исследований, об основных современных концепциях организации мозговых структур, управляющих поведением.*

Константин Владимирович, уже столько лет люди пытаются моделировать интеллектуальные функции, но успехи довольно скромны. Не кажется ли вам, что в этой области наметился идейный тупик?
- Кажется. Но как любил говорить Черчилль: «любой кризис это новые возможности». Поэтому именно сегодня я с оптимизмом смотрю на проблему моделирования поведения и интеллекта.

Несмотря на очевидный крах высоких начинаний в области ИИ?
- Мы, видимо, говорим о разных вещах. Вы, вероятно, имеете в виду классический искусственный интеллект, КИИ, или, как его называют на Западе, GOFAI - Good Old Fashioned Artificial Intelligence (старый добрый ИИ). Действительно, это направление столкнулось с серьезной критикой в 1980-90 годы и оказалось в застое. Но иначе и быть не могло. В методических предпосылках этого направления уже изначально содержались серьезные изъяны.

То есть?
- Одним из условий, приведших к неудачам в КИИ, было игнорирование важнейшего факта: интеллект впервые появился вовсе не у человека. Классический искусственный интеллект всегда понимали как задачу обратной разработки (reverse engineering): взять готовый конечный продукт, интеллект человека, реконструировать его базовые алгоритмы и создать на этой основе искусственный продукт. При этом терялись многие важные свойства, вытекающие из принципиального факта: человеческий интеллект является продуктом преемственной эволюции усложнявшегося адаптивного поведения. Интеллектом как способностью к решению проблем и адаптации в непредсказуемой среде, способностью к предвидению и выбору между разными вариантами действий, к обучению и формированию знаний обладают многие живые существа, даже с примитивной нервной системой. Поэтому фундаментальные модели интеллекта - это обязательно модели свойств, которые имеются не только у человека. Такая эволюционная ориентация позволяет, в частности, шкалировать сложные задачи: если сходу не удается создать мозг человека - попробуйте создать мозг обезьяны; если и это сложно - попробуйте дрозофилу или хотя бы воспроизведите адаптивное поведение инфузории.

Ну и в какой же точке на этой шкале мы сейчас находимся?
- Поведенческое моделирование появилось очень давно, вспомните хотя бы «черепах» Грея Уолтера (Gray Walter) в 1950-е годы. Но в 1960-е годы оно было вытеснено КИИ, где вопрос об ориентации на законы развития и эволюции поведения почти не ставился. К этой задаче вновь вернулись лишь в конце 1980-х, когда многие специалисты по ИИ вместо сверхсложных экспертных систем занялись попытками создавать пластичные, быстро работающие компьютерные модели или роботы,

выживающие в той же среде, где существуют реальные животные. Эти модели, в отличие от КИИ, не имеют предварительной схемы мира, они должны строить ее под существующую среду. Такие системы адаптируются к среде с помощью набора элементарных поведенческих модулей, каждый из которых отвечает за какую-то независимую элементарную форму поведения, а затем методом проб и ошибок происходит отбор и фиксация таких модулей по результатам их действий. Планы, цели, схемы поведения не закладываются. Они эмерджентны[1] складываются «снизу вверх» из результатов адаптивного поведения. Такие системы получили название искусственных животных - «аниматов» (от англ. animal - животное и automat - автономно действующее устройство) и стали особенно известны после состоявшейся в 1990 году в Париже конференции «Симуляция адаптивного поведения: от животных к аниматам». Квинтэссенцию различий КИИ и поведенческого моделирования, на мой взгляд, остроумно отражает подмеченная еще Бертраном Расселом национальная разница в школах немецкой гештальтпсихологии и американского бихевиоризма: «Животные, которых наблюдали немцы, спокойно сидят, думают и в конце концов выдают решение из своего внутреннего сознания. А животные, которых изучали американцы, бешено прыгают с невероятным напором и темпераментом и, наконец, случайно получают желаемый результат».

Сегодня идет поиск общих принципов шкалируемой архитектуры адаптивного поведения. Это очень интересная задача. Первыми существами, чье поведение пытались моделировать, были насекомые - пауки, пчелы, осы. Например, в лаборатории ИИ в MIT (http://www.ai.mit.edu/), руководимой Родни Бруксом (Rodney Brooks), было создано самодвижущееся шестиногое 35-сантиметровое «насекомое-робот» Genghis, которое, передвигаясь по неровному ландшафту с препятствиями, адаптивно принимало те или иные решения. В Париже есть лаборатория аниматов (AnimatLab, animatlab.lip6.fr/index_en.html), работающая с похожим насекомоподобным ходящим роботом Hexapod. В Англии, в университете Сассекса (University of Sussex),[2] работает большой центр вычислительных нейронаук и робототехники (CCNR - www.cogs.sussex.ac.uk/ccnr/index.html), создающий насекомоподобные «интеллектуальные существа».

