ФИЛОСОФИЯ ТЕХНИКИ

Е.В. Смирнов

КИТАЙСКАЯ КОМНАТА, КИТАЙСКИЙ РОБОТ И "СИСТЕМНЫЙ ОТВЕТ": ПЕРСПЕКТИВЫ "СИЛЬНОГО" ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация: *в данной статье рассматриваются и анализируются две основных стратегии возражения по известному мысленному эксперименту, получившему название «аргумент китайской комнаты»: так называемое возражение «от робота» и системное возражение. Проведенный анализ позволяет оценить правомерность критики противников аргумента, занимающих данные позиций, а также положений, защищаемых в рамках «аргумента китайской комнаты». В завершение статьи приводится мысленный эксперимент, модифицирующий исходный аргумент для демонстрации ошибочности ряда содержащихся в нем выводов, в частности для опровержения имплицитно подразумеваемой эквивалентности китайской комнаты и находящегося в ней субъекта.*

Ключевые слова: *философия, аргумент китайской комнаты, искусственный интеллект, сознание, функционализм, мышление, алгоритм, коннективизм, тест Тьюринга, программа.*

С момента публикации статьи «Minds, Brains and Programs»[1] Дж. Сёрла, в которой он изложил ставший впоследствии известным не только в среде философов, но и далеко за её пределами, «аргумент китайской комнаты», прошла почти треть века. В течение всего этого времени и сам аргумент и некоторые отдельные положения статьи подвергались разносторонней критике. И хотя интерес к данному мысленному эксперименту американского философа, сохранившийся по сей день со стороны его коллег, свидетельствует о том, что несмотря на многочисленные атаки со стороны оппонентов, ему удалось устоять, тот факт, что большинство упоминаний данного аргумента в работах сопровождается его критикой, ставит под сомнение истинность выводов, полученных в содержащей его работе. В данной статье мы рассмотрим и проанализируем позиции отдельных философов, представляющие наиболее интересные, на наш взгляд, линии возражения по «аргументу китайской комнаты». Это поможет нам не только более точно сформулировать собственную позицию по вопросу «аргумента китайской комнаты», но и показать какое место можно отвести ей среди множества других возражений данному аргументу.

Как отмечает сам Дж. Сёрл, целью его аргумента является критика когнитивистской парадигмы искусственного интеллета (ИИ), которая выражается в притязаниях последнего на то, что машины, обладающие соответствующим программным обеспечением способны в буквальном смысле понимать, а также их программы способны объяснять ментальные психические процессы. Для обозначения ИИ, выраженного в двух этих притязаниях, он вводит понятие «сильный» ИИ в отличие от так называемого «слабого» ИИ, который подразумевает взгляд на компьютерную программу как лишь на очень мощный инструмент, помогающий в исследовании сознания. Данные термины Дж. Сёрл вводит в связи с рассмотрением им работ Роджера Шэнка и его коллег — создателей, так называемых машин Тьюринга, предназначенных для прохождения теста на способность успешно эмулировать речевое поведение человека. Такие машины принимают на входе рассказы, содержащие некоторую информацию на заданную тему, и, затем, на основе содержащихся в них «представлений» по данной теме и в соответствии с программой выдают ответы на вопросы, задаваемые по этой теме. Программа считается успешно прошедшей тест Тьюринга,

[1] Searle, J. Minds, Brains and Programs / J. Searle // Behavioral and Brain Sciences. 1980. № 3 (3). P. 417-457.

если её ответы на вопросы неотличимы с точки зрения человека-эксперта от ответов человека. При этом успешность машин при прохождении теста Тьюринга, по мнению создателей программ, позволяет утверждать, что машины буквальным образом понимают рассказы и дают ответы на вопросы, а также что то, что делают машины, объясняет человеческую способность понимать рассказы и отвечать на вопросы.

