А. М. Тьюринг (1950) Вычислительная техника и интеллект. Mind 49: 433-460.

ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНТЕЛЛЕКТ

Автор: А. М. Тьюринг

1. Имитирующая игра

Я предлагаю рассмотреть вопрос: «Могут ли машины думать?» Это следует начать с определений значения терминов «машина» и «думать». Определения можно оформить так, чтобы как можно точнее отражать обычное употребление слов, но такая позиция опасна; если смысл слов «машина» и «думать» должен быть установлен путем анализа того, как они обычно используются, трудно избежать заключения, что смысл и ответ на вопрос «Могут ли машины думать?» следует искать в статистическом опросе вроде опроса Галлопа. Но это абсурдно. Вместо попытки такой дефиниции я заменю вопрос другим, который тесно связан с ним и выражен в относительно неразборчивых словах.

Новая форма проблемы может быть описана как игра, которую мы назовем «имитационной игрой». В ней участвуют три лица: мужчина (A), женщина (B) и опрашивающий (C), который может быть любого пола. Опрашивающий остается в комнате отдельно от двух других. Цель игры для опрашивающего — определить, кто из двух остальных является мужчиной, а кто — женщиной. Он знает их под ярлыками X и Y, и в конце игры говорит либо «X — это A, а Y — это B», либо «X — это B, а Y — это A». Опрашивающий имеет право задавать вопросы A и B таким образом:

C: Пожалуйста, скажите, длинна ли его или её волосы у X?

Предположим теперь, что X на самом деле является A, тогда A должен ответить. Цель A в игре — заставить C неправильно идентифицировать. Его ответ может быть следующим:

«Мои волосы уложены, и наиболее длинные пряди примерно девять дюймов в длину.»

Чтобы интонация голоса не помогала опрашивающему, ответы должны быть записаны на письме или, ещё лучше, набраны на машинке. Идеальная схема — телепринтер, соединяющий обе комнаты. Иначе вопросы и ответы может повторить посредник. Целью игры для третьего игрока (B) является помощь опрашивающему. Лучшая стратегия для неё, вероятно, — давать правдивые ответы. Она может добавить такие слова, как «Я — женщина, не слушай его!», но это ничего не даст, поскольку мужчина может сделать аналогичные замечания.

Теперь задаём вопрос: «Что произойдёт, если машина займёт место A в этой игре?» Будет ли опрашивающий ошибаться так же часто, как и при игре между мужчиной и женщиной? Эти вопросы заменяют наш первоначальный вопрос: «Могут ли машины думать?» "

2. Критика новой проблемы

Помимо того, что задают вопрос: «Каков ответ на эту новую форму вопроса?», можно спросить: «А стоит ли исследовать этот новый вопрос?» Мы исследуем этот последний вопрос без промедления, тем самым прерывая бесконечную регрессию.

Новая проблема имеет преимущество в том, что она четко разграничивает физические и интеллектуальные способности человека. Ни один инженер или химик не утверждает, что способен создать материал, неотличимый от человеческой кожи. Возможно, что когда-нибудь такое будет сделано, но даже если бы изобретение было доступно, мы сочли бы малоинформативным пытаться сделать «мышление машины» более человечным, одевая его в такую искусственную плоть. Форма, в которой мы поставили проблему, отражает этот факт в условии, которое препятствует опрашивающему видеть или трогать других конкурентов или слышать их голоса. Другие преимущества предлагаемого критерия могут быть показаны по образцам вопросов и ответов. Таким образом:

Q: Пожалуйста, напишите мне сонет на тему моста Форт (Forth Bridge).

A: От этого дела я отойду. Я никогда не умел писать поэзию.

Q: Сложить 34957 и 70764.

A: (Пауза около 30 секунд, затем ответ) 105621.

Q: Вы умеете играть в шахматы?

A: Да.

Q: У меня король на h1, и больше фигур нет. У вас только король на f6 и ладья на a1. Ваш ход. Что вы играете?

A: (После паузы в 15 секунд) P-P8 мат.

Метод вопросов и ответов, по-видимому, подходит для введения почти любой из областей человеческого дела, которую мы хотим включить. Мы не хотим наказывать машину за её неспособность блистать в конкурсах красоты, и не хотим наказывать человека за проигрыш в гонке с самолётом. Условия нашей игры делают эти недостатки несущественными. «Свидетели» могут похвастаться, если сочтут нужным, чем угодно о своей привлекательности, силе или героизме, но опрашивающий не может требовать практических демонстраций.

Игру, возможно, можно критиковать на основании того, что шансы слишком сильно уложены против машины. Если человек попытается притвориться машиной, он очевидно будет выглядеть очень плохо. Ему выдать всё рано будет медлительностью и неточностью в арифметике. Разве не могут машины выполнять то, что следует назвать мышлением, но что очень отличается от того, что делает человек? Этот возражение

очень сильное аргумент, но по крайней мере можно сказать, что если, несмотря на это, машина может быть сконструирована для удовлетворительной игры в имитационную игру, нас не должно беспокоить это возражение.

Можно было бы утверждать, что при игре в "имитационную игру" наилучшая стратегия для машины может быть чем-то иным, чем имитация поведения человека. Это может быть, но я думаю, что маловероятно, что имеет место какой-то значительный эффект такого рода. В любом случае здесь нет намерения исследовать теорию игры, и будет предположено, что лучшая стратегия — пытаться предоставить ответы, которые естественно дали бы человек.

3. Машины, участвующие в игре

Вопрос, который мы поставили в пункте 1, будет довольно неопределён, пока мы не уточним, что мы имеем в виду под словом «машина». Естественно, что мы хотим позволить использовать в наших машинах любой вид инженерной техники. Мы также хотим позволить возможность того, что инженер или команда инженеров может сконструировать машину, которая работает, но чьё устройство не может быть удовлетворительно описано конструкторами, потому что они применили метод, который в основном экспериментальный. Наконец, мы хотим исключить из машин людей, рождённых обычным способом. Трудно сформулировать определения так, чтобы удовлетворить эти три условия. Например, можно настаивать на том, что команда инженеров должна состоять из лиц одного пола, но это было бы не совсем удовлетворительно, ведь, вероятно, возможно вырастить целостного индивида из одной клетки кожи человека. Сделать это — подлинное достижение биотехнической техники, достойное самых высоких похвал, но мы не склонны рассматривать это как случай «построения думающей машины». Это побуждает нас отказаться от требования, чтобы была допустима любая техника. Мы тем более готовы сделать это, учитывая тот факт, что нынешний интерес к «мыслящим машинам» вызван конкретным видом машины, обычно называемой «электронным компьютером» или «цифровым компьютером». Следуя этому предложению, мы допускаем участие в нашей игре только цифровых компьютеров.

На первый взгляд это ограничение кажется крайне радикальным. Я попытаюсь показать, что на деле это не так. Для этого потребуется короткий рассказ о природе и свойствах этих компьютеров.

Можно также сказать, что это соответствие машин цифровым компьютерам, как и наш критерий «мышления», будет неудовлетворительным лишь в том случае, если (что противоречит моим убеждениям) окажется, что цифровые компьютеры не могут хорошо проявить себя в игре.

Уже есть ряд цифровых компьютеров в рабочем состоянии, и можно спросить: «Почему бы не провести эксперимент сразу? Было бы легко выполнить условия игры. Можно использовать ряд опрашивателей и собрать статистику, чтобы показать, как часто даётся правильная идентификация». Краткий ответ: мы не спрашиваем, хорошо ли справились бы все цифровые компьютеры в игре, или хорошо ли справлялись бы прямо сейчас доступные компьютеры, а спрашиваем, существуют ли вообразимые компьютеры, которые бы справились. Но это только краткий ответ. Мы увидим этот вопрос в ином свете позже.