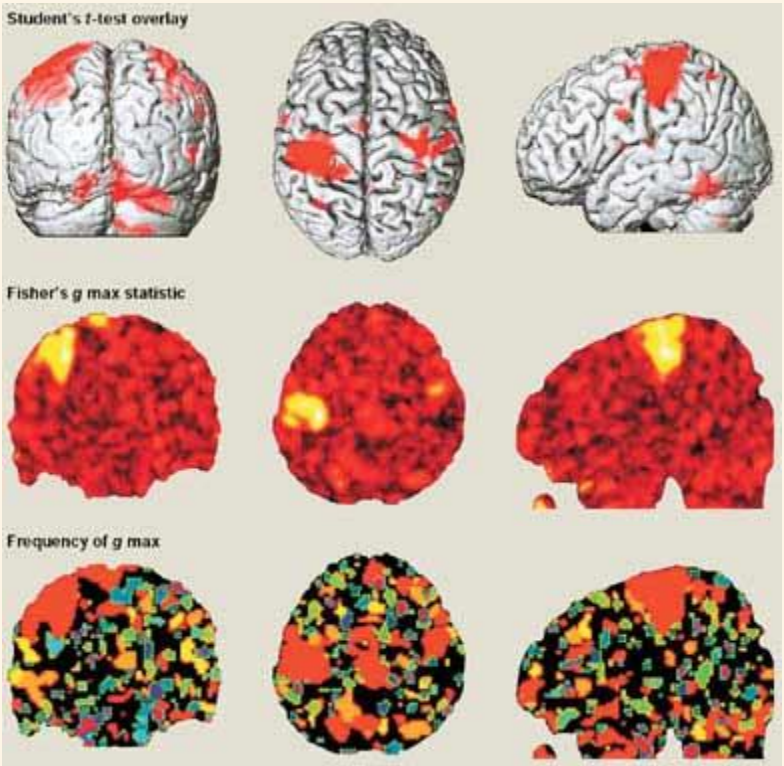
- Да, но не будем забывать, что это лишь первые шаги нового подхода. Сегодня, например, в Animat Lab разрабатывается проект Psikharpax, где в роботе синтезируются некоторые из адаптивных механизмов и нервных структур, ответственных за пространственную навигацию у крыс. Способности этой крысы-робота будут расти за счет «обучения без учителя», то есть анимат будет сам строить когнитивную карту среды и вырабатывать адаптивные стратегии поведения по механизмам, схожим с теми, что использует мозг крысы. В группе гуманоидной робототехники (Humanoid Robotics Group, www.ai.mit.edu/projects/humanoid-robotics-group) из лаборатории ИИ в MIT сегодня разрабатываются обезьяноподобные и мобильные роботы (Kismet, Coco) с гораздо более сложным, чем у первых насекомоподобных роботов, поведенческим репертуаром, куда, в частности, входят способности к социальным взаимодействиям и аффективным эмоциональным реакциям.

Кроме того, эксперименты с такими роботами могут стимулировать появление новых идей, проливающих свет на принципы адаптивного поведения. Как минимум, такие эксперименты позволяют отсекать заведомо нереалистичные теории. Если согласиться с Джоном Арчибальдом Уилером (J. A. Wheeler), который считал, что задача ученого в том, чтобы совершать ошибки как можно быстрее, то моделирование как средство элиминации ошибок - мощный инструмент в познании работы мозга. Поэтому многие нейробиологи настаивают на том, что теории работы мозга должны быть сформулированы алгоритмично, чтобы допускать моделирование. Это очень сильное условие! Один из ведущих нейробиологов-

теоретиков, Нобелевский лауреат Джеральд Эделман (Gerald Edelman), стал и одним из пионеров эволюционного обучения роботов. Эделман, создавший фундаментальную теорию работы мозга и биологических основ сознания[3] в своем институте в Калифорнии (Neuroscience Institute) разрабатывает серию роботов NOMAD (Neurally Organized Mobile Adaptive Devices). Эти роботы имеют еще и родовое имя «Дарвин» (сейчас уже существует четвертое поколение этой серии). Каждый новый «Дарвин» появляется на свет практически необученным, но сталкиваясь с объектами внешнего мира и имея какое-нибудь врожденное предпочтение, начинает вырабатывать собственные абстрактные категории (научается, например, отличать друг от друга разные шары). У робота появляются знания, которые он может использовать и в других задачах. То есть начинает работать один из принципов, по которым, судя по всему, шла эволюция механизмов интеллекта.

Смущает, что об успехах в таких экспериментах пишут как о «прорывах», о «первых шагах». Однако о «вторых шагах» что-то не слышно. Может быть, поэтому меня и, думаю, многих читателей не меньше компьютерного моделирования в этой области интересует прямое изучение процессов мышления, восприятия, других высших функций мозга. Вы наблюдаете за тем, какие процессы происходят в работающем мозге, заглядываете внутрь. Что вы там видите?

- Наш мозг состоит из 1011 нейронов, образующих между собой примерно 1014 связей и имеющих гиперастрономическое число возможностей взаимодействия. Чтобы понять, как работает такой сложнейший объект, бесполезно просто заглядывать внутрь и описывать, что мы там увидим. Даже если описать работу всех до единой клеток мозга в любой момент времени, мы все равно не поймем, как возникают мышление, восприятие, психика. Следовательно, нам прежде всего требуется ясное представление о том, какой вид ответа нам нужен, что нас устроит в качестве объяснения.



На фотографиях показана активность мозга человека, визуализируемая с помощью функционального магнитно-резонансного нейрокартирования (ФМРН) (из: John D. Van Horn and Michael S. Gazzaniga, Databasing fMRI studies - towards a 'discovery science' of brain function, Nature Reviews Neuroscience, 2002, V 3, 314-318).