Чтобы показать ошибочность таких суждений Дж. Сёрл предлагает мысленный эксперимент, заключающийся в следующем. Человек, знающий английский язык, но не понимающий по-китайски, заперт в комнате и не имеет перцептивного контакта с внешним миром. Он получает тексты на китайском языке (вопросы), и, совершая над символами текста формальные операции в соответствии с некоторым алгоритмом, выдаёт преобразованные последовательности символов на китайском языке. Таким образом, если используемая программа настолько хороша, что позволяет человеку в комнате выдавать такие последовательности китайских символов, которые будут интерпретироваться носителями китайского языка как осмысленные ответы на вопросы, то для внешних наблюдателей, общающихся с испытуемым, будет казаться, что он понимает по-китайски. Однако, как утверждает Сёрл, данный мысленный эксперимент показывает, что неспособность наблюдателей отличить ответы человека от ответов компьютера не означает, что последний что-то действительно понимает. Более того, Дж. Сёрл утверждает, что данный эксперимент показывает неправомерность второго притязания «сильного» ИИ, то есть того, что данные программы способны объяснять механизмы человеческого понимания и познания. Итак, попробуем разобраться в описаной выше ситуации. Безусловно, данный аргумент, апеллирующий, прежде всего, к интуиции, представляет вымышленное положение вещей таким образом, чтобы выводы, получаемые его автором, казались очевидными. Однако прежде чем выносить какое-либо решение по вопросу «аргумента китайской комнаты», мы должны рассмотреть, какие возражения, более других заслуживающие внимания, предпринимались оппонентами Дж. Сёрла в разное время.

Нужно отметить, что все основные формы возражений против собственного аргумента Дж. Сёрл, предвидя наперёд, рассмотрел и прокомментировал в той же статье, что содержит и сам аргумент. Он приводит шесть видов возражений, направленных против предложенного им аргумента, однако в данной статье мы остановимся лишь на двух, представляющих на наш взгляд наибольший интерес для исследования и опасность для «аргумента китайской комнаты». В частности мы опустим рассмотрение так называемых возражений «от других сознаний» и «от нескольких обиталищ» и др., поскольку нас устраивают аргументы, приведённые в отношении их Дж. Сёрлом. Одним из представляющих на наш взгляд интерес возражений «аргументу китайской комнаты» является так называемый «ответ от робота». Дж. Сёрл излагает его следующим образом: «*Предположим, мы написали программу, отличную от программы Шэнка. Предположим, мы поместили компьютер внутрь некоего робота, и этот компьютер не просто воспринимал бы формальные символы на входе и выдавал бы формальные символы на выходе, а на самом деле руководил бы роботом так, что тот делал бы нечто очень похожее на сенсорное восприятие… Такой робот, в отличие от компьютера Шэнка, обладал бы настоящим пониманием и другими ментальными состояниями*»[2]. По мнению Дж. Сёрла, возражения, содержащиеся в данном «ответе» не играют значимой роли, поскольку поступающую извне информацию можно таким же образом представить в виде последовательности неинтерпретированных символов. Однако такая аргументация, на наш взгляд, способна отвести лишь часть возражений, направленных против «аргумента китайской комнаты», выдержанных в данном ключе. Здесь мы остановимся и подробнее проанализируем вариант возражения «от робота», принадлежащий В.В. Васильеву. Его позиция по вопросу Аргумента китайской комнаты изложена в статье «*Кока-кола и секрет китайской комнаты*»[3], а также, в дополненном и уточнённом виде — в книге «*Трудная проблема сознания*»[4].

По мнению В. В. Васильева «*…китайская комната при её адекватном толковании, доказывает прямо противоположное тому, что хотел сказать с её помощью Сёрл. Напомним,*

[2] Ibid.

[3] Васильев В.В. Кока-кола и секрет Китайской комнаты / В.В. Васильев // Философия сознания: классика и современность. М., 2007. С. 86-94.

[4] Васильев В.В. Трудная проблема сознания / В.В. Васильев. М.: Прогресс-Традиция, 2009. 272 с.