4. Цифровые вычислители

Идея цифровых вычислителей может быть объяснена тем, что эти машины предназначены для выполнения любых операций, которые мог бы выполнить человеческий вычислитель. Предполагается, что человеческий вычислитель следует фиксированным правилам; у него нет полномочий отклоняться от них в деталях. Можно предположить, что эти правила приведены в книге, которая изменяется всякий раз, когда он приступает к новой работе. У него также есть неограниченный запас бумаги, на которой он выполнять свои вычисления. Он также может выполнять умножения и сложение на «письмоводной машине», но это не важно.

Если использовать приведённое выше объяснение как определение, мы окажемся в опасности зацикленности аргумента. Мы избегаем этого, дав схему способов достижения желённого эффекта. Цифровой компьютер обычно можно рассматривать как состоящий из трёх частей:

(i) Хранение.

(ii) Исполнительное устройство.

(iii) Управление.

Хранение — это хранилище информации и соответствует бумаге человеческого вычислителя, будь то та бумага, на которой он выполняет вычисления, или та, на которой напечатана его книга правил. Насколько человеческий вычислитель выполняет вычисления в уме, часть хранения будет соответствовать его памяти.

Исполнительное устройство — это часть, которая выполняет различные отдельные операции, вовлечённые в вычисление. Какие именно операции будут выполняться, будет зависеть от машины к машине. Обычно можно выполнить довольно длинные операции, например «Умножить 3540675445 на 7076345687», но в некоторых машинах возможны только очень простые операции, например «Записать 0».

Мы упоминали, что «книга правил», поставляемая компьютеру, в машине замещается частью хранилища. Тогда она называется «таблицей инструкций». Обязанность управления состоит в том, чтобы обеспечить выполнение этих инструкций правильно и в правильном порядке. Управление сконструировано так, что это обязательно происходит.

Информация в хранилище обычно разбита на пакеты умеренно малого размера. В одной машине, например, пакет может состоять из десяти десятичных цифр. Числа присваиваются частям хранилища, в которых хранятся различные пакеты информации, каким-то систематическим образом. Типичная инструкция могла бы говорить:

"Прибавить число, сохранённое в позиции 6809, к числу в 4302 и вернуть результат в ту же позицию хранения."

Естественно, это не встретится в машине, выраженной на английском языке. Скорее всего она будет закодирована в виде, например, 6809430217. Здесь 17 указывает, какой из множества возможных

над двумя числами должна быть выполнена операция. В данном случае это та операция, которая описана выше, а именно: «Прибавьте число. . . .» Будет замечено, что инструкция занимает 10 цифр и потому очень удобно образует один пакет информации. Управление будет

обычно передавать инструкции для исполнения в порядке позиций, в которых они хранятся, но иногда встречается инструкция такого рода, как

«Сейчас выполняй инструкцию, хранящуюся в позиции 5606, и продолжай оттуда»

может встретиться, или снова

«Если в позиции 4505 находится 0, то дальше выполни инструкцию, хранящуюся в 6707, иначе продолжай прямо дальше.»

Инструкции этих последних типов очень важны, потому что они позволяют снова и снова заменять последовательность операций до тех пор, пока не будет выполнено некоторое условие, но при этом подчиняться не свежим инструкциям при каждом повторении, а одним и тем же, снова и снова. Приведём бытовую аналогию. Допустим, мама хочет, чтобы Томми каждый день по дороге в школу заходил к сапожнику, чтобы узнать, готовы ли её туфли; она может спрашивать его снова каждое утро. В качестве альтернативы она может повесить объявление раз и навсегда в коридоре, которое он увидит, когда будет уходить в школу, и которое говорит ему зайти за

туфлями, а также уничтожить объявление, когда он вернётся, если он принёс туфли с собой.

Читатель должен принять как факт, что цифровые компьютеры можно построить и что действительно они построены в соответствии с принципами, которые мы описали, и что они могут

фактически очень точно имитировать действия человека-компьютера.

Книга правил, которую мы описали как используемую человеком-компьютером, конечно, является удобной выдумкой. Реальные люди-компьютеры действительно запоминают, что им нужно делать. Если кто-то хочет, чтобы машина имитировала поведение человеческого компьютера при некоторой сложной операции, надо спросить его, как это делается, а затем перевести ответ в форму таблицы инструкций. Составление таблиц инструкций обычно называют «программированием». Под «программированием машины для выполнения операции А» понимают

внести в машину соответствующую таблицу инструкций, чтобы она выполнила А.

Интересным вариантом идеи цифрового компьютера является «цифровой компьютер со случайным элементом». В них инструкции предусматривают бросание игральной кости или некоторый эквивалентный электронный процесс; одна такая инструкция, например, может быть: «Брось кость и помести получившееся число в память 1000». Иногда такую машину описывают как обладающую свободной волей (хотя я бы сам не использовал эту фразу), Обычно невозможно определить, наблюдая машину, есть ли у неё случайный элемент, потому что сходный эффект можно получить и с помощью таких устройств, которые заставляют выбор зависеть от цифр

десятичного для .

Большинство реальных цифровых компьютеров имеют только конечную память. В теоретическом плане нет трудности в идее компьютера с неограниченной памятью. Конечно, в любой момент может быть использована только конечная часть. Точно так же можно построить только конечное число таких компьютеров, но мы

я могу представить, что будет добавляться всё больше и больше по мере необходимости. Такие компьютеры имеют особый теоретический интерес и будут называться компьютерами бесконечной ёмкости.

Идея цифрового компьютера стара как мир. Чарльз Бэббидж, Лукасский профессор математики в Кембридже с 1828 по 1839 год, спроектировал такую машину, названную Аналитическая машина, но она так и не была завершена. Хотя у Бэббиджа были все основные идеи, тогда его машина не представляла собой столь же привлекательную перспективу. Скорость, которая могла быть доступна, безусловно была бы быстрее человеческого калькулятора, но примерно в 100 раз медленнее манчестерской машины, самой одной из медленнейших современных машин. Хранение должно было быть purely mechanical, using wheels and cards.

Тот факт, что Аналитическая машина Бэббиджа была полностью механической, поможет нам избавиться от суеверия. Важность часто приписывается тому, что современные цифровые компьютеры электрические, и что нервная система тоже электрическая. Так как машина Бэббиджа не была электрической, и поскольку все цифровые компьютеры в каком-то смысле эквивалентны, мы видим, что использование электричества не имеет теоретического значения. Конечно, электричество обычно появляется там, где быстрое сигналиование, поэтому неудивительно, что мы находим его и в этих соединениях. В нервной системе химические явления по крайней мере так же важны, как электрические. В некоторых компьютерах система хранения в основном акустическая. Следовательно, характеристика использования электричества воспринимается как очень поверхностное сходство. Если хотим найти такие сходства, следует больше искать математические аналогии функций.

5. Универсальность цифровых компьютеров

Последовательно рассмотренные в прошлом разделе цифровые компьютеры можно классифицировать как «дискретно-состояния машины». Это те машины, которые переходят резкими скачками или щелчками из одного конкретного состояния в другое. Эти состояния достаточно различны, чтобы можно было проигнорировать возможность путаницы между ними. Строго говоря, таких машин нет. Всё на самом деле движется непрерывно. Но существует множество видов машин, которые можно благоприятно рассматривать как дискретно-состояния машины. Например, при рассмотрении выключателей для осветительной системы удобна фиктивная мысль, что каждый переключатель должен быть либо включён, либо выключен. Должны быть промежуточные положения, но по большинству задач их можно забыть. В качестве примера дискретно-состояний машины можно рассмотреть колесо, которое вращается вокруг 120 раз в секунду, но может быть остановлено тумблером, который можно привести извне; в дополнение лампа загорается в одном из положений колеса. Эта машина может быть описана абстрактно следующим образом. Внутреннее состояние машины (которое описывается положением колеса) может быть q1, q2 или q3. Входной сигнал i0 или i1 (положение тумблера). Внутреннее состояние в любой момент определяется по последнему состоянию и входному сигналу согласно таблице.

(ТАБЛИЦА УДАЛЕНА)

Выходные сигналы, единственное внешнее видимое указание внутреннего состояния (свет), описаны таблицей

Состояние q1 q2 q3

выход o0 o0 o1

Этот пример типичен для машин с дискретными состояниями. Их можно описать такими таблицами, если у них лишь конечное число возможных состояний.