Каким требованиям должно отвечать такое решение, если принять изложенную выше эволюционную перспективу взгляда на психику и интеллект? Допустим, мы хотим получить ответ на знаменитый вопрос Сеченова в «Элементах мысли»: как мозг формирует абстрактные категории? Сегодня у нас для этого есть великолепные методы функционального магнитно-резонансного нейрокартирования, и мы можем увидеть, какие области человеческого мозга вовлекаются в выработку абстрактных категорий. Но будет ли описание этих

структур мозга отвечать нужным нам образом на вопрос о фундаментальных механизмах категоризации? Нет.

Почему я так говорю? Некоторое время назад американские психологи Ричард Гернштейн (Richard Herrnstein) и Джон Серелла (John Cerella) провели удивительные эксперименты: они показывали голубям серии цветных слайдов и награждали птиц пищей только в том случае, если те тыкались клювом в слайды с изображением деревьев. Вскоре голуби начинали безошибочно узнавать деревья в любых новых слайдах, независимо от того, стоит ли дерево отдельно или это кромка леса, или даже отражение дерева в воде. При этом птицы четко отличали деревья от кустарников или, например, от крупноплановой фотографии ботвы сельдерея. И дело здесь вовсе не в выработанной эволюцией способности опознавать деревья как экологически значимый объект - с такой же легкостью голуби вырабатывали, например, зрительную категорию «рыба» на подводных слайдах морских рифов.

Теперь мы знаем, что формировать абстрактные категории могут не только люди, но и птицы. Однако мозг птиц устроен иначе, чем у высших млекопитающих. Такую же категоризацию, как мы, птицы выполняют иными структурами своего мозга. Это значит, что ответ на вопрос о принципиальных механизмах этой высшей психической функции не может быть дан в терминах уникальных структур головного мозга млекопитающих или мозга птиц.

То же самое касается и других когнитивных механизмов. Возьмем, например, способность к счету. У человека за нее отвечают некоторые области коры головного мозга, в частности нижнетеменная. Но опыты Л. А. Зориной и ее сотрудников из МГУ показывают, что вороны тоже способны считать. Однако эти птицы не имеют сколько-нибудь развитой коры и поэтому должны производить счет совсем другими мозговыми структурами. Еще один пример - механизмы рабочей памяти и поддержания цели. У человека эти функции зависят от лобных отделов коры головного мозга, а у птиц те же задачи, по-видимому, решаются расположенными в затылочной области зонами латерального каудального неостриатума. Мы вновь и вновь видим, что фундаментальные свойства интеллекта способны реализовываться у разных видов животных с самой разной архитектурой нервной системы.

Доводя анализ до логического конца, я должен сказать, что ответ, который мы ищем, вряд ли может быть получен в терминах описания специфических нервных структур. Допустим, что в мозге птиц и млекопитающих еще можно искать некое морфологическое сходство структур, выполняющих одинаковую когнитивную функцию. Но цель поведения отлично «держат в голове» и пчелы. Немецкий нейробиолог Рандольф Менцель (Randolf Menzel) показал, что пчелы способны удерживать в рабочей памяти то, какие цветы они уже облетали и к каким они направляются, даже если убрать эти цветы из поля их прямой видимости. Мозг пчелы при этом имеет в миллион раз меньше нейронов, чем мозг человека, и в нем нет ни одной структуры, похожей на высшие отделы человеческого мозга, отвечающие за поддержание цели и рабочей памяти.

Это значит, что фундаментальные механизмы психологических функций мозга должны быть сформулированы в терминах, не зависящих от конкретного строения мозга того или иного организма, несущего эти функции. Пользуясь компьютерной терминологией, можно сказать, что теория высших функций мозга должна быть построена на базе платформнонезависимого языка.

Почему я уделяю этому требованию такое внимание? Потому что существование этого языка - основная надежда для возможности переноса свойств интеллекта с одной (естественной) платформы на другую (искусственную). Теория высших функций

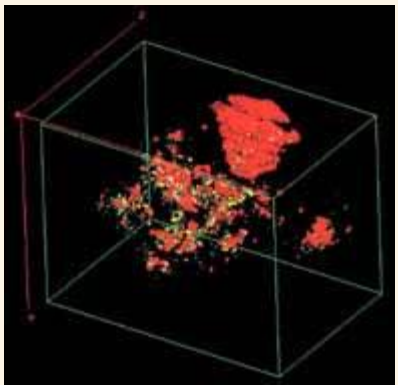
мозга, сформулированная на платформнонезависимом языке, - единственный мост, который может реально связать науки о мозге и искусственном интеллекте.

Продолжение интервью с Константином Петровичем Анохиным (Институт нормальной физиологии им. П. К. Анохина).