что он собирался продемонстрировать ошибочность концепции «сильного искусственного интеллекта»... При этом он не отрицал возможности механического эмулирования поведения, то есть «слабого искусственного интеллекта»... мы, однако, приходим к выводу, что она помогает уяснить нереальность проектов подобного эмулирования, — и соответственно заключить, что если программирование разумного поведения возможно, то реализующая такую программу система с необходимостью будет обладать сознательными состояниями»[5]. Из приведённого выше отрывка видно, что позиция В.В. Васильева по вопросу возможности «слабого» ИИ сводится к тому, что 1) — «слабый» ИИ невозможен, поскольку 2) — в случае создания системы, эмулирующей поведение человека, она с необходимостью попадёт в разряд представителей «сильного» ИИ. Рассмотрим аргументацию В.В. Васильева, приводящую к выводу, содержащемуся во втором тезисе.

Ключевым моментом в правильном понимании «аргумента китайской комнаты» является разделение «обычного» варианта с комнатой и варианта с роботом. Существенным отличием между ними, по мнению В.В. Васильева являются ограниченные эмулятивные способности комнаты, в частности неспособность отвечать на индексикальные вопросы: «*Допустим, я знаю китайский и веду диалог с Сёрлом, запертым в его Комнате вместе с этой программой и целыми горами иероглифов. Допустим также, что перед экспериментом я зашел в комнату и поставил банку кока-колы на стол, за которым будет сидеть Сёрл. И вот мы начинаем наш диалог. Я спрашиваю Сёрла по-китайски: «Скажите, что находится прямо перед Вами?»*[6] *По идее, он должен обратиться к программе, в которой должен содержаться ответ: «банка кока-колы». Но как мог человек, составивший его программу, знать, что я принесу в Китайскую комнату именно этот предмет? Такую программу могло бы написать только всезнающее существо...*»[7]. В то же время робот, обладающий сенсорно-моторными способностями, заведомо не имеет таких ограничений. Вследствие этого, наложенные таким образом ограничения на «китайскую комнату», по мнению В. В. Васильева, делают логически невозможным создание на её основе системы, эмулирующей человеческое поведение, в то время как в отношении робота такая возможность сохраняется. Таким образом, в отношении робота выполняются необходимые условия для создания как «сильного», так и «слабого» искусственного интеллекта, представляющие собой требование возможности эмуляции человеческого поведения. Дальнейший ход рассуждений, приводящий к выполнению достаточных условий «сильного» и исключению возможности «слабого» искусственного интеллекта основывается на аргументе каузальных траекторий, рассмотрение которого в рамках данной статьи не входит в наши планы. Однако, на наш взгляд, различия между комнатой и роботом в действительности не столь значительны, как считает В.В. Васильев, а акцентуация их неэквивалентности не имеет значимых последствий для аргумента.

Основным преимуществом робота перед комнатой, по мнению В.В. Васильева, является возможность автономно ориентироваться в окружающем мире благодаря своим сенсорно-моторным способностям, то есть, иными словами, такой робот может проходить «*тотальный тест Тьюринга (подразумевающий не только языковую, но сенсорно-моторную компетентность)*»[8]. Такая способность подразумевает наличие квазисемантического блока, выполняющего функцию установки соответствия между символами и образами внешнего мира и внутренней информацией. По условиям аргумента эта часть робота оказывается за пределами функциональных обязательств помещённого в него человека, который, судя по всему, получает в качестве входных данных выходные данные квазисемантического блока. Таким образом, содержание блока недоступно человеку, находящемуся внутри китайского робота, соответственно, по мнению В.В. Васильева «*он не понимает задаваемые ему вопросы. Но можно предположить, что их понимает весь робот, одним из механизмов которого оказывается человек внутри него*»[9]. Таким образом, он чётко определяет место человека в системе,

[5] Васильев В.В. Кока-кола и секрет Китайской комнаты / В.В. Васильев // Философия сознания: классика и современность. М., 2007. С. 94.

[6] Другим примером вопроса, на который, по мнению В.В. Васильева, не способна ответить комната является вопрос «Сколько сейчас времени?».

[7] Там же. С. 88.

[8] Там же. С. 90.

[9] Васильев В. В. Трудная проблема сознания / В.В. Васильев. М.: Прогресс-Традиция, 2009. С. 89.

представляющей собой робота. Каково же будет место человека в традиционной форме «аргумента китайской комнаты»?