Кажется, что, зная начальное состояние машины и входные сигналы, всегда можно предсказать все будущие состояния. Это напоминает взгляд Лапласа, согласно которому по полному состоянию Вселенной в один момент времени, описанному положениями и скоростями всех частиц, можно было бы предсказать все будущие состояния. Однако рассматриваемое нами предсказание куда ближе к практической осуществимости, чем то, которое рассматривал Лаплас. Система «Вселенной как целого» такова, что совсем небольшие ошибки в начальных условиях могут в более поздний момент дать огромный эффект. Смещение одного электрона на миллиардную долю сантиметра в один момент может через год означать разницу между тем, будет ли человек убит лавиной или спасётся. Существенным свойством механических систем, которые мы назвали «машинами с дискретными состояниями», является то, что этот феномен не возникает. Даже когда мы рассматриваем реальные физические машины вместо идеализированных, достаточно точное знание состояния в один момент даёт достаточно точное знание спустя любое число шагов.

Как уже упоминалось, цифровые компьютеры относятся к классу машин с дискретными состояниями. Но число состояний, на которые способна такая машина, обычно огромно. Например, число состояний машины, работающей сейчас в Манчестере, составляет около 2 165,000, то есть примерно 10 50,000. Сравните это с нашим примером щёлкающего колеса, описанного выше, у которого было три состояния. Нетрудно понять, почему число состояний должно быть столь огромным. Компьютер включает в себя память, соответствующую бумаге, используемой человеком-вычислителем. В памяти должно быть возможно записать любую из комбинаций символов, которые могли бы быть записаны на бумаге. Для простоты предположим, что в качестве символов используются только цифры от 0 до 9. Вариации почерка игнорируются. Предположим, компьютеру разрешено 100 листов бумаги, каждый по 50 строк, каждая строка вмещает 30 цифр. Тогда число состояний равно 10 100x50x30, то есть 10 150,000. Это примерно число состояний трёх машин Манчестера, взятых вместе. Логарифм по основанию два от числа состояний обычно называют «ёмкостью хранения» машины. Таким образом, машина Манчестера имеет ёмкость хранения около 165,000, а колесная машина из нашего примера — около 1.6. Если две машины соединить вместе, их ёмкости нужно сложить, чтобы получить ёмкость результирующей машины. Это приводит к возможности таких утверждений, как: «Машина Манчестера содержит 64 магнитные дорожки, каждая ёмкостью 2560, восемь электронных ламп ёмкостью 1280. Прочее хранилище составляет около 300, что в сумме даёт 174,380.»

Имея таблицу, соответствующую машине с дискретными состояниями, можно предсказать, что она будет делать. Нет причин, по которым этот расчёт нельзя было бы выполнить с помощью цифрового компьютера. Если бы его можно было выполнить достаточно быстро, цифровой компьютер мог бы имитировать поведение любой машины с дискретными состояниями. Тогда в имитационную игру можно было бы играть с данной машиной (как B) и имитирующим цифровым

компьютер (как A) и опрашивающий не смогут различить их. Разумеется, цифровой компьютер должен иметь достаточную емкость памяти, а также работать достаточно быстро. Более того, его нужно перепрограммировать заново для каждой новой машины, которую предполагается подражать.

Эта особая способность цифровых компьютеров, заключающаяся в том, что они могут подражать любой машине с конечным состоянием, описывается как то, что они являются универсальными машинами. Существование машин с такими свойствами имеет важное последствие: неважно, какие по скорости к расчётам, не обязательно разрабатывать различные новые машины для разных вычислительных процессов. Все это можно выполнить на одном цифровом компьютере, надлежащим образом запрограммированном для каждого случая. Можно увидеть, что как следствие этого все цифровые компьютеры в определённом смысле равноправны.

Теперь мы можем снова рассмотреть пункт, поднятый в конце §3. Было предположено временно, что вопрос «Могут ли машины думать?» следует заменить на «Есть ли вообразимые цифровые компьютеры, которые будут хорошо справляться с имитационной игрой?» Если мы захотим, можем сделать это более общим на вид и спросить: «Есть ли дискретно-состоящие машины, которые будут хорошо справляться?» Но учитывая свойство универсальности, мы видим, что любой из этих вопросов эквивалентен следующему: «Зафиксируем наше внимание на одном конкретном цифровом компьютере C. Верно ли, что путём модификации этого компьютера для обеспечения достаточного объёма памяти, надлежащего увеличения скорости его действия и предоставления ему соответствующей программы, C можно заставить надёжно сыграть роль A в имитационной игре, роль B будет занимать человек?»

6. Противоречивые взгляды на основной вопрос

Теперь можно рассмотреть, что площадка очищена, и мы готовы приступить к дебатам по нашему вопросу: "Могут ли машины думать?" и вариант, приведённый в конце предыдущего раздела. Мы не можем полностью отказаться от исходной формы проблемы, потому что мнения будут различаться по поводу уместности замены, и мы должны как минимум выслушать, что нужно сказать по этому поводу.

Сделаем чтение проще: сначала изложу свои убеждения по этому вопросу. Рассмотрим сначала более точную формулировку вопроса. Я считаю, что примерно через пятьдесят лет будет возможно программировать компьютеры с объёмом памяти примерно 10^9, чтобы они так хорошо играли имитационную игру, что у среднего опрашивающего не будет более 70 процентов шанса правильно идентифицировать после пяти минут вопросов. Исходный вопрос «Могут ли машины думать?» я считаю слишком бессмысленным для обсуждения. Тем не менее, я считаю, что к концу века употребление слов и общепринятое образованное мнение изменятся настолько, что можно будет говорить о мышлениях машин, не ожидая возражений. Я также считаю, что conceal эти убеждения бесполезно. Популярный взгляд на то, что учёные неуклонно переходят от устоявшихся фактов к устоявшимся фактам, никогда не влияя на какие-либо улучшенные гипотезы, является совершенно ошибочным. При условии, что ясно различать доказанные факты и гипотезы, вреда не будет. Гипотезы имеют огромное значение, поскольку они предлагают полезные направления исследований.

Я сейчас перейду к рассмотрению мнений, противоположных моему.

(1) Богословское возражение

Мысль — функция бессмертной души человека. Бог дал бессмертную душу каждому мужчине и каждой женщине, но не любому другому животному или машинам. Следовательно, ни одно животное или машина не может думать.

Я не могу принять ни часть этого, но попытаюсь ответить богословскими терминами. Я нашёл бы аргумент более убедительным, если животные были бы отнесены к человеку, потому что, на мой взгляд, существует большее различие между типичным одушевлённым и неодушевлённым, чем между человеком и другими животными. Произвольный характер ортодоксального взгляда становится яснее, если рассмотреть, как он мог бы выглядеть члену другой религиозной общины. Как христиане относятся к мусульманскому взгляду на то, что женщины не имеют душ? Но давайте оставим этот вопрос и вернёмся к основному аргументу. Мне кажется, что приведённый выше аргумент подразумевает серьёзное ограничение всеведущей мощи Всемогущего. Признаётся, что есть вещи, которые Он не может сделать, например сделать одного равным двум, но не следует ли нам верить, что Он имеет свободу даровать душу слону, если посчитает нужным? Мы могли бы ожидать, что Он бы применял эту силу только в сочетании с мутацией, которая обеспечила бы слону соответствующе улучшенный мозг, чтобы удовлетворять потребности такого рода. Аргумент точно такой же формы можно привести и для случая машин. Это может казаться иным, потому что это сложнее «проглотить». Но это действительно просто означает, что мы думаем, что менее вероятно, что Он сочтёт обстоятельства подходящими для дарования души. Обстоятельства, о которых идёт речь, обсуждаются в остальной части этой статьи. При попытке сконструировать такие машины мы не должны неуважительно узурировать Его силу творить души, и не более чем в деле продолжения рода: в любом случае мы являемся инструментами Его воли, обеспечивающими .дома для душ, которых Он создаёт.