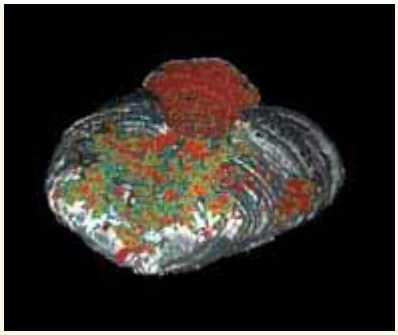
В вашем рассказе об истории развития научных концепций работы мозга упоминается школа Анохина П.К.. Известно, что создание Анохиным теории функциональных систем (ТФС) стало одним из важных этапов этой истории. В чем суть ТФС, находящей сегодня множество экспериментальных подтверждений?[4]

- Важно понять, что ТФС относится не к физиологии или психологии, она адресуется к особому промежуточному слою процессов мозга, которые определяют когнитивные свойства и поведение. Нервные клетки человека, находящегося под наркозом или в коме, продолжают работать. При некоторых видах наркоза активность даже выше, чем в обычном состоянии. Но свойства сознания, психики и разума у таких животных и человека отсутствуют. Эти свойства возникают не из-за самого факта работы массы клеток мозга, а из-за их объединения в организованные надклеточные образования. Можно сказать, что языком работы бодрствующего мозга, языком психики и поведения является не язык отдельных нервных клеток, а язык таких систем. Эти системы и получили названия функциональных систем (ФС), из-за тесной связи их деятельности с выполнением тех или иных функций организма. В ТФС была поставлена задача - изучить специфические принципы самоорганизации системных процессов, ведущих к появлению адаптивного поведения со всеми его психологическими атрибутами.

Но эти процессы долгое время нельзя было исследовать объективно. Не существовало методов, позволявших залезть в мозг и посмотреть, как нервные клетки объединяются в такие системы. Только в 1960-х годах стало возможно регистрировать нервные клетки мозга животных в свободном поведении. В нашем институте этим занимались несколько групп. Пожалуй, наиболее подробно изучал эти процессы Вячеслав Борисович Швырков со своими сотрудниками, образовавшими потом лабораторию нейрональных основ психики в Институте психологии. Они установили (сейчас на Западе начинают об этом вспоминать), что в разных областях мозга бодрствующих животных в поведении синхронно активируются миллионы нервных клеток, работающих по принципу ФС. Тем самым два важнейших предположения ТФС - что такие системы являются распределенными (элементы из разных структур объединяются для достижения общего результата) и синхронизованными - были подтверждены экспериментальными наблюдениями.



Можно ли в буквальном смысле увидеть эти системы в мозге?



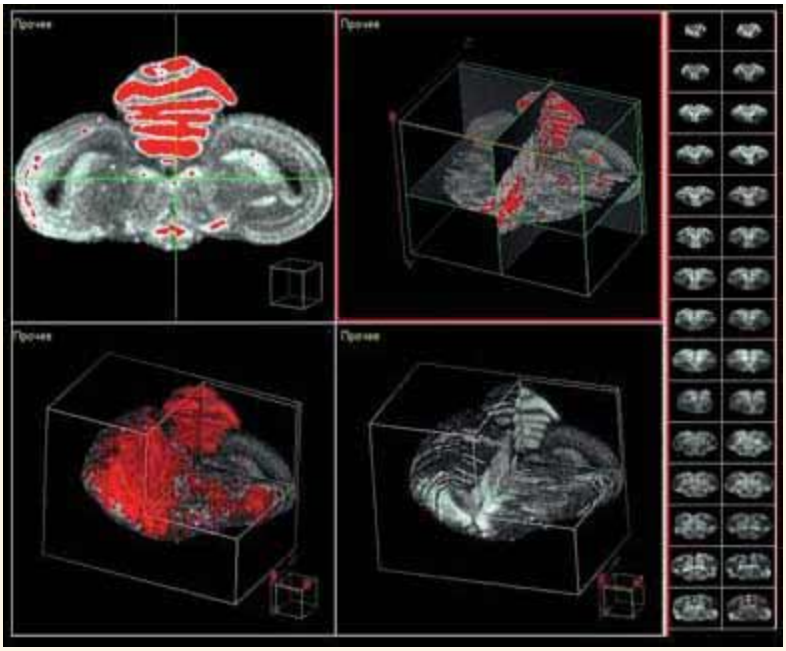
- В середине 1980-х мы начали искать методы такой визуализации. Так как в каждый момент в мозге работает множество уже сформированных ФС, при регистрации электрической активности невозможно выделить вклад каждой из них. Поэтому мы стали изучать процесс формирования отдельных ФС. Например, в описанной Конрадом Лоренцом классической модели импринтинга (запечатления), цыпленок должен запомнить образ курицы и всюду следовать за ней. Мы изучали формирование функциональных систем, которые это обеспечивают. Другой пример - ФС, которые складываются целиком в индивидуальном опыте организма, без каких-либо эволюционных предпосылок. Например, чтобы обеспечить нажимание животным в клетке педали для получения корма. Мы исследовали, что общего у тех клеток, которые впервые объединяются в такие ФС: нет ли в момент формирования ФС неких общих химических процессов у всех клеток этой системы? Мы искали ответ в ядре клетки, используя методы молекулярной биологии и генетики - и нам удалось найти гены, которые обычно молчат в мозге, и начинают работать только в тот момент, когда клеткам надо надолго запомнить новую информацию. Зонды, выявляющие работу этих генов в мозге, позволяют увидеть нервные клетки разных отделов мозга, образующие ФС в момент ее самого первого формирования. Это визуализируется на срезах мозга, где четко видны нейроны, в которых произошла эта реакция.