По условиям мысленного эксперимента человек помещён внутрь комнаты и является исполнителем совершенной программы, позволяющей китайской комнате демонстрировать удивительную для наблюдателя лингвистическую компетентность. Таким образом, в условиях эксперимента изначально заложено ограничение на демонстрацию собственной антропоморфности. Очевидно, что этот момент, обусловленный стремлением Дж. Сёрла создать мысленный аргумент, отвергающий тест Тьюринга в качестве критерия понимания, делает такую систему подобной не собственно человеку, а, например, человеку, общающемуся с собеседником через средства электронной связи. Более того, поскольку поступающая внутрь комнаты информация представляет собой исключительно текст на китайском языке, то сама такая «разумная» комната похожа на человека, получающего информацию извне также, исключительно в виде текста[10]. И в таких условиях приводимые В.В. Васильевым в качестве примера индексикальные вопросы «Сколько сейчас времени?» и «Что находится перед вами?» (в случае варианта с банкой кока-колы) уже не представляют угрозы для китайской комнаты. Ответом на первый вопрос может быть указание на обстоятельства, в которых находится собеседник, то есть ответ в духе: «В комнате, которой я нахожусь очень темно, нет окон и часов, поэтому я не знаю текущего времени». Что касается второго вопроса, то он вообще оказывается неправомерным: человек, исполняющий программу, перед которым поставлена банка кока-колы, неэквивалентен комнате, которой задаётся вопрос[11]. Следует отметить, что аналогичной позиции по вопросу правомерности модификации аргумента китайской комнаты с использованием банки кока-колы, придерживается в недавно вышедшей монографии «*Бостонский зомби: Д. Деннет и его теория сознания*»[12] коллега В.В. Васильева по центру исследования сознания Д.Б. Волков: «*Васильев проводит действительно оригинальную атаку на аргумент Сёрла. Но, мне кажется, можно её отбить. Можно показать, что этот ход незаконный, что он нарушает аналогию между комнатой и рациональным агентом. В самом деле, сколько нейронов у вас в голове? … А как выглядит орган, соединяющий ваше правое и левое полушарие? А вдруг вместо него там находится банка с кока-колой? … Неспособность агента дать ответы о своём внутреннем устройстве и внутренних изменениях не дают основания полагать, что агент не рационален*»[13]. Несмотря на это, он всё же полагает, что линия возражения, основанная на неэквивалентности комнаты роботу перспективна, и пытается отыскать вопрос, демонстрирующий эмулятивную ограниченность комнаты. Справедливо замечая, что вопрос о времени (например, «который час?»), неправомерен, Д.Б. Волков обращается к внутреннему ощущению времени системы, то есть предлагает модифицировать вопрос, предлагая системе сравнить два интервала времени. Поскольку «*чтобы казаться рациональным он (предполагаемый узник комнаты — прим. автора) должен уметь отличить 10 минут от 10 дней*»[14], комната, претендующая на прохождение теста Тьюринга должна обладать инструкциями, позволяющими сравнивать временные интервалы, однако, по мнению автора «*… нельзя написать инструкции так, чтобы предусмотреть правильный ответ*»[15]. На наш взгляд, такая попытка «разоблачения» комнаты, обращённая к внутреннему «чувству» времени, также не достигает своей цели. Условия мысленного эксперимента без ущерба для его содержания можно трансформировать так, чтобы комната справлялась с таким вопросом. Например, можно вменить в обязанности человеку, находящемуся в комнате систематически совершать какое-либо контрольное действие, делая определённые пометки в своих книгах по его завершении (в соответствии с предусмотренными для такого случая инструкциями). Программный модуль, реализующий в

[10] Вероятно, непросто представить такую ситуацию: даже будучи в звуконепроницаемых наушниках и находясь в абсолютно неосвещённой комнате, человек сохранит тактильное восприятие. Это, однако, не имеет решающего значения для нашей линии возражения — ответы на вопросы, связанные с воспринимаемой информацией и направленные на «разоблачение» собеседника в условиях их удалённости друг от друга могут быть в достаточной мере произвольны, сохраняя при этом правдоподобность.