Однако это лишь спекуляции. Я не очень впечатлён богословскими аргументами, какими бы не были их применения. Такие аргументы часто оказывались неудовлетворительными в прошлом. Во времена Галилея утверждалось, что тексты: «И солнце стоят и не торопится зайти за день» (Иисус Навин 10:13) и «Он заложил основания земли, чтобы она не поколебалась во ни время» (Псалом 102:5) — были достаточным опровержением коперниковской теории. С нашим нынешним знанием такой аргумент кажется бесполезным. Когда это знание было недоступно, оно производило совершенно другое впечатление.

(2) Возражение «Головы в песок»

Последствия мышления машин были бы слишком ужасны. Давайте надеяться и верить, что они не смогут это сделать.

Этот аргумент редко выражается так откровенно, как в приведённой форме. Но он затрагивает большинство из нас, кто хотя бы немного об этом задумывается. Нам нравится думать, что Человек по какому-то тонкому признаку превосходит остальное творение. Лучше всего, если можно показать, что он обязательно выше, потому что

тогда не будет угрозы потерять им командующее положение. Популярность теологического аргумента явно связана с этим чувством. Это, возможно, будет довольно сильно у интеллектуальных людей, так как они ценят силу мышления выше других и склонны основывать свою веру в превосходство Человека на этой силе.

Я не считаю, что этот аргумент достаточно значителен, чтобы требовать опровержения. Утешение было бы более уместным: возможно, его следует искать в перевоплощении душ.

(3) Математическое возражение

Существует ряд результатов математической логики, которые можно использовать для показа того, что существуют ограничения возможностей дискретно-состояниевых машин. Самый известный из этих результатов известен как теорема Геделя (1931) и показывает, что в любой достаточно могущественной логической системе формулируются высказывания, которые не могут быть доказаны и не могут быть опровергнуты в рамках этой системы, если только сама система не противоречива. Есть и другие, в некоторых отношениях подобные, результаты у Чурча (1936), Креле (1935), Россер и Тьюринга (1937). Последний результат наиболее удобен для рассмотрения, поскольку он напрямую относится к машинам, тогда как остальные могут быть использованы лишь в сравнительно косвенном аргументах: например, если теорема Геделя должна быть использована, нам нужно дополнительно иметь средства описания логических систем через машины, и машины через логические системы. Результат в данном контексте относится к типу машины, которая по существу является цифровым компьютером с бесконечной способностью. Он утверждает, что существуют определённые вещи, которые такая машина не может сделать. Если её запрограммировать так, чтобы давать ответы на вопросы, как в имитационной игре, то найдутся вопросы, на которые она либо даст неправильный ответ, либо не сможет дать ответ вовсе, несмотря на возможное время ответа. Конечно, может быть множество таких вопросов, и вопросы, на которые одной машине ответить нельзя, могут быть удовлетворительно отвечены другой. Мы, конечно, сейчас предположим, что вопросы такие, на которые уместен ответ «Да» или «Нет», а не вопросы вроде «Что ты думаешь о Пикассо?» Вопросы, на которые мы знаем, что машины должны терпеть неудачу, относятся к этому типу: «Рассмотрите машину, указанную следующим образом. . . . Ответит ли эта машина когда-нибудь на какой-либо вопрос «Да»?» Точки заменяются описанием некоторой машины в стандартной форме, которая может быть подобна той, которая использовалась в §5. Когда описанная машина имеет определенные сравнительно простые отношения к машине, которая находится под допросом, можно показать, что ответ либо неверен, либо не даётся. Это и есть математический вывод: утверждается, что он доказывает неспособность машин к которой человеческий интеллект не подвержен.

Краткий ответ на этот аргумент заключается в том, что хотя установлены ограничения Powers Для любой конкретной машины, это утверждение сделано без какого-либо доказательства, что такие ограничения не применимы к человеческому интеллекту. Но я не думаю, что эту точку зрения можно так легкомысленно отвергнуть. Каждый раз, когда одна из этих машин задаёт соответствующий критический вопрос и даёт определённый ответ, мы знаем, что этот ответ должен быть неверным, и это даёт нам определённое чувство превосходства. Это чувство иллюзорно? Безусловно, оно вполне подлинно, но я не думаю, что следует придавать ему слишком большое значение. Мы слишком часто сами даём неправильные ответы на вопросы, чтобы быть удовлетворёнными собой.

при таком свидетельстве о присущей машинам ошибочности. Далее, наше превосходство можно ощутить лишь в таком случае по отношению к той одной машине, над которой мы одержали наш ничтожный триумф. Речь не шла бы о том, чтобы побеждать одновременно все машины. Короче говоря, тогда могут быть люди умнее любой заданной машины, но с другой стороны могут быть и другие машины, снова умнее, и так далее.

Те, кто придерживается математического аргумента, я думаю, в основном были бы готовы принять имитационную игру как основу для обсуждения. Те, кто верит в два предыдущих возражения, вероятно, не будут заинтересованы ни в каких критериях.

(4) Аргумент от сознания

Этот аргумент очень хорошо изложен в лекции Lister Oration профессора Джефферсона за 1949 год, откуда я привожу цитату. «Не прежде чем машина сможет написать сонет или сочинить концерт благодаря испытываемым мыслям и эмоциям, а не случайному падению символов, мы могли бы согласиться, что машина равна мозгу — то есть не только написать это, но и знать, что она это написала. Никакой механизм не мог бы чувствовать (и не просто искусственно сигнализировать, простым ухищрением) удовольствие от своих успехов, горе, когда его клапаны сплавляются, согреваться лестью, становиться несчастным из-за своих ошибок, приходить в очарование от секса, злиться или впадать в депрессию, когда он не может получить то, чего хочет»."

Похоже, что этот аргумент отрицает обоснованность нашего теста. Согласно самой крайней форме этого взгляда единственный способ удостовериться в том, что машина думает, — это быть машиной и чувствовать, как ты сам думаешь. Тогда можно было бы описать эти чувства миру, но, конечно, никто не был бы вправе принимать это во внимание. Точно так же согласно этому взгляду единственный способ узнать, что человек думает, — это быть именно тем человеком. На самом деле это солипсистская точка зрения. Возможно, это самый логичный взгляд, чтобы его придерживаться, но он делает общение идей затруднительным. А склонен верить «А думает, но В не думает», тогда как В верит «В думает, но А не думает». Вместо того чтобы постоянно спорить об этом, обычно придерживаются вежливого соглашения, согласно которому каждый думает.

Я уверен, что профессор Джефферсон не желает придерживаться крайней и солипсистской точки зрения. Вероятно, он был бы вполне готов принять имитационную игру как тест. Игра (с исключением игрока В) часто используется на практике под названием viva voce, чтобы выяснить, действительно ли кто-то понимает что-то или «выучил это попугайским способом». Послушаем фрагмент такого viva voce:

Интеррогатор: В первой строке вашего сонета, которая гласит «Сравню ли я тебя с летним днём», не подошло бы «весенний день» так же или даже лучше?

Свидетель: Оно не просится по размеру.

Интеррогатор: А как насчёт «зимнего дня» — это будет по размеру.

Свидетель: Да, но никто не хочет, чтобы его сравнивали с зимним днём.

Допросчик: Скажите ли вы, что мистер Пиквик напоминал вам Рождество?

Свидетель: В какой-то мере.

Допросчик: Однако Рождество — это зимний день, и я не думаю, что мистеру Пиквику понравится сравнение.

Свидетель: Я не думаю, что вы серьёзны. Под «зимним днём» подразумевают обычный зимний день, а не особенный, как Рождество.

И так далее. Что бы сказал профессор Джефферсон, если машина для сочинения сонетов могла бы отвечать так на устном экзамене? Не знаю, счёл бы он машину «просто искусственно сигнализирующей» эти ответы, но если ответы были бы такими же удовлетворительными и устойчивыми, как в приведённом выше отрывке, я не думаю, что он назвал бы это «легким устройством». Эта фраза, по моему мнению, призвана охватывать такие устройства, как включение в машину записи того, как кто-то читает сонет, со соответствующим переключением, чтобы включать её время от времени.