Но что именно запоминает нейрон?[5]

- Есть два фактора возникновения новой ФС - **новизна** и **синхронизация**. После возникновения новой проблемы, новой ситуации, запускаются **процессы обучения** - мозг начинает перебирать новые варианты ФС, в нейронах которых и происходит экспрессия этих генов. Одновременный приход сигналов по многим входам в нейронах таких новых систем и является условием запуска этой генетической реакции, с помощью которой нейроны запоминают тех «собратьев», вместе с которыми они участвовали в данной ФС. При этом критическую роль здесь играет фактор «новизны» - то, что клетки, входящие в новую пробную систему, впервые связались в такой конфигурации. Клетка способна «понимать», что она никогда еще не входила в такую систему связей, и бешено экспрессирует соответствующие гены. Потом экспрессия быстро ослабевает - во второй, третий раз она еще заметна, затем постепенно исчезает. Основное, что нас сейчас интересует - как связать эти ФС с поведением.

То есть показать, что если на эти пятьсот тысяч клеток синхронно подать вот такие колебания, то человек заиграет на пианино?..

- Нет, такой экстравагантный эксперимент по стимуляции нейронов распределенной по всему мозгу ФС пока сделать нельзя. Но можно, кстати, сделать обратное - «стереть» такую систему из памяти. Дело в том, что в мозге можно, с помощью «антисмысловых» олигонуклеотидов (синтетических фрагментов ДНК) направленно выключить работу генов, которые экспрессируются при запоминании.[6] Мы это делали; такие животные способны нормально учиться, но не способны запомнить то, чему научились - у них **не формируется долговременная память**.



Однако главный смысл наших поисков состоит в другом. Конечно, с помощью найденного нами инструмента мы теперь можем изучать ранее совершенно недоступные стороны работы мозга. Например, можно составлять мозговые карты тех или иных форм памяти, задавать вопрос о том, как те или иные структуры мозга участвуют в том или ином поведении или когнитивном процессе. Более того, мы можем исследовать таким образом работу мозга млекопитающих и птиц, потому что обнаруженные нами гены работают в очень широком спектре организмов. Но если мы будем собирать информацию лишь о конфигурациях нервных структур, активирующихся при тех или иных когнитивных процессах у тех или иных организмов, мы едва ли приблизимся к задаче построения фундаментальной теории работы мозга. Такой феноменологический путь, по которому движется западная когнитивная нейронаука, увлекаемая фантастическими возможностями нейрокартирования психологических процессов в мозге человека, ведет, по моему глубокому убеждению, к неизбежному теоретическому кризису. В его конце нас может ожидать лишь исчерпывающая функциональная топография мозга, без единой концепции механизмов его высших функций. Для их объяснения нам необходим «платформнонезависимый» язык. В ТФС мы отождествляем такой язык с закономерностями **промежуточного системного слоя**, лежащего **между физиологическими процессами в нервных клетках и уровнем психологических феноменов**. И если мы ясно видим все эти опасности и перспективы, значит нам уже сегодня нужно, применяя новые инструменты исследования мозга, учиться задавать с их помощью вопросы о специфических законах организации этого системного слоя.

Чтобы получить порцию пищи, собака обучается бегать от кормушки к педали и назад. Вопрос - как возникла в ее мозге и опыте функциональная система, связывающая нажатие педали с открыванием кормушки? Почему нейроны в удаленных друг от друга областях мозга сливаются при этом в единую систему? По каким принципам происходит извлечение таких распределенных систем из памяти? Как осуществляется переключение с одних функциональных систем на другие после достижения результатов? Можем ли мы смоделировать эволюцию таких систем, объяснить, как они появляются в развитии мозга? Вот вопросы, которые мы пытаемся решить.

Одним из самых узких мест в такой программе исследований на сегодняшний день мне представляется, как ни странно, не ограниченность методов изучения работы мозга, а отсталость наших подходов к объективному анализу поведения. Поведение, согласно ТФС, представляет собой континуум системных процессов, очень структурированный, разбитый на отдельные поведенческие акты. Все они имеют одинаковую структуру - планирование действий, модели результата, оценка, переход к следующему действию. Все это организовано в сложную иерархию, но мы пока очень мало знаем о ней. Нас сегодня интересует возможность, объективно регистрируя поведение агента (человека или животного) «читать» эту внутреннюю организацию, выявлять эту структуру.

Такое разложение поведения подтверждено на физиологическом уровне?
- Про науку вообще нельзя говорить, что в ней что-либо «подтверждено». Это составляет часть нашей теории и нашей исследовательской программы. Например, если регистрировать нейроны мозга в поведении, можно увидеть нейроны «ошибочных» систем. Животное бегает от педали к кормушке, чтобы нажать педаль. Иногда оно промахивается. Почему? В экспериментах Леонида Бобровникова и Натальи Швырковой в нашем институте выяснилось, что **при ошибочном действии в мозге работают другие нейроны**. То есть уже в момент планирования, поворота от кормушки к педали сложилась такая функциональная система, которая приведет к промаху! Кролик еще не дошел до педали, а мы, регистрируя нейроны, можем сказать, что через две секунды он промахнется. Такой же феномен позже обнаружили у обезьян при сложной манипуляторной деятельности группы Патриции Голдман-Ракич (P. Goldman-Rakic) из Йельского университета и Вольфрама Шульца (W. Schultz) из Фрайбургского университета. И такую же вещь в ноябре прошлого года описали американские исследователи в работе лобных отделов коры головного мозга человека, используя методы функционального нейрокартирования. Я кое-что упрощаю для ясности, но по сути все обстоит так.

