

ДЖ. фон НЕЙМАН ОБЩАЯ И ЛОГИЧЕСКАЯ ТЕОРИЯ АВТОМАТОВ³⁶

В естественных науках автоматы играли роль, значение которой непрерывно возрастало и которая к настоящему времени стала весьма значительной. Этот процесс развивался в течение нескольких десятилетий. В конце этого периода автоматы стали захватывать и некоторые области математики, в частности (но не только) математическую физику и прикладную математику. Их роль в математике представляет интересный аналог некоторых сторон жизнедеятельности организмов в природе. Как правило, живые организмы гораздо более сложны и тоньше устроены и, следовательно, значительно менее понятны в деталях, чем искусственные автоматы. Тем не менее рассмотрение некоторых закономерностей устройства живых организмов может быть весьма полезно при изучении и проектировании автоматов. И наоборот, многое из опыта нашей работы с искусственными автоматами может быть до некоторой степени перенесено на наше понимание естественных организмов.

I. ПРЕДВАРИТЕЛЬНЫЕ СООБРАЖЕНИЯ

Подразделение проблемы, природа элементов, аксиоматический подход к их синтезу

При сравнении живых организмов и, в частности, наиболее сложно организованной системы — нервной системы человека — с искусственными автоматами следует иметь в виду следующее ограничение. Естественные системы чрезвычайно сложны, и ясно, что проблему их изучения необходимо подразделить на несколько частей. Один метод такого расчленения, особенно важный в нашем случае, заключается в следующем. Организмы можно рассматривать как составленные из частей, из элементарных единиц, которые в определенных пределах автономны. Поэтому можно считать первой частью проблемы исследование структуры и функционирования таких элементарных единиц в отдельности. Вторая часть проблемы состоит в том, чтобы понять, как эти элементы организованы в единое целое и каким образом функционирование целого выражается в терминах этих элементов.

Первая часть нашей проблемы в настоящее время является основной проблемой физиологии. Она тесно связана с наиболее трудными главами органической и физической химии и в свое время, по-видимому, будет решена в значительной мере с помощью квантовой механики. Я недостаточно компетентен, чтобы входить в обсуждение этих вопросов, и в настоящей работе не буду рассматривать эту часть проблемы.

³⁶ Работа впервые опубликована в книге: *Cerebral Mechanisms in Behavior*. The Nixon Symposium. Edited by Lloyd A. Jeffress, New York—London, 1951, p. 2070–2098. Книга представляет собой отчет о симпозиуме на тему «Механизмы мозга в поведении», состоявшемся в Калифорнийском технологическом институте в сентябре 1948 г. Симпозиум был организован комитетом так называемого Хиксоновского фонда (основан в 1938 г. на средства Ф. П. Хиксона с целью поощрения исследовательских работ в области изучения поведения человека). Группа участников симпозиума состояла главным образом из специалистов в области психологии, неврологии и психиатрии. Нейман выступил на симпозиуме с докладом «Общая и логическая теория автоматов» <The General and Logical Theory of Automata>. Кроме доклада Неймана, на симпозиуме было заслушано и обсуждено еще шесть докладов, в том числе доклад У. Маккаллока <W.S.McCulloch>, профессора психиатрии Иллинойского медицинского колледжа и Иллинойского невропсихиатрического института, на тему «Почему разум сосредоточен в голове» и доклад У. Хальстеда <W. C. Halstead>, профессора экспериментальной психологии Чикагского университета, на тему «Мозг и ум». Все доклады и дискуссии, имевшие место на симпозиуме, были опубликованы в названной выше книге, изданной под редакцией проф. Джеффриса, организатора симпозиума. Доклад Неймана на симпозиуме был перепечатан (с незначительным сокращением) в упомянутом выше (стр. 102) томе «Мира математики». Перевод работы Неймана для настоящего издания сделан с текста издания Ньюмана.

С другой стороны, вторая часть нашей проблемы является задачей, которая, вероятно, привлечет тех из нас, кто имеет подготовку и вкусы математика или логика. В связи с этим мы склонны отвлечься от первой части проблемы с помощью метода аксиоматизации и сосредоточить свое внимание на второй.