Короче говоря, я думаю, что большинство тех, кто поддерживает аргумент сознания, можно убедить отказаться от него, чем вынудить принять солипсистскую позицию. Они, вероятно, будут готовы принять наш тест.

Я не хочу создавать впечатление, что считаю, будто в сознании нет загадки. Есть, к примеру, нечто парадоксальное, связанное с любой попыткой локализовать его. Но я не думаю, что эти загадки обязательно нужно разрешать до того, как мы сможем ответить на вопрос, который рассматривается в этой работе.

(5) Аргументы из разных ограничений

Эти аргументы принимают форму: «Я признаю, что вы можете заставить машины выполнять все перечисленные вами вещи, но вы никогда не сумеете заставить одну сделать X». В этом отношении здесь предлагается выбор многочисленных признаков X:

Будьте добры, находчивы, красивы, дружелюбны, проявляйте инициативу, имейте чувство юмора, различайте добро и зло, делайте ошибки, влюбляйтесь, наслаждайтесь клубникой со сливками, заставляйте кого-то влюбиться в вас, учитесь на опыте, правильно употребляйте слова, будьте предметом своих мыслей, имейте такое же разнообразие поведения, как у человека, создавайте что-то действительно новое.

Обычно для этих утверждений не приводят доказательств. Я считаю, что они в основном основаны на принципе научной индукции. Человек видел тысячи машин за свою жизнь. Судя по тому, что он видит, он делает ряд общих выводов. Они некрасивы, каждая — предназначена для очень ограниченной цели, при необходимости для минуточно другой цели они бесполезны, разнообразие поведения любой из них очень мало и т. д. Естественно, он делает вывод, что это необходимые свойства машин в целом. Многие из этих ограничений связаны с очень малой емкостью памяти большинства машин. (Я предполагаю, что идея емкости памяти как-то распространяется на

покрывает машины, отличные от машин с дискретными состояниями. Точное определение не важно, поскольку в настоящем обсуждении не заявляется математическая точность,) Несколько лет назад, когда о цифровых компьютерах было известно совсем немного, можно было вызвать массу недоверия к ним, если упоминать их свойства, не описывая их конструкцию. Это, по-видимому, объяснялось сходным применением принципа научной индукции. Эти применения принципа, конечно, в значительной степени бессознательны. Когда обожжённый ребёнок боится огня и показывает, что боится его, избегая его, я должен сказать, что он применял научную индукцию. (Я, разумеется, мог бы также описать его поведение во многих других формах.) Труды и обычаи человечества, по-видимому, не являются особенно подходящим материалом для применения научной индукции. Если нужно получить надёжные результаты, необходимо исследовать очень большую часть пространства-времени. Иначе мы можем (как это делают большинство английских «детей») решить, что все говорят по-английски, и что учить французский — глупо.

Однако о многих из упомянутых «недостатков» следует сделать особые замечания. Неспособность получать удовольствие от клубники со сливками могла показаться читателю легкомысленной. Возможно, можно было бы создать машину, которая будет получать удовольствие от этого восхитительного блюда, но любые попытки заставить её делать это были бы идиотскими. Важно то, что этот недостаток способствует некоторым другим недостаткам, например, трудности возникновения такого же рода дружелюбия между человеком и машиной, как между белым человеком и белым человеком или между чёрным человеком и чёрным человеком.

Утверждение, что «машины не могут допускать ошибок», кажется странным. Возникает искушение возразить: «А разве от этого они хоть немного хуже?» Но давайте займём более сочувственную позицию и постараемся понять, что именно имеется в виду. Я думаю, что эту критику можно объяснить через «игру в подражание». Утверждается, что вопрошающий мог бы отличить машину от человека просто задав им ряд задач по арифметике. Машину разоблачили бы из-за её смертельно точных результатов. Ответ на это прост. Машина (запрограммированная для игры) не будет пытаться давать правильные ответы на задачи по арифметике. Она намеренно будет вводить ошибки так, чтобы это было рассчитано на то, чтобы запутать вопрошающего. Механическая неисправность, вероятно, проявилась бы через неподходящее решение о том, какой именно тип ошибки следует допустить в арифметике. Даже это толкование критики не является достаточно сочувственным. Но мы не можем позволить себе уделять этому слишком много места. Мне кажется, что эта критика опирается на смешение двух видов ошибок. Назовём их «ошибками функционирования» и «ошибками заключения». Ошибки функционирования возникают из-за какой-то механической или электрической неисправности, которая заставляет машину вести себя иначе, чем предполагалось при её проектировании. В философских дискуссиях обычно любят игнорировать возможность таких ошибок; поэтому обсуждаются «абстрактные машины». Эти абстрактные машины — математические вымыслы, а не физические объекты. По определению они неспособны совершать ошибки функционирования. В этом смысле мы можем по праву сказать, что «машины никогда не могут допускать ошибок». Ошибки заключения могут возникать только тогда, когда к выходным сигналам машины придаётся определённый смысл. Машина, например, могла бы напечатать математические уравнения или фразы на английском языке. Когда напечатается ложное утверждение, мы говорим, что машина совершила ошибку заключения. Совершенно нет оснований утверждать, что машина не может допустить такой вид ошибки. Она могла бы делать лишь одно — снова и снова печатать «О = I». Взять менее

извращенный пример, возможно, у него есть некоторый метод делать выводы с помощью научной индукции. Мы должны ожидать, что такой метод иногда приведет к ошибочным результатам.

Утверждение, что машина не может быть предметом своих мыслей, разумеется, можно ответить только если показать, что у машины есть некоторые мысли по какому-то предмету. Тем не менее «предмет деятельности машины» действительно имеет смысл, по крайней мере для людей, работающих с ней. Например, если машина пыталась найти решение уравнения x^2 - 40x - 11 = 0, можно было бы устами сказать, что это уравнение является частью предмета деятельности машины в этот момент. В таком смысле машина безусловно может быть своим собственным предметом. Её можно использовать для помощи в составлении собственных программ или для предсказания эффекта изменений в её собственной структуре. Наблюдая за результатами своего поведения, она может модифицировать свои программы так, чтобы достигать целей эффективнее. Это возможности ближайшего будущего, а не утопические мечты.

Критика того, что машина не может иметь большого разнообразия поведения, по сути, является способом сказать, что у неё не так много памяти. До недавнего времени ёмкость памяти даже в тысячу цифр была редкостью.

Критика, которую мы здесь рассматриваем, часто принимает замаскированную форму аргумента об сознании. Обычно утверждение, что машина может выполнить одно из таких действий, и описание того рода метода, который машина может использовать, не производит сильного впечатления. Считается, что метод (каким бы он ни был, потому что он должен быть механическим) на самом деле достаточно низок. Сравните скобки в заявлении Джефферсона, приведенном на странице 22.

(6) Возражение леди Лавлейс

Наиболее подробная информация о Аналитической машине Бэббиджа поступает из мемуаров леди Лавлейс (1842). В них она пишет: «Аналитическая машина не претендует на создание чего-то нового. Она может делать то, что мы знаем, как ей приказать выполнять» (курсив автора). Это утверждение приводит Хартер (1949), который добавляет: «Это не говорит о том, что может быть невозможно сконструировать электронное оборудование, которое будет «думать за себя», или в биологических терминах, можно установить условный рефлекс, который послужит основой для «обучения». Вопрос о принципиальной возможности или невозможности это — возбуждающий и интригующий вопрос, на который указывают некоторые из недавних разработок. Но в то время казалось, что машины, построенные или спроектированные, не обладали этим свойством.»