Так вот, мы столкнулись с тем, что существующего языка анализа поведения сегодня не хватает, чтобы соотнести разрешение, получаемое нами на уровне мозга, с теми феноменами поведения, которые мы наблюдаем. Анализ поведения животных и человека в этологии, экспериментальной психологии, психофармакологии, нейропсихологии до сих пор строится либо на словесном описании поведения, либо на простых количественных измерениях времен реакции, пройденного расстояния и т. п., то есть показателей, которые неизмеримо беднее истинного содержания поведения. Поэтому мы сегодня работаем над возможностью, используя системы видеорегистрации и компьютерного анализа изображений, извлекать объективную информацию о поведении, его структуре, организации - о языке поведения. А затем хотим попытаться соотнести этот отпрепарированный, расквантованный на дискретные акты континуум поведения с работой мозга.

Все это связано только с движением?
- Да, в основном с движением животных. Но у животных почти любое поведение связано с движением. У человека, конечно, мыслительные процессы нельзя включить в классическое определение движения.

Насколько концепция ФС общепринята? Например, я знаю о концепции нейрофизиолога из Беркли Уолтера Фримена (Walter Freeman, см. «КТ» #365), основанной на аттракторах хаотических процессов, которые он наблюдает в мозге.
- Фримен - выдающийся нейрофизиолог, но он проводил свои классические работы, регистрируя электроэнцефалограмму (ЭЭГ) - тогда ничего другого не было. А ЭЭГ - очень грубый метод, в нем усредняются сигналы от сотен и тысяч клеток, хотя Фримен и старался максимально повысить разрешающую способность метода. Поэтому при помощи ЭЭГ можно выявить только наличие самоорганизованных структур, но не их специфику, системную семантику процесса. В целом же наше направление и изучение аттракторов - взаимообогащающие. Идеология теории хаоса, аттракторов имеет абстракции и принципы, которые были бы полезны и нам. Мне, например, было бы очень интересно поработать со специалистами по хаотическим процессам, чтобы понять, можно ли использовать этот язык для описания поведения. Впрочем, подход на основе аттракторов отталкивает многих нейрофизиологов. Они считают, что это неадекватный язык описания работы мозга.

Какие еще концепции существуют в изучении мозга - кроме ФС и аттракторов?

- Я бы прежде всего назвал **теорию селекции нейрональных групп** Эделмана, о которой мы уже упоминали. Она предполагает, что **мозг состоит из мини-групп клеток**, которые обладают свойствами распознавания и формируют некий изначальный репертуар - так же, как антитела в иммунологии позволяют распознать каждый потенциальный антиген, даже если в природе его пока не существует - просто за счет того, что их очень много. Эделман считает, что из мини-групп образуются системы, позволяющие мозгу организовывать процессы восприятия и поведения на основании отбора, узнавания нужных мини-групп и замыкания связей между ними. По сути, в результате этого в мозге появляются те же ФС. Подход у Эделмана другой, но результаты очень близки.

Во всем мире сейчас идут споры о том, что такое сознание и какова его нейрофизиологическая основа. Самая модная тема связана с работой конца 1980-х годов, сделанной в Институте мозга им. Макса Планка во Франкфурте, которым руководит Вольф Зингер (Wolf Singer). Зингер с сотрудниками установили, что **когда происходят процессы восприятия, то клетки в разных областях головного мозга синхронизуются, и в них возникают осцилляции на частоте 40 Гц**. Сразу появились теории, что нейрофизиологической основой сознания являются такие группы клеток, а сама осцилляция является моментом осознания процесса.

В чем вы видите прогресс в понимании этой проблематики? Ведь все это обдумывалось веками. Понято сегодня нечто новое о том, что такое сознание, разум - благодаря тому, что мы можем использовать, например, позитронную томографию? Вы говорите - сознание связано с синхронной осцилляцией определенных клеток. Если мы синхронно возбудим эти клетки у человека, он что-нибудь осознает?
- Допустим, я искусственно приведу в возбуждение несколько тысяч определенных элементов мозга, сделаю это по определенным правилам (связь, осцилляция на определенной частоте и т.п.), и в результате этого возбуждения я получу у вас заданное ощущение. Если я подсмотрел, как вы «чувствовали красное», а потом смогу заставить вас таким способом чувствовать красное, на что бы вы ни смотрели - вас это убедит, что мы нашли некие критические условия работы мозга, создающие субъективные ощущения?

Компьютерная реконструкция лица человека.
Так же, как методы функционального магнитно-резонансного нейрокартирования дают возможность изучать активность человеческого

Такие опыты проводились?
- Нет. Но я вам предлагаю это как мысленный эксперимент. Сочли бы вы это важным шагом в понимании того, как мозг создает нашу психику?

Конечно, это был бы важный шаг - но опять-таки лишь первый шаг.
- А какой следующий?

Не знаю, это трудный вопрос. То, что вы предлагаете, напоминает картирование ФС. Сейчас вы скартировали обучение простейшим актам движения. Допустим, удастся скартировать системы, реализующие какие-то абстрактные функции - например, как я решаю математическую задачу, доказываю теорему о средней линии треугольника. Установлено, что при этом сработали определенные нейроны. Потом надо доказать теорему Пифагора. Вопрос: какие нейроны надо включить? Те же или другие?
- То есть вам было бы интересно понять, как это происходит в мозгу и какие системы за это отвечают?