Аксиоматическая процедура

Аксиоматизация поведения элементов означает следующее. Мы принимаем, что элементы имеют некоторые вполне определенные внешние функциональные характеристики, т. е. что их следует считать «черными ящиками». Это означает, что их рассматривают как автоматы, внутреннюю структуру которых нет необходимости раскрывать и которые, по предположению, реагируют на некоторые точно определенные раздражители (стимулы) посредством некоторых точно определенных реакций.

Установив это, мы можем перейти к изучению более сложных организмов, которые можно построить из этих элементов,— их структуры, функционирования, связей между элементами и общих теоретических закономерностей, которые можно обнаружить в том сложном синтезе, который представляют собой рассматриваемые организмы.

Нет необходимости подчеркивать ограниченность этой процедуры. С помощью исследований этого типа можно показать, что применяемая система аксиом удобна и по крайней мере в отношении тех результатов, которые она дает, соответствует действительности. Однако такой подход не дает идеального, а возможно даже, и достаточно эффективного метода проверки правильности этих аксиом. Установление верности аксиом относится главным образом к первой части проблемы. В действительности оно предполагает определение — с помощью соответствующих физиологических (или химических, или физико-химических) средств — природы и свойств элементов организма.

Значимые порядки величин

И все-таки, несмотря на эту ограниченность, «вторая часть» нашей проблемы, как указывалось выше, является важной и трудной. При любом разумном определении того, что следует считать элементом, естественные организмы предстанут перед нами как весьма сложные агрегаты таких элементов. Число клеток в человеческом теле имеет порядок 10^{15} или 10^{16} . Число нейронов в центральной нервной системе приблизительно равно 10^{10} . У нас нет совершенно никакого предшествующего опыта относительно систем такой степени сложности. Все искусственные автоматы, созданные человеком, состоят из частей, число которых при любом сравнительно схематическом подсчете имеет порядок от 10^8 до 10^6 . К тому же те искусственные системы, которые работают с той степенью логической гибкости и автономии, какую мы обнаруживаем в естественных организмах, находятся отнюдь не на вершине этой шкалы. Прототипами таких систем служат современные вычислительные машины, и в этом случае любое разумное определение того, что следует рассматривать в качестве элемента, приводит к величине, лишь в несколько раз большей чем 10^8 или 10^4 .

II. НЕКОТОРЫЕ ЧЕРТЫ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

Вычислительные машины. Типичные операции

После этих общих замечаний разрешите мне быть более определенным и перейти к той части предмета, о которой я могу говорить уже подробно, с привлечением данных специального характера и технических деталей. Как я уже отмечал, эта часть проблемы

связана с искусственными автоматами и особенно с вычислительными машинами. Последние имеют некоторое сходство с центральной нервной системой или по крайней мере с определенной стороной ее функций. Разумеется, вычислительная машина гораздо проще нервной системы, проще в том смысле, который имеет здесь значение. Тем не менее представляет определенный интерес проанализировать проблему организмов и организации с точки зрения этих относительно простых искусственных автоматов и установить черты сходства последних с центральной нервной системой в этой, так сказать, «лягушачьей перспективе».

Начну с некоторых утверждений относительно вычислительных машин как таковых.

Идея использовать в целях вычисления автомат относительно нова. Хотя вычислительные автоматы не являются наиболее сложными искусственно созданными автоматами, если рассматривать их с точки зрения тех окончательных результатов, которые можно получить с их помощью, тем не менее они отличаются наивысшей степенью сложности в том смысле, что производят наиболее длинную цепь событий, следующих друг за другом и определяющих одно другое.

В настоящее время имеются вполне определенные и обоснованные соображения относительно того, когда имеет смысл применять быстродействующие вычислительные машины, а когда нет. Этот критерий обычно выражается через число умножений, входящих в математическую проблему. Считают, что применение быстродействующей вычислительной машины в общем оправдывает себя, если вычислительная задача содержит миллион или более последовательных умножений.

В более фундаментальных логических терминах это выражается следующим образом. В рассматриваемых областях (т. е. в тех частях — обычно прикладной — математики, где используются такие машины) математический опыт указывает на желательность доведения точности вычислений приблизительно до десяти десятичных разрядов.