[11] Этот вопрос более подробно будет рассмотрен далее, при анализе так называемого «системного ответа».

[12] Волков Д.Б. Бостонский зомби: Д. Деннет и его теория сознания / Д.Б. Волков. М.: Книжный дом «ЛИБРОКОМ», 2012. 320 с. (Философия сознания).

[13] Там же. С. 87.

[14] Там же. С. 88.

[15] Там же. С. 88.

такой системе внутреннее «чувство» времени может быть написан исходя из прямой зависимости между количеством таких контрольных действий и истёкшим за период их выполнения временем. В нём же, с помощью нехитрого алгоритма, может быть достигнута необходимая степень огрубления «восприятия» времени системой, позволяющая ей отличать интервалы времени, значительно отличающиеся по величине и вынуждающая для правдоподобия ошибаться при сопоставлении близких по величине интервалов. При необходимости ответить на вопрос, подобный тому, который приводит Д.Б. Волков, человек в такой комнате, в соответствии с инструкциями обратится к сделанным ранее пометкам и без труда выдаст подходящий ответ, не понимая как содержание вопроса, так и ответа, то есть, не нарушая условия эксперимента. Таким образом, на наш взгляд, линия возражения аргументу Дж. Сёрла, основывающаяся на демонстрации неэквивалентности комнаты роботу (вследствие эмулятивной ограниченности) недостаточно обоснована, поскольку её сторонникам не удаётся отыскать пример вопроса, с очевидностью ставящий такую комнату в тупик. Поэтому мы полагаем, что при правильном учёте специфики условий мысленного эксперимента ограничения, накладываемые на эмулятивные способности китайской комнаты, в виде неспособности давать ответы на индексикальные вопросы, снимаются.

Рассмотрев один из вариантов критики «аргумента китайской комнаты», относящийся в классификации Дж. Сёрла к «ответам от робота» мы пришли к заключению, что эмулятивные возможности комнаты не уступают таковым в её роботизированном варианте, а значит, возражения «от робота» не представляют для него серъёзной опасности. В дополнение к сказанному выше мы считаем необходимым коснуться ещё одного вопроса, тесно связанного с данной линией возражения мысленному эксперименту Дж. Сёрла. Речь идёт о соотношении синтаксиса и семантики применительно к интеллектуальным система, рассматриваемым в нём и, в частности, к вышеописанному роботу. Одним из аргументов Дж. Сёрла является указание на то, что «*понимание семантично, в то время как компьютерная программа исключительно синтаксична*»[16]. Как в приведённом выше возражении В.В. Васильева, так и в возражениях других философов, использующих вариант роботизированной китайской комнаты для возражения аргументу Дж. Сёрлу, одним из основных пунктов является указание на то, что программа такого робота *может* быть семантичной.

В возражениях В.В. Васильева под «семантической» частью программы понимается так называемый квазисемантический блок, определяющий «*отношения иероглифов к определённым неиероглифическим, физическим данным, получаемым его телекамерой…*»[17]. Эта характерная черта робота, по мнению сторонников этой позиции, наделяет инсталлированную в него программу семантикой, что возможно, позволяет такому роботу понимать информацию из внешнего мира. В то же время Дж. Сёрл видит ту же ситуацию с роботом совершенно иначе: «*…вы даете мне еще больше китайских символов с еще большим количеством инструкций на английском языке насчет того, как сопоставлять одни китайские символы с другими и выдавать китайские символы вовне. Предположим далее, что некоторые китайские символы, приходящие ко мне от телекамеры, встроенной в робота, и другие китайские символы, которые выдаю я, служат для того, чтобы включать моторы, встроенные в робота, так чтобы двигались ноги и руки робота, но я ничего этого не знаю. Важно подчеркнуть, что я делаю только одно — манипулирую формальными символами: я не знаю никаких дополнительных фактов*»[18].