Конечно.
- Мне здесь кажется критичным вот что. Если при рассматривании одной картинки включаются одни нейроны, при рассматривании другой - другие, то по самому факту существования этих двух карт мы не поймем, как работает система. Нам нужно понять, почему возникли эти системы и почему они приводят к решению этих задач. Нам требуется не просто перечисление клеток, а содержательное описание образующихся из них функциональных систем. Нам нужен язык работы с этими системными процессами. Поэтому я бы скептически отнесся к такой постановке задачи: давайте прокартируем и то и другое. Это бессмысленно, если нет идей и инструментов, позволяющих это **перенести на алгоритмический уровень**.

Может ли быть, что ответ на вопрос «что такое разум» мы получим от вас, нейрофизиологов - а не от философов? Вы скажете: разум - это то, что происходит в таких-то нейронах при таких-то условиях…
- Я бы принял здесь точку зрения Алана Тьюринга. Помните, свою знаменитую работу «Вычислительные устройства и интеллект» он начал со слов: «Я предлагаю рассмотреть вопрос, могут ли машины мыслить?» А дальше он пишет, что это плохой вопрос, потому что он может привести только к бесплодным дискуссиям вокруг определений. Нужно добиваться прогресса в эмпирическом понимании явления, говорит он, и вводит затем свой классический тест.[7] Так вот, я считаю, что вопрос о том, что такое сознание и разум, еще очень долго будет оставаться философским и зависящим от определений. Как говорил Витгенштейн, философские проблемы - это проблемы языка. Разные определения, которые возникли вокруг разума на протяжении веков, могут отражать совершенно разные аспекты работы мозга. А мы объединяем это все под одним термином. Мы сможем понять сознание, только когда скажем - черт с вами, будем считать, что все эти смыслы равнозначны и легитимны, а мы для каждого из них подберем соответствующий процесс в мозге. Не надо искать единый нейрофизиологический ответ на все символические смыслы сознания.

Вернемся к тому, с чего мы начали наш разговор - как же быть с моделированием всего этого? Получается, что если у нас есть множество элементов, которые могут синхронно колебаться, - и это все, что нужно для моделирования сознания, пусть даже не сознания, хотя бы простых функций поведения - что же нам мешает все это смоделировать? Взять нейронную сеть улитки - при современных компьютерах это никакого труда не составляет, там всего-то несколько тысяч

мозга, новые методы компьютерного анализа изображений позволяют автоматически распознавать и анализировать черты лица и мимику человека (из: Arthur W. Toga, Neuroimage databases: the good, the bad and the ugly, Nature reviews Neuroscience, 2002, V 3, 302).

нейронов…
- Этого будет недостаточно. Вы же должны создать все поведенческие петли, которые будут включать нейроны со специализацией самого разного свойства. Допустим теперь, что вы можете создать сеть со свойствами нейронов гиппокампа, который хранит пространственную память, нейронов хвостатого ядра, которое связывает десятки структур в планировании действий, что вы сможете наделить эту сеть поведением, которое в глазах наблюдателя будет выглядеть как целенаправленное и адаптивное. Проблема в том, что после этого вы не будете точно знать, думает эта сеть или нет. Ведь при моделировании явно неразумных объектов иногда появляются действия, которые нами, людьми, интерпретируются как осмысленные. Например группа Терзопулоса (Demetri Terzopoulos) из Университета Торонто, когда работала над компьютерной анимацией подводного мира, моделировала нервные системы своих объектов. Вместе с рыбами и другими животными они создали еще и подводного человека, «водяного» (Merman). Этот объект постепенно учился «мыслить» - вырабатывать сложные стратегии поведения, которые при программировании в него не закладывались: например, он научился занимать позицию за препятствием так, чтобы его не видела акула (а на ранних стадиях обучения она его всегда пожирала). Наблюдателем это расценивается как разумное поведение. Теперь вопрос: если появится существо, которое прилетит на летающей тарелке и будет обладать такими свойствами поведения - вы будете считать его разумным?

По-моему, разумность - это нечто такое, что все время ускользает: как только дано определение, она от него уворачивается. Я о другом: будет ли практический смысл у модели, например, мозга улитки. Может быть, логичнее ставить задачу так: дана функция, как ее выполнить? Для этого не обязательно имитировать то, что есть в наших мозгах. Или для медицины, биологии есть смысл в моделях реально существующих когнитивных нейронов?
- Я думаю, что смысл есть, но кроме того, меня ведь интересует, как устроен настоящий мозг. Причем ответ должен иметь форму теории, которая может быть смоделирована. Моделирование для меня - очень эффективный инструмент убивания собственных гипотез. Иногда на самых первых этапах моделирования уже становится ясно, что так быть не может.