Одно умножение представляется поэтому требующим по меньшей мере $10 \cdot 10$ шагов (цифровых умножений), откуда следует, что миллион умножений требует по меньшей мере 10^8 операций. В действительности, однако, умножение двух десятичных знаков не является элементарной операцией. Существуют различные способы разбиения его на элементарные операции, и все эти способы имеют приблизительно одинаковую степень сложности. Наиболее простой способ оценки степени сложности умножения двух чисел в десятичной системе счисления состоит в следующем: вместо того, чтобы учитывать число десятичных разрядов, рассматривают число разрядов, которое требуется в двоичной системе счисления (с основанием 2 вместо 10) при той же точности вычислений. Одна десятичная цифра соответствует приблизительно трем двоичным, следовательно, десять десятичных знаков соответствуют приблизительно тридцати двоичным. В силу этого умножение, о котором говорилось выше, состоит не из $10 \cdot 10$, а из $30 \cdot 30$ элементарных шагов, т. е. не из 10^2 , а из 10^3 шагов. (Двоичные цифры подчиняются принципу «да или нет», «все или ничего», так как имеют только значения 0 или 1. Поэтому их умножение является действительно элементарной операцией. Заметим, кстати, что эквивалентом 10 десятичных знаков является, более точно, не 30, а 33 двоичных знака, но $33 \cdot 33$ приблизительно равно 10^3 .) Отсюда, следовательно, вытекает, что имеется больше оснований считать, что миллиону умножений (в указанном выше смысле) соответствует 10^9 элементарных операций.

Требования точности и надежности

Я не знаю никакой другой области человеческой деятельности, где результат действительно зависит от последовательности, состоящей из миллиарда (10^9) шагов в каком бы то ни было искусственном устройстве, и где к тому же каждая операция на самом деле имела бы значение для результата (или по крайней мере могла бы иметь для него значение со значительной вероятностью). А именно эта ситуация имеет место в

области вычислительных машин, именно это является их наиболее характерной чертой, вызывающей наибольшие затруднения.

Действительно, за последние два десятилетия созданы автоматы, которые выполняют сотни миллионов или даже миллиарды операций, прежде чем выдают результат. Однако действие этих автоматов не последовательное. Такое большое количество операций является следствием того, что по разным соображениям желательно проделать один и тот же эксперимент несколько раз. Такие кумулятивные, повторяющиеся, накапливающиеся процедуры могут, например, повышать «вес» результата, т. е. (и это существенный аспект) увеличивать значимый - результат — «сигнал» — по сравнению с искажающим его «шумом». Так, например, при любом разумном методе подсчета числа реакций, происходящих в микрофоне до того, как будет произведен членораздельный звуковой сигнал, мы придем к величине, равной многим десяткам тысяч. Аналогичные оценки для телевидения приводят к величине, равной десяткам миллионов, а в случае радарной установки, возможно, даже многим миллиардам. Для любого из автоматов такого рода справедливо, однако, следующее: влияние ошибок, которые автомат совершает в своей работе, на окончательный результат обычно соответствует той доле, которую ошибочные шаги составляют в общем числе шагов, осуществленных автоматом. (Для всех примеров, которые приведены выше, это не совсем верно, однако это утверждение достаточно для качественной характеристики ситуации). Таким образом, чем большее число операций требуется для того, чтобы получить результат, тем меньше оказывается значимость каждой операции в отдельности.

К вычислительным машинам это правило не применимо. Каждая операция в них является (или может оказаться) столь же существенной, что и результат в целом; любая ошибка может сделать полностью недействительным весь результат. (Справедливость этого утверждения не абсолютна; однако, по-видимому, почти 30% всех операций, которые обычно применяются в вычислительных машинах, — такого типа.) Таким образом, вычислительные машины занимают исключительное место среди продуктов человеческой деятельности. Помимо того, что они должны быстро выполнять миллиарды и более операций, от них требуется, чтобы в значительной части вычислительной процедуры (часть эта строго определяется заранее) не было ни одной погрешности.

Действительно, для того чтобы быть уверенным в том, что машина в целом исправна и не имеет дефектов, могущих расстроить всю ее работу, существующая практика обычно требует, чтобы во всей процедуре не встречалось ни единой погрешности.