Я полностью согласен с Хартри по этому поводу. Заметим, что он не утверждает, что машины в вопросе не обладали этим свойством, а скорее что доступные леди Лавлейс доказательства не убеждали её поверить, что они обладали им. Совершенно возможно, что машины в вопросе в некотором смысле обладали этим свойством. Предположим, что некоторая машина с конечным числом состояний обладает этим свойством. Аналитическая машина была универсальным цифровым компьютером, так что, если её ёмкость памяти и скорость будут достаточны, её можно с помощью надлежащего программирования заставить имитировать машина в вопросе. Вероятно, этот аргумент

не пришло в голову графине или Бэббиджу. В любом случае не было обязательства перед ними претендовать на все, что можно предъявлять.

Этот вопрос будет рассмотрен снова под заголовком «обучающие машины».

Вариант возражения леди Люблейс гласит, что машина «никогда не может делать ничего по-настоящему нового». Это можно на мгновение опровергнуть пословицей: «нет ничего нового под солнцем». Кто может быть уверен, что «оригинальная работа», которую он сделал, не есть просто рост семени, вложенного в него обучением, или следствие следования хорошо известным общим принципам. Лучший вариант возражения говорит, что машина никогда не «удивляет нас». Это заявление является более прямым вызовом и может быть встречено прямо. Машины удивляют меня с большой частотой. Это во многом потому, что я не провожу достаточных расчетов, чтобы определить, чего ожидать от них, или, точнее, потому, что хотя я и делаю расчет, делаю его торопливо и небрежно, рискуя. Возможно, я думаю про себя: «Полагаю, здесь напряжение должно быть таким же, как там: в любом случае предположим, что так.» Естественно, я часто ошибаюсь, и результат оказывается для меня сюрпризом, потому что к моменту завершения эксперимента эти предположения забыты. Эти признания открывают мне дорогу к лекциям по теме моих порочных привычек, но не бросают сомнение в мою достоверность, когда я свидетельствую о сюрпризмах, которые испытываю.

Я не ожидаю, что этот ответ утихомирит моего критика. Вероятно, он скажет, что сюрпризы происходят из-за какого-то творческого умственного акта с моей стороны и не отражают никакой славы машине. Это возвращает нас к аргументации из сознания и далеко не к идее изумления. Это линия рассуждений, которую мы должны считать закрытой, но стоит отметить, что оценка чего-либо как удивительного требует «творческого умственного акта» в той же мере, независимо от того, происходит ли удивительный факт от человека, книги, машины или чего-либо другого.

Взгляд на то, что машины не могут порождать сюрпризы, по моему мнению, связан с заблуждением, которому особенно подвержены философы и математики. Это предположение, что как только факт представлен разуму, все последствия этого факта возникают в сознании одновременно с ним. Это очень полезное предположение во многих случаях, но слишком легко забывается, что оно ложно. Естественным следствием этого является предположение, что нет доблести в простом проработке последствий из данных и общих принципов.

(7) Аргумент от непрерывности в нервной системе

Нервная система, безусловно, не является дискретно-состоящей машиной. Малая ошибка в информации о размере нервного импульса, воздействующего на нейрон, может вызвать значительную разницу в размере исходящего импульса. Можно привести аргумент: если так, то невозможно ожидать, что поведение нервной системы можно эмулировать с помощью дискретной(state) системы.

Верно, что дискретная машина должна отличаться от непрерывной. Но если мы придерживаемся условий имитационной игры, опрашиватель не сможет воспользоваться этой разницей. Ситуацию можно сделать яснее, если рассмотреть звукоподобность.

более простая непрерывная машина. Дифференциальный анализатор вполне подойдет. (Дифференциальный анализатор — это определенный вид машины, не типа с дискретными состояниями, используемого для некоторых видов вычислений.) Некоторые из них выдают свои ответы в напечатанном виде, и поэтому подходят для участия в игре. Было бы невозможно для цифрового компьютера предсказать точно, какие ответы дифференциальный анализатор даст на задачу, но он вполне способен дать ответы нужного типа. Например, если попросят дать значение (на самом деле около 3.1416), то разумно случайным образом выбрать между значениями 3.12, 3.13, 3.14, 3.15, 3.16 с вероятностями 0.05, 0.15, 0.55, 0.19, 0.06 (скажем). При этих обстоятельствах экзаменатору будет очень трудно отличить дифференциальный анализатор от цифрового компьютера.

(8) Аргумент от неформальности поведения

Невозможно составить набор правил, которые претендовали бы на то, чтобы описывать, что человек должен делать в любой мыслимой ситуации. Например, можно установить правило: нужно остановиться, когда видишь красный сигнал светофора, и идти, если видишь зеленый, но что если из-за какой-нибудь неисправности оба появятся одновременно? Возможно, решат, что самое безопасное — остановиться. Но из этого решения позже вполне может возникнуть еще одна трудность. Пытаться предусмотреть правила поведения, охватывающие каждую случайность, даже возникающую из-за светофоров, представляется невозможным. Со всем этим я согласен.

Из этого утверждают, что мы не можем быть машинами. Я попробую воспроизвести этот аргумент, но боюсь, что вряд ли сделаю это должным образом. Похоже, он строится примерно так. «если бы у каждого человека был определенный набор правил поведения, с помощью которых он регулировал бы свою жизнь, он не был бы лучше машины. Но таких правил не существует, значит, люди не могут быть машинами». Нераспределенный средний термин бросается в глаза. Я не думаю, что этот аргумент когда-либо формулировали именно так, но полагаю, что, тем не менее, используется именно он. Однако может быть некоторая путаница между «правилами поведения» и «законами поведения», которая затуманивает вопрос. Под «правилами поведения» я имею в виду предписания вроде «Остановись, если увидишь красные огни», на которые можно действовать и о которых можно помнить. Под «законами поведения» я имею в виду законы природы, примененные к телу человека, такие как «если ущипнуть его, он пискнет». Если в процитированном аргументе подставить «законы поведения, которые регулируют его жизнь» вместо «законов поведения, с помощью которых он регулирует свою жизнь», нераспределенный средний термин больше не является непреодолимым. Ибо мы полагаем, что верно не только то, что человек, находящийся под управлением законов поведения, является некоторого рода машиной (хотя и не обязательно машиной с дискретными состояниями), но и то, что, наоборот, такая машина означает, что она регулируется этими законами. Однако нам не так легко убедить себя в отсутствии полных законов поведения, как в отсутствии полных правил поведения. Единственный способ, который нам известен, чтобы найти такие законы, — это научное наблюдение, и мы, конечно, не знаем ни обстоятельств, при которых можно было бы сказать: «Мы достаточно искали. Таких законов нет.»

Мы можем доказать еще убедительнее, что любое такое утверждение было бы необоснованным. Предположим, что мы могли бы быть уверены в том, что найдем такие законы, если бы они существовали. Тогда, применительно к машине с дискретными состояниями, должно быть совершенно возможно, наблюдая, выяснить о ней достаточно, чтобы предсказать ее дальнейшее поведение — и сделать это за разумное время, скажем, за тысячу лет. Но, похоже, дело обстоит иначе. Я настроил на манчестерском компьютере небольшую

программа, использующая всего 1 000 единиц памяти, при этом машина, получив шестнадцатизначное число, отвечает другим числом в течение двух секунд. Я озадачу любого: научиться по этим ответам достаточно понять программу, чтобы уметь предсказывать любые ответы на непроверяемые значения.

(9) Аргумент в пользу экстрасенсорного восприятия

Я полагаю, что читатель знаком с идеей экстрасенсорного восприятия и значением четырех его элементов, а именно телепатия, ясновидение, предвидение и психокинез. Эти тревожные явления, похоже, опровергают все наши обычные научные представления. Как бы мы хотели их дискредитировать! К сожалению, статистические данные, по крайней мере для телепатии, обескураживают. Очень трудно перестроить свои идеи, чтобы вписать эти новые факты. Как только приняты они, не кажется большим шагом поверить в призраков и домовых. Идея о том, что наши тела движутся просто по известным законам физики, вместе с некоторыми другими, еще не обнаруженными, но несколько похожими, была бы одним из первых факторов к исчезновению.