То есть тут большое поле деятельности для компьютерных наук?
- Безусловно. В свое время когнитивная наука возникла из того, что исследователи в области компьютерных наук, психологии, философии, лингвистики и нейронаук поняли, что они задают много одинаковых вопросов о природе человеческого разума и используют комплементарные методы для их решения. Но за прошедшие после этого несколько десятилетий в компьютерных дисциплинах и науках о мозге произошли грандиозные изменения. Возьмите хотя бы разрабатываемый нами метод визуализации когнитивных процессов у мышей, в геном которых пересажен специально сконструированный «репортерный» ген, кодирующий молекулу GFP (Green Fluorescent Protein). Этот ген заставляет нервные клетки у таких трансгенных животных флюоресцировать каждый раз, когда они запоминают новую информацию, и по этому свечению мы можем судить, как работают функциональные системы в мозге при том или ином поведении.

Такие возможности исследований просто не могли предвидеть сорок лет назад. Поэтому я убежден, что сегодня наступило время нового синтеза на стыке компьютерных наук и наук о мозге. У нас в стране я вижу уникальные условия для этого, и недавно мы начали три таких новых «синтетических проекта»: «Язык поведения» - соединение методов видеорегистрации и распознавания изображений для объективного анализа поведения; «Открытый мозг» - синтез методов молекулярной биологии и компьютерной трехмерной графики для визуализации процессов в мозге при поведении; и «Мозг анимата» - моделирование поведения аниматов исходя из

принципов работы ФС в мозге животных. Кроме того, мы стали проводить регулярный междисциплинарный семинар «МОЗГ» (http://www.neurogene.ru).

Первую экспериментальную концепцию работы мозга выдвинул в начале XX века Иван Петрович Павлов. Он предположил, что единицами работы мозга являются рефлексы, организованные по четким физиологическим механизмам поступательного распространения нервного возбуждения от рецепторов к рабочим органам. «Основным исходным понятием у нас, - писал Павлов, - является декартовское понятие, понятие рефлекса.[8] Конечно, оно вполне научно, так как явление, им обозначаемое, строго детерминизируется». Однако, выдвинув эту ясную гипотезу, Павлов вскоре столкнулся с фактами, вступившими с ней в драматическое противоречие.[9]

Одна из главных проблем состояла в том, что концепция рефлекса не оставляла места для объяснения свойств предвидения, целеполагания, мышления и психики у животных. Декарт, разделив весь мир на res extensa (протяженную субстанцию) и res cogitans (мыслящую субстанцию), наделил последней лишь человека. Животные же, по мнению Декарта, обладали только машинными рефлексами: «…машины, имеющие органы и внешний вид обезьяны или другого неразумного существа… вполне той же природы, как эти животные».

Пойманный в тиски этой дуалистической логики рефлекторной теории Павлов, которого по его собственному признанию интересовало «в жизни только одно: наше психическое содержание», был вынужден категорически отвергать разговоры и исследовательские программы, связанные с психикой животных. Вместо этого он требовал: «Надо показывать пальцем: где было раздражение, куда оно перешло? Если вы живо себе это представите, то поймете всю силу и правду учения об условных рефлексах, которое совершенно исключило из своего круга психологические понятия, а все время имеет дело только с объективными фактами, то есть с фактами, существующими во времени и в пространстве». Однако постепенно стало совершенно ясно, что такой подход, где психика выступала как ненужный феномен, ввергает рефлекторную гипотезу в глубокий кризис.[10] Психическое в этой концепции было своеобразным дымом из трубы, который сопровождает процессы в мозгу, но причинно на них не влияет, потому что вся реальная детерминация поведения разворачивается на физиологическом уровне возбуждений и торможений нервных клеток.

Но психологи, наоборот, считали, что психика играет детерминирующую роль в поведении. Принимая решение, вы своей волей запускаете то или иное поведение. Кроме того, мы видели, что животные обладают сложными психологическими функциями и это было ясно исследователям уже в первой половине XX века. Например американский психолог Эдвард Толман (Edward Tolman) в 1930-е годы ввел представление о когнитивных картах у животных. Он изучал поведение крыс при поиске пищи в лабиринтах, и показал, что для его объяснения недостаточно простых механических рефлексов, определяющих повороты по рукавам лабиринта. Масса данных говорила о том, что животное, исследуя новое пространство, формирует некую «мысленную» когнитивную карту этого пространства, и когда принимает решение для следующего поведения, извлекает из этой карты образы, цели и задачи. Все это требовало включить психику в картину работы мозга. Нужен был подход, который позволил бы изучать и мозг и поведение и не терять определенный уровень самоорганизации, связанный с этим. Такой подход был предложен несколькими авторами и выделился в 1970-е годы в отдельное направление - когнитивные нейронауки (cognitive neuroscience). В его основе - работы Хебба (Donald Hebb), Саймона (Herbert Simon), Миллера (George Miller), Сперри (Roger Sperry) многих других психологов и нейробиологов. Этот подход вводил представление об эмоциях,

памяти, целеполагании в язык понимания и изучения функционирования мозга.

Однако в отечественной науке то, что сегодня называется когнитивной нейронаукой, возникло гораздо раньше, чем на Западе. У нас уже в 1930-е годы существовали школа Бериташвили, знаменитого грузинского физиолога, который изучал поведение своими методами, вне русла павловской системы, и школа Анохина.[11] И когда на Западе только начала возникать когнитивная наука, у нас в 1950 году уже прошла та самая Объединенная Павловская сессия (Академии наук и Академии медицинских наук), на которой разгромили школы Бериташвили и Анохина. Обоих ученых сняли со всех должностей и закрыли их исследования за отклонение от традиционных павловских концепций.[12]