На наш взгляд, часть программы, ставящая в соответствие физическим данным, получаемым из внешнего мира определённые символы, всё же не может считаться семантической лишь в силу этих её способностей. Поскольку под установлением семантических связей подразумевается раскрытие значения символов, то такое установление должно подразумевать субъекта, интерпретирующего данное значение. Иными словами, «семантическая» часть программы может представлять собой лишь алгоритм распознания или перевода одних символов в другие и не может претендовать на подлинную семантичность, подразумевающую установление значения символов. Поэтому позиция Дж. Сёрла в данном вопросе кажется нам более обоснованной.

[16] Searle, J. Minds, Brains and Programs / J. Searle // Behavioral and Brain Sciences. 1980. № 3 (3). P. 417-457.

[17] Васильев В.В. Кока-кола и секрет Китайской комнаты / В.В. Васильев // Философия сознания: классика и современность. М., 2007. С. 89.

[18] Searle, J. Minds, Brains and Programs / J. Searle // Behavioral and Brain Sciences. 1980. № 3 (3). P. 417-457.

Вторым, наиболее часто встречающимся возражением по «аргументу китайской комнаты», независимо от возможных его модификаций, является справедливое на наш взгляд указание на неэквивалентность человека в комнате субъекту возможного понимания. Поэтому, большая часть критики данного мысленного эксперимента, так или иначе, сводятся к «системному ответу», то есть указанию на то, что из отсутствия понимания у человека в китайской комнате не следует отсутствие понимания у всей системы в целом. Известные нам модификации такой линии возражения, как правило, не подкрепляются серьёзной аргументацией и выглядят недостаточно демонстративно. На наш взгляд, возражение, основанное на «системном ответе», можно сделать более демонстративным, видоизменив мысленный эксперимент с китайской комнатой соответствующим образом. При этом материал для трансформации «аргумента китайской комнаты» мы позаимствуем из другого мысленного эксперимента, принадлежащего одному из наиболее последовательных его критиков — Д. Деннету. Речь идёт о мысленном эксперименте «Где Я?», изложенном в сборнике философских статей «*Глаз разума*»[19].

В оригинальном мысленном эксперименте мозг героя в целях безопасности был отделён от его тела, посылаемого на специальное задание в урановую шахту, и помещён в специальную колбу. Связь мозга с периферийной нервной системой тела осуществлялась с использованием комбинации радиопередатчиков и приёмников. Более того, для надёжности, учёные создают точную функциональную копию мозга героя, способную в ответ на входные импульсы от тела генерировать выходные импульсы, идентичные тем, что формирует реальный мозг. Сам эксперимент представляет собой описание различных вариантов сочетаний «искусственный / естественный мозг — тело», однако нас интересуют лишь исходные условия эксперимента, изложенные выше. Итак, предположим, что в нашем случае, героя деннетовского эксперимента отправляют с миссией на Марс. Снова, чтобы не подвергать чрезмерной опасности, повсеместно угрожающей на чужой планете, его мозг отделяют от тела и помещают в хорошо защищённое место на космическом корабле. Как и в оригинальной версии эксперимента, мозг получает сигналы от сенсоров тела, а также передаёт сигналы обратно посредством радиосвязи. Поскольку его мозг периодически нуждается в отдыхе, для его замещения, как и в оригинальной версии эксперимента, учёные создают цифровой дубликат, получающий на входе и генерирующий на выходе тот же набор сигналов, что и реальный мозг. Чтобы снизить риск провала дорогостоящей программы, учёные предусматривают ручной режим работы электронного дубликата мозга. Для этих целей они нанимают человека из китайской комнаты, обладающего сверхспособностями в отношении быстроты выполнения алгоритмов, и обучают алгоритмам формирования выходных импульсов в ответ на входные. Иными словами, его задачей является формировние выходных импульсов в ответ на входные в соответствии с алгоритмами, реализуемыми каузальными структурами мозга (то есть, в основу алгоритма его работы положен коннекционистский подход). Как и с инструкциями в комнате, этот человек достигает такого мастерства, что становится способен работать в ручном режиме столь же быстро, как и реальный мозг. Таким образом, команда корабля состоит из двух человек: из исполнителя миссии, чьи мозг и тело разделены в пространстве и специалиста, находящегося в комнате, целью которого является ручное управление телом в случае отключения передатчиков реального и электронного мозга. Обозначим первого из них — А, а второго — В. Очевидно, что оба специалиста являются действующими субъектами выполнения операции. Однако, помимо них на борту корабля будет присутствовать субъект операции С, представляющий собой систему «комната для управления телом — тело» с помещённым в неё специалистом В. Поскольку такая система по условиям эксперимента способна к автономному существованию и решению поставленных задач, то правомерность притязаний на то, чтобы считать её субъектом операции не должна вызывать сомнений. В даном мысленном эксперименте мы постараемся показать, что С и В в действительности представляют собой два различных субъекта деятельности.