В силу этого требования большие, высокосложные вычислительные машины предстают перед нами в совершенно новом свете. Это делает, в частности, сравнение вычислительных машин с деятельностью живых организмов не лишенным определенного смысла.

Принцип моделирования

Все вычислительные автоматы могут быть разделены на два больших класса на основании признака, который непосредственно очевиден и который, как вы увидите несколько ниже, переносится и на живые организмы. Это классы моделирующих и цифровых машин.

Рассмотрим сначала принцип моделирования. Действие вычислительной машины может основываться на представлении чисел посредством определенных физических величин. В качестве таких величин мы можем, например, использовать силу электрического тока, величину электрического потенциала или число градусов угла поворота диска (возможно, с учетом числа полных оборотов, сделанных диском) и т. д. Тогда такие операции, как сложение, умножение и интегрирование, можно производить, подбирая различные физические процессы, воздействующие на эти величины нужным образом. Токи можно перемножать, пропуская их по обмоткам обычного ваттметра; если

якорь ваттметра связать с реостатом, то за счет поворота якоря (возникающего в результате прохождения перемножаемых токов по его обмоткам) можно получить изменение сопротивления реостата; наконец, включив этот реостат в цепь, в которой некоторый независимый источник тока поддерживает постоянную разность потенциалов, можно изменение сопротивления преобразовать в изменение тока в этой цепи. Весь агрегат представляет собой «черный ящик», в который вводятся два тока и который выдает ток, равный их произведению. Разумеется, вам известны и многие другие способы, с помощью которых разнообразные естественные процессы можно использовать для выполнения умножения и многих других математических операций.

Первой хорошо смонтированной большой вычислительной машиной, которая когда-либо была сделана, явилась моделирующая машина — дифференциальный анализатор *В. Буша*. Кстати сказать, эта машина производила вычисления не с помощью электрических токов, а с помощью вращающихся дисков. Я не буду рассматривать остроумные приемы, позволявшие различным образом сочетать углы поворота дисков машины в соответствии с характером выполняемых математических операций.

Я не буду пытаться перечислять, классифицировать или систематизировать все многообразие принципов моделирования и механизмов, которые можно использовать для вычислений. Их чрезвычайно много. Руководящим принципом, без которого невозможно понять ситуацию, является классический принцип всей теории передачи сообщений *<communication theory>* — принцип увеличения отношения «сигнал/шум». Это означает, что во всякой процедуре моделирования критическим является вопрос, как велики неконтролируемые флуктуации механизма, составляющие «шум», по сравнению с значимыми «сигналами», выражающими числа, над которыми оперирует машина. Пригодность любого принципа моделирования зависит от того, насколько малой возможно сделать относительную величину неконтролируемых флуктуации — «уровень шума», применяя данный принцип.

Выразим это иначе. Не существует ни одной моделирующей машины, которая действительно давала бы произведение двух чисел. То, что она дает, — это произведение плюс небольшая, но неизвестная величина, которая выражает случайный шум механизма и протекающих в нем физических процессов. Вся проблема заключается в том, чтобы понизить эту величину. Вся техника, связанная с переработкой информации, подчиняется этому принципу. Например, наличие этого принципа привело к тому, что получили распространение сложные и неуклюжие на вид механические устройства вместо более простых и изящных электрических. (Такое положение существовало, по крайней мере, на протяжении большей части последних 20 лет. Сравнительно недавно в некоторых приложениях, где требуется лишь весьма ограниченная точность, электрические устройства вновь вышли на первый план.) Если сравнивать механические и электрические моделирующие процессы, то в общих чертах справедливо следующее. Механические устройства могут доводить отношение уровня шума к «максимальному уровню сигнала» до величины порядка $1 : 10^4$ или $1 : 10^5$. В электрических устройствах это отношение редко бывает лучше, чем $1 : 10^2$. Разумеется, эти отношения характеризуют ошибки в элементарных операциях вычисления, а не конечные результаты. Очевидно, что в последних ошибки будут существенным образом больше.