Этот аргумент, на мой взгляд, довольно убедителен. Можно ответить, что многие научные теории остаются применимыми на практике, несмотря на столкновение с ЭСП; что, по сути, можно жить очень спокойно, если забыть об этом. Это довольно холодное утешение, и мы опасаемся, что мышление — это именно тот вид явления, где ESP может быть особенно актуальным.

Более конкретный аргумент на основе ЭСП мог бы звучать так: «Давайте сыграем в имитационную игру, в качестве свидетелей возьмем человека, который хорош как телепатический приёмник, и цифровой компьютер. Опрашивающий может задать такие вопросы, как: “К какому масти принадлежит карта в моей правой руке?” Человек по телепатии или ясновидению даёт правильный ответ 130 раз из 400 карт. Машина может только гадать наугад и, возможно, набирает 104 правильных, поэтому опрашиватель делает правильную идентификацию». Здесь открывается интересная возможность. Предположим, цифровой компьютер содержит генератор случайных чисел. Тогда естественно использовать это для решения того, какой ответ дать. Но тогда генератор случайных чисел будет подчиняться психокинетическим силам опрашивателя. Возможно, эта психокинезия заставит машину давать правильный ответ чаще, чем ожидалось бы по теории вероятностей, так что опрашиватель может по-прежнему не суметь сделать правильную идентификацию. С другой стороны, он может угадать правильно и без any questioning, clairvoyance. С ЭСП может произойти всё, что угодно.

Если телепатия принята, нужно сделать наш тест более строгим. Ситуацию можно рассматривать как аналог той, которая произошла бы, если опрашивающий говорит сам с собой, а один из соперников слушает у стены. Помещение конкурентов в «помещение без телепатии» удовлетворило бы всем требованиям.

7. Обучающиеся машины

Читатель наверняка ожидал, что у меня нет очень убедительных аргументов положительного характера в поддержку моих взглядов. Если бы они были, я не стал бы так старательно указывать на заблуждения противоположных взглядов. Доказательства, которые у меня есть, я приведу далее.

Давайте ненадолго вернемся к возражению леди Лавлейс, которое утверждало, что машина может делать только то, что мы ей скажем. Можно сказать, что человек может "ввести" идею в машину, и что она в определенной степени ответит, а затем впадет в состояние покоя, как рояльная струна, ударенная молотком. Другая аналогия — атомный реактор (pile) меньшего, чем критический, размера: введенная идея должна соответствовать нейтрону, который входит в реактор извне. Каждый такой нейтрон вызывает некоторое возмущение, которое в конце концов затухает. Однако если размер реактора достаточно увеличить, то возмущение, вызванное таким входящим нейтроном, весьма вероятно будет продолжать нарастать снова и снова, пока весь реактор не будет разрушен. Есть ли соответствующее явление для умов и есть ли оно для машин? Похоже, что для человеческого ума оно есть. Большинство из них, судя по всему, являются "докритическими", то есть в этой аналогии соответствуют реакторам докритического размера. Идея, представленная такому уму, в среднем порождает менее одной идеи в ответ. Небольшая часть является надкритической. Идея, представленная такому уму, может породить целую "теорию", состоящую из вторичных, третичных и более отдаленных идей. Умы животных, судя по всему, весьма определенно являются докритическими. Следуя этой аналогии, мы спрашиваем: "Можно ли сделать машину надкритической?"

Аналогия с "кожурой луковицы" тоже полезна. Рассматривая функции ума или мозга, мы обнаруживаем некоторые операции, которые можно объяснить в чисто механических терминах. Мы говорим, что это не соответствует реальному уму: это своего рода кожура, которую нужно снять, если мы хотим найти реальный ум. Но затем в том, что остается, мы находим еще одну кожуру, которую нужно снять, и так далее. Продвигаясь таким образом, мы вообще когда-нибудь приходим к "реальному" уму, или в конце концов приходим к кожуре, в которой ничего нет? В последнем случае весь ум является механическим. (Однако это не была бы машина с дискретными состояниями. Мы это обсуждали.)

Последние два абзаца не претендуют на то, чтобы быть убедительными аргументами. Их скорее следует описать как "повторения, стремящиеся вызвать веру."

Единственная по-настоящему удовлетворительная поддержка для взгляда, высказанного в начале §6, будет той, которую можно получить, если дождаться конца века и затем провести описанный эксперимент. Но что можно сказать тем временем? Какие шаги следует предпринять сейчас, если эксперимент должен оказаться успешным?

Как я объяснил, проблема в основном заключается в программировании. Потребуются и достижения в области инженерии, но кажется маловероятным, что они окажутся недостаточными для требований. Оценки емкости памяти мозга варьируются от 1010 до 1015 двоичных разрядов. Я склоняюсь к меньшим значениям и полагаю, что для высших типов мышления используется лишь очень небольшая доля. Вероятно, большая часть используется для сохранения зрительных впечатлений; я был бы удивлен, если бы для удовлетворительной игры в имитационную игру, во всяком случае против слепого человека, потребовалось бы более 109. (Примечание: емкость Encyclopaedia Britannica, 11-го издания, составляет 2 X 109) Емкость памяти 107 было бы вполне осуществимой возможностью даже при нынешних методах. Вероятно, вовсе не нужно увеличивать скорость операций машин. Части современных машин, которые можно рассматривать как аналоги нервных клеток, работают примерно в тысячу раз быстрее, чем последние. Это должно обеспечить "запас прочности", который мог бы покрыть потери

скорость возникает различными способами, наша задача состоит в том, чтобы выяснить, как запрограммировать эти машины на игру. При моем нынешнем темпе работы я получаю примерно тысячу цифр progriatine в день, так что около шестидесяти рабочих, стабильно работавших в течение пятидесяти лет, могли бы выполнить работу, если бы ничего не оказалось в корзине макулатуры. Кажется желательным более Expedite метод.

В процессе попытки имитировать взрослый человеческий разум мы обязаны думать о процессе, который привел его к тому состоянию, в котором он находится. Мы можем отметить три компонента.

(a) начальное состояние ума, скажем, при рождении,

(b) образование, которому его подвергали,

(c) Другой опыт, который нельзя назвать образованием, которому его подвергали.

Вместо того чтобы пытаться создать программу для моделирования взрослого разума, почему бы не создать такую, которая моделирует ребенка? Если затем поместить ее в соответствующий курс обучения, можно получить взрослый мозг. Предположительно детский мозг похож на блокнот, который покупают в почтовом отделении. Гораздо меньше механизмов и много пустых страниц. (Механизм и письмо с нашей точки зрения почти синонимы.) Надежда наша состоит в том, что в детском мозге так мало механизма, что нечто вроде этого может быть легко запрограммировано. Количество работы в образовании, которое мы можем предположить, как первый приближенный эквивалент, будет примерно таким же, как и у человеческого ребенка.

Таким образом, мы разделили нашу задачу на две части. Программа для ребенка и процесс образования. Эти две части остаются очень тесно связанными. Мы не можем рассчитывать на то, чтобы на первом же попытке найти хорошую машину для ребенка. Нужно экспериментировать с обучением такой машины и смотреть, насколько хорошо она учится. Затем можно попробовать другую и посмотреть, лучше она или хуже. Существует очевидная связь между этим процессом и эволюцией, по принципу идентификаций.

Структура детской машины = наследственный материал

Изменения детской машины = мутация,

Естественный отбор = суждение экспериментатора

Однако можно надеяться, что этот процесс будет более быстрым, чем эволюция. Выживание наиболее приспособленного — медленный метод измерения преимуществ. Экспериментатор, применяя интеллект, должен суметь ускорить этот процесс. Не менее важно то, что он не ограничен случайными мутациями. Если он сможет проследить причину некоторой слабости, он, вероятно, подумает о виде мутации, который улучшит ее.

Не будет возможно применить точно такой же процесс обучения к машине, как и к обычному ребенку. Например, ей не дадут ног, поэтому ее не будут просить

выйти наружу и заполнить уголок для угля. Возможно, у него может не быть глаз. Но как бы хорошо ни были преодолены эти недостатки благодаря умной инженерии, нельзя отправить существо в школу, пока другие дети не станут смеяться над ним слишком часто. Ему нужно дать какое-то обучение. Нам не стоит слишком беспокоиться о ногах, глазах и т.д. Пример мисс Хелен Келлер показывает, что образование может происходить при условии, что общение в обоих направлениях между учителем и учеником может происходить каким-либо образом.