Итак, по достижении пункта назначения худшие опасения учёных оправдываются: электронный мозг выходит из строя и замещение реального мозга приходится проводить в ручном режиме. В заданный момент времени специалисты приступают к стадии переключения: для этого входные и выходные сигналы мозга и электронной комнаты должны стабильно совпадать в течение продолжи-

[19] Хофштадтер Д. Глаз разума / Д. Хофштадтер, Д. Деннет. Самара: Бахрах-М, 2003. 432 с.

тельного промежутка времени. В определённый момент времени такое совпадение достигается и система переключает управление телом на электронную комнату. Человек в ней хорошо подготовлен и успешно справляется с задачей. Но что происходит в момент, предшествующий переключению?

Человек в комнате (В) получает извне точно такие же сигналы, как и мозг хозяина тела (А), генерирует точно такие же импульсы на выходе, оперируя при этом, как и в китайской комнате неинтерпретированными символами. Следуя интуициям, на которые наталкивает нас мысленный эксперимент Дж. Сёрла, мы можем утверждать что он, выполняя программу посредством операций над формально определёнными элементами, *не понимает* поступающей ему информации. В то же время его напарник (субъект А), разделённый для безопасности в пространстве, должен сохранить такое понимание. И причиной этого, как и отмечалось выше, является тот факт, что человек В, заключённый в электронной комнате *неэквивалентен* субъекту А. Поскольку система «комната — тело» (субъект С) функционально эквивалентна системе «мозг — тело» (субъекту А), а А неэквивалентен В, следовательно, субъект В оказывается неэквивалентным субъекту С, что и требовалось доказать.

Подводя итоги сказанному выше, следует отметить следующее. Прежде всего, возражения по «аргументу китайской комнаты» «от робота», направленные на демонстрацию эмулятивной ограниченности системы (комнаты), фигурирующей в оригинальной версии мысленного эксперимента не достигают целей. Как было показано, отличия рассмотренных модификаций когнитивных систем, представляющих собой китайскую комнату и робота, сводятся к отличиям между их имплицитно подразумеваемыми прототипами: носителями китайского языка, в первом случае запертым в тёмной комнате и имеющий возможность общения с внешним миром лишь посредством электронных сообщений и, во втором случае, не имеющим таких ограничений. Когнитивно более ограниченная китайская комната, безусловно, уступает в интеллектуальности роботу, однако, вполне удачно может справиться с эмуляцией поведения своего прототипа. Касаясь вопроса, могут ли такие системы обладать по аналогии со своими прототипами квалитативными состояниями (то есть, переходя к анализу «системного возражения») необходимо отметить, что аргумент не даёт на него однозначного отрицательного ответа. Перефразируя Дж. Сёрла, применительно к предложенному в данной статье мысленному эксперименту можно говорить, что В является гомункулом комнаты, то есть субъекта С. При этом информация, получаемая им, не является воспринимаемой непосредственно, в виде квалиа, а является внешним информационным нейронным кодом и требует интерпретации[20]. Таким образом, очевидно, что доказательство отсутствия понимания у человека в исходной версии аргумента не даёт права на выводы о невозможности «сильного» ИИ, а значит, нам удалось показать, что китайская комната не достигает цели, по крайней мере, в опровержении притязания «сильного» ИИ на обладание пониманием. Что касается второго притязания «сильного» ИИ на то, что программы таких систем способны объяснять механизмы человеческого понимания и познания, то в рамках предложенного мысленного эксперимента оно не подтверждается, а значит в отношении него «аргумент китайской комнаты» сохраняет свою силу. Коннекционистский подход, положенный в основу алгоритма работы описываемой в нём системы подразумевает воспроизведение каузальных свойств структур мозга, и поэтому такая программа способна объяснять сознательный процесс лишь на уровне нейронных процессов. Для того чтобы программа могла помочь в понимании феноменологической стороны психических процессов она должна быть реализована в рамках символьного подхода, предполагающего наличие строгой последовательности операций, условных переходов, ветвлений, циклов и т.д. Однако против возможности реализации системы, описываемой в эксперименте, с использованием такого подхода выступает аргумент о неформализуемости поведения, убедительного опровержения которого в настоящее время не существует и, на наш взгляд, маловероятно в будущем.