Цифровой принцип

Цифровая машина работает по хорошо известному методу представления чисел как совокупностей цифр. Это, кстати, та же процедура, которую все мы применяем в наших индивидуальных механических вычислениях, когда мы выражаем числа в десятичной системе счисления. Строго говоря, нет необходимости проводить вычисление непременно в десятичной системе. В качестве основания для цифрового обозначения чисел можно использовать любое целое число, большее единицы. Десятичная система (с основанием

10) является наиболее употребительной, и все цифровые машины, построенные до настоящего времени, работают на основе этой системы. Однако кажется вероятным, что в конце концов будут доказаны преимущества двоичной системы (с основанием 2). В настоящее время строится ряд цифровых машин, в которых используется двоичная система³⁷. Основными операциями в цифровой машине обычно являются четыре действия арифметики: сложение, вычитание, умножение и деление. На первый взгляд может показаться, что цифровая машина выполняет эти действия с абсолютной точностью (в отличие от моделирующих машин, о которых говорилось выше). Однако, как мы увидим ниже, это неверно.

Возьмем умножение. Цифровая машина, производя умножение одного десятизначного числа на другое десятизначное число, даст двадцатизначное число — их произведение — без какой бы то ни было ошибки. На этом этапе ее точность абсолютна, даже несмотря на то, что электрические и механические компоненты арифметического органа машины как таковые обладают ограниченной точностью. До тех пор, пока не вышла из строя какая-нибудь компонента, т. е. до тех пор, пока действие каждой компоненты сопровождается лишь флуктуациями в допустимых пределах, результат будет абсолютно точным. В этом, разумеется, состоит значительное и характерное преимущество цифровой процедуры. Ошибка как явление, сопровождающее нормальную операцию, а не только лишь (как указывалось выше) в виде случайности, происходящей вследствие какой-то конкретной неисправности, тем не менее вкрадывается; это происходит следующим образом. Абсолютно точным произведением двух 10-значных чисел является 20-значное число. Если машина построена для того, чтобы работать только с 10-значными числами, ей приходится отбрасывать последние 10 цифр этого 20-значного числа и оперировать только с первыми десятью знаками. (Здесь можно не рассматривать то небольшое, хотя и распространенное в практике уточнение, которое дает так называемое «округление» числа.) Если же машина может оперировать с 20-значными числами, то умножение двух таких чисел даст 40 знаков, и снова 20 цифр придется отбросить и т. д. (Это означает, что, каково бы ни было максимальное число знаков, на которое рассчитана машина, в процессе последовательных умножений этот максимум рано или поздно будет достигнут. И как только он будет достигнут, следующее умножение даст число цифр, которое превосходит этот максимум, и в произведении будет сохранена лишь половина его цифр [первая половина цифр, соответствующим образом округленная]. Поэтому ясно, что ситуация с максимальным числом знаков, равным 10, типична, и мы с полным основанием можем воспользоваться ею для того, чтобы разъяснить суть дела.)

Таким образом, необходимость округления (точного) 20-значного произведения до установленного (максимального) числа из десяти цифр приводит в случае цифровых машин к качественно той же ситуации, какая была обнаружена выше в случае моделирующих машин. То, что выдает машина в качестве результата, когда требуется найти произведение, есть не само произведение, а произведение плюс малый добавочный член — ошибка округления. Понятно, что эта ошибка не является случайной величиной, подобной, например, «шуму» в моделирующей машине. Арифметически она полностью определена в каждом отдельном случае. Тем не менее способ определения этой ошибки настолько сложен, а ее варьирование в ряде случаев столь незакономерно, что обычно можно с высокой степенью приближения рассматривать ее как случайную величину.

(Эти рассуждения относятся к умножению. В случае деления ситуация оказывается даже немного хуже, так как частное — в общем случае — нельзя выразить с абсолютной точностью, какое бы конечное число цифр ни брать. Следовательно, в случае деления округление обычно оказывается необходимым уже после первой операции. С другой стороны, для сложения и вычитания эта трудность не возникает. Сумма или разность

³⁷ Большинство современных универсальных вычислительных машин работает по двоичной системе. (Стр. 67.)

имеет [если величины не превышают установленного максимума] то же число цифр, что и сами слагаемые. Переполнение разрядной сетки может создавать трудности, которые добавляются к рассмотренным выше, но здесь я не буду их рассматривать.)