Обычно мы связываем наказания и вознаграждения с процессом обучения. Некоторые простые детские машины могут быть сконструированы или запрограммированы на основе подобного принципа. Машина должна быть так сконструирована, чтобы события, которые незадолго до появления сигнала наказания, с маловероятностью повторялись, тогда как сигнал вознаграждения увеличивал вероятность повторения тех событий, которые привели к нему. Эти определения не предполагают никаких чувств со стороны машины, я провёл несколько экспериментов с одной такой детской машиной и сумел научить её нескольким вещам, но метод обучения был слишком нетрадиционен, чтобы эксперимент считался действительно успешным.

Использование наказаний и вознаграждений может быть по крайней мере частью процесса обучения. Говоря вкратце, если у учителя нет других средств общения с учащимся, количество информации, которое может до него дойти, не превышает общее количество применённых вознаграждений и наказаний. К моменту, когда ребёнок научится повторять «Казабианка», он, вероятно, почувствует себя очень обиженным, если текст можно открыть только с помощью техники «Двадцать вопросов», где каждый «НЕТ» принимает форму удара. Таким образом, необходимо иметь ещё какие-то «безэмоциональные» каналы общения. Если они доступны, то возможно обучить машину с помощью наказаний и вознаграждений подчиняться приказам, данным на каком-либо языке, например, символическому языку. Эти приказы должны передаваться через «безэмоциональные» каналы. Использование этого языка значительно уменьшит число необходимых наказаний и вознаграждений.

Мнения по поводу сложности, которая подходит для детской машины, могут различаться. Можно постараться сделать её как можно проще, сохраняя общие принципы. Либо можно реализовать полную систему логических выводов «встроенную в неё». В таком случае склад будет в основном занят определениями и положениями. Положения будут иметь различный статус, например хорошо установленные факты, гипотезы, математически доказанные теоремы, высказывания, данные авторитетом, выражения, имеющие логическую форму положения, но не имеющие ценности убеждения. Некоторые положения могут быть описаны как «императивы». Машина должна быть сконструирована так, чтобы как только императив отнесён к «хорошо установленным», соответствующее действие автоматически происходило. Чтобы проиллюстрировать это, предположим, учитель говорит машине: «Сделай домашнее задание сейчас». Это может привести к тому, что «Учитель говорит: 'Сделай домашнее задание сейчас'» будет включено в число хорошо установленных фактов. Другой подобный факт может быть: «Всё, что говорит учитель, истинно». Объединение этих фактов может в конечном итоге привести к императиву «Сделай домашнее задание сейчас», который, по конструкции машины, означает, что домашнее задание действительно начнётся, но эффект будет очень удовлетворительным. Процессы вывода, используемые машиной, не обязательно должны быть такими, как удовлетворят самых строгих логиков. Например, может не быть иерархии типов. Но это не означает, что будут возникать логические ошибки, не больше чем мы обязаны споткнуться о них.

огороженные скалы. Подходящие императивы (выраженные внутри систем, не являющиеся частью правил системы), такие как «Не используйте класс, если он не является подклассом того, который был упомянут учителем», могут иметь подобный эффект на «Не подходите слишком близко к краю».

Императивы, которым может следовать машина без конечностей, будут в первую очередь интеллектуальными, как в приведённом выше примере (выполнение домашнего задания). Среди таких императивов важнейшими будут те, которые регулируют порядок применения правил соответствующей логической системы. На каждом этапе использования логической системы существует очень большое количество альтернативных шагов, any из которых разрешено применить в рамках подчинения правилам логической системы. Эти выборы сильно различают выдающегося рассудителя от мелочного, а не различают надёжного от ошибочного. Пропозиции, ведущие к императивам такого рода, могут звучать как «Когда упоминается Сократ, используй силлогизм по Барбаре» или «Если один метод доказан быстрее другого, не используйте более медленный метод». Некоторые из них могут быть «даны по власти», но другие могут быть произведены самой машиной, например, путём научной индукции.

Идея обучающейся машины может показаться парадоксальной для некоторых читателей. Как могут изменяться правила работы машины? Они должны полностью описывать, как машина отреагирует независимо от её истории и любых изменений. Правила, следовательно, во времени инвариантны. Это совершенно верно. Объяснение парадокса состоит в том, что правила, которые изменяются в процессе обучения, являются более скромного рода и заявляют лишь мимолётную валидность. Читатель может провести параллель с Конституцией Соединённых Штатов.

Важной особенностью обучающейся машины является то, что её учитель часто будет в значительной степени не знать, что происходит внутри, хотя он всё же может отчасти предсказать поведение ученика. Это следует применить сильнее к позднему обучению машины, возникающей из детской машины хорошо продуманного дизайна (или программы). Это ясно контрастирует с обычной процедурой использования машины для вычислений: цель — иметь чёткое мысленное представление о состоянии машины в каждый момент вычисления. Эта цель может быть достигнута только через усилия. Утверждение «машина может сделать только то, что мы знаем, как ей приказать» кажется странным на фоне этого. Большинство программ, которые мы можем поместить в машину, приведут к тому, что она сделает что-то, что нам трудно понять (вообще невозможно или мы считаем поведение полностью случайным). Интеллектуальное поведение, по всей видимости, состоит в отклонении от полностью дисциплинированного поведения, связанного с вычислениями, но в довольно слабом отклонении, которое не приводит к случайному поведению или к бессмысленным повторяющимся циклам. Ещё один важный результат подготовки нашей машины к роли в игре подражания через процесс обучения и обучения состоит в том, что «человеческая человеческая ошибка» может быть естественным образом опущена, то есть без специального «тренинга». (Читатель должен соотнести это с точкой зрения на стр. 23 и 24.) Процессы, которые обучаются, не дают сто процентной уверенности в результате; если бы они дали, их нельзя было бы «разучить».

Наверное, разумно включить случайный элемент в обучающую машину. Случайный элемент довольно полезен, когда мы ищем решение какой-то задачи. Предположим, например, что мы хотели бы найти число между 50 и 200, которое равно квадрату суммы его цифр: мы могли бы начать с 51, затем попробовать 52 и продолжать, пока не получим число, которое подошло. В качестве альтернативы мы могли бы выбирать числа случайным образом, пока не попадём на хорошее. У этого метода есть преимущество в том, что не нужно отслеживать значения, которые уже пробовали, но есть и недостаток: можно попробовать одно и то же дважды; однако это не очень важно, если существует несколько решений. Систематический метод имеет недостаток в том, что в исследуемой в первую очередь области может оказаться огромный участок без единого решения. Теперь процесс обучения можно рассматривать как поиск формы поведения, которая удовлетворит учителя (или какой-либо другой критерий). Поскольку, вероятно, существует очень большое число удовлетворительных решений, случайный метод, похоже, лучше систематического. Следует заметить, что он используется в аналогичном процессе эволюции. Но там систематический метод невозможен. Как можно было бы отслеживать различные генетические комбинации, которые уже были опробованы, чтобы не пытаться их снова?

Можно надеяться, что в конечном счёте машины будут конкурировать с людьми во всех чисто интеллектуальных областях. Но с каких лучше всего начинать? Даже это — трудный выбор. Многие считают, что лучше всего подойдёт очень абстрактная деятельность, например игра в шахматы. Также можно утверждать, что лучше всего снабдить машину лучшими органами чувств, которые можно купить за деньги, а затем научить её понимать и говорить по-английски. Этот процесс мог бы идти по примеру обычного обучения ребёнка: предметы будут показывать и называть и т. д. Снова же, я не знаю, какой ответ правильный, но думаю, что оба подхода следует попробовать.

Мы можем видеть лишь на небольшое расстояние вперёд, но там достаточно того, что нужно сделать.