[20] Для наглядной демонстрации соотношения типов поступающей информации субъектам А и Б, на наш взгляд, очень удачно подходит терминология информационной концепции сознания Д.И. Дубровского. Если субъект А воспринимает так называемую «чистую» информацию, подразумевающую феноменологическое понимание, то В получает «чуждый» информационный код, реализованный в мозговых нейронных модулях.

Список литературы:

1. Васильев В.В. Кока-кола и секрет Китайской комнаты / В.В. Васильев // Философия сознания: классика и современность. М., 2007. С. 86-94.
2. Васильев В.В. Трудная проблема сознания / В.В. Васильев. М.: Прогресс-Традиция, 2009. 272 с.
3. Волков Д.Б. Бостонский зомби: Д. Деннет и его теория сознания / Д.Б. Волков. М.: Книжный дом «ЛИБРОКОМ», 2012. 320 с. (Философия сознания).
4. Дубровский Д.И. Психические явления и мозг: философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики / Д.И. Дубровский. М.: Наука, 1971. 386 с.
5. Дубровский Д.И. Сознание, мозг и искусственный интеллект: сб. статей / Д.И. Дубровский. М.: ИД Стратегия-Центр, 2007. 272 с.
6. Дубровский «Трудная» проблема сознания (в связи с книгой В.В. Васильева «Трудная проблема сознания») / Д. И. Дубровский // Вопросы философии. 2011. № 9. С. 136-148.
7. Хофштадтер Д. Глаз разума / Д. Хофштадтер, Д. Деннет. Самара: Бахрах-М, 2003. 432 с.
8. Searle, J. Minds, Brains and Programs / J. Searle // Behavioral and Brain Sciences. 1980. № 3 (3). P. 417-457.

References (transliteration):

1. Vasil'ev V.V. Koka-kola i sekret Kitayskoy komnaty / V.V. Vasil'ev // Filosofiya soznaniya: klassika i sovremennost'. M., 2007. S. 86-94.
2. Vasil'ev V.V. Trudnaya problema soznaniya / V.V. Vasil'ev. M.: Progress-Traditsiya, 2009. 272 s.
3. Volkov D.B. Bostonskiy zombi: D. Dennet i ego teoriya soznaniya / D.B. Volkov. M.: Knizhnyy dom «LIBROKOM», 2012. 320 s. (Filosofiya soznaniya).
4. Dubrovskiy D.I. Psikhicheskie yavleniya i mozg: filosofskiy analiz problemy v svyazi s nekotorymi aktual'nymi zadachami neyrofiziologii, psikhologii i kibernetiki / D.I. Dubrovskiy. M.: Nauka, 1971. 386 s.
5. Dubrovskiy D.I. Soznanie, mozg i iskusstvennyy intellekt: sb. statey / D.I. Dubrovskiy. M.: ID Strategiya-Tsentr, 2007. 272 s.
6. Dubrovskiy «Trudnaya» problema soznaniya (v svyazi s knigoy V.V. Vasil'eva «Trudnaya problema soznaniya») / D.I. Dubrovskiy // Voprosy filosofii. 2011. № 9. S. 136-148.
7. Khofshtadter D. Glaz razuma / D. Khofshtadter, D. Dennet. Samara: Bakhrakh-M, 2003. 432 s.
8. Searle, J. Minds, Brains and Programs / J. Searle // Behavioral and Brain Sciences. 1980. № 3 (3). S. 417-457.