Роль цифрового метода в понижении уровня шума

Существенное различие между уровнем шума цифровой машины и уровнем шума моделирующей машины вовсе не является качественным. Оно носит количественный характер. Как отмечалось выше, относительный уровень шума в моделирующей машине никогда не бывает ниже $1 : 10^5$, а во многих случаях доходит до $1 : 10^2$. В 10-разрядной цифровой машине, о которой речь шла выше, относительный уровень шума (являющийся результатом округления) равен $1 : 10^{10}$. Таким образом, практическое значение цифрового метода состоит в том, что он уменьшает уровень вычислительного шума до такого предела, который недостижим ни при какой другой (моделирующей) процедуре. Кроме того, дальнейшее понижение уровня шума сопряжено в случае моделирующего механизма с увеличением трудностей, между тем как в случае цифровой машины оно достигается со все возрастающей легкостью. В моделирующей машине легко достигнуть точности $1 : 10^3$. Точность $1 : 10^4$ достигается с некоторым трудом, точность же $1 : 10^5$ получить уже очень трудно, а $1 : 10^6$ — при современном состоянии техники — невозможно. В случае цифровой машины такое же уменьшение уровня шума достигается всего лишь за счет того, что строят машину, рассчитанную на 3, 4, 5 и 6 десятичных разрядов соответственно. В этом случае переход от одной ступени к следующей действительно становится все более легким. Переход от машины, рассчитанной на 3 разряда (если бы кто-нибудь захотел построить такую машину), к 4-разрядной машине означает возрастание сложности на 33%, переход от 4-разрядной машины к машине, рассчитанной на 5 разрядов, означает увеличение еще на 25%, а переход от 5-разрядной машины к 6-разрядной — на 20%. Переход от 10 разрядов к 11 разрядам означает увеличение сложности всего лишь на 10%. Ясно, что с точки зрения уменьшения «случайного шума» здесь обстановка совершенно иная, чем в случае моделирующей процедуры, основанной на применении физических процессов. Именно в этом — а не в его абсолютной надежности, которая практически недостижима, — кроется важность цифрового метода.

III. СРАВНЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН С ЖИВЫМИ ОРГАНИЗМАМИ

Смешанный (моделирующе-цифровой) характер живых организмов

Изучение центральной нервной системы обнаруживает в ней наличие элементов обеих процедур — цифровой и моделирующей.

Нейрон передает импульс. По-видимому, в этом состоит основная его функция, хотя последнее слово относительно этой функции и о том, ограничивается ли ею роль нейрона в центральной нервной системе, еще далеко не сказано. Нервный импульс в основном подчиняется принципу «включено — выключено», «все или ничего», и его можно сравнивать с двоичной цифрой. Таким образом, наличие цифрового элемента очевидно, но также очевидно, что это еще не все. Многое из того, что происходит в организме, обусловлено не явлениями этого рода, а зависит от общего химического состава крови и других гуморальных сред. Хорошо известно, что в организме имеется множество сложных функциональных цепей, в которых переход от первоначального раздражения к конечному эффекту осуществляется через целый ряд этапов; некоторые из этих этапов являются нейронными, т. е. цифровыми, другие гуморальными, т. е. моделирующими. Цифровые и

моделирующие участки такого рода цепи могут различным образом чередоваться. В некоторых случаях этого типа цепь может фактически иметь обратную связь «на себя», т. е. стимул на ее выходе может в свою очередь оказывать воздействие на ее начальный вход.

Хорошо известно, что такие смешанные (частично нейронные, а частично гуморальные) цепи обратной связи могут порождать весьма важные процессы. Так, механизм, поддерживающий постоянство кровяного давления, является механизмом смешанного типа. Нерв, который воспринимает давление крови и сигнализирует о его величине, осуществляет это в виде последовательности нейронных импульсов, т. е. цифровым способом. Мышечное сокращение, вызванное этой системой импульсов, можно описать как результат наложения многих цифровых импульсов. Однако влияние этого сокращения на ток крови является гидродинамическим, и следовательно, моделирующим. Воздействие изменившегося давления на нерв, сигнализирующий о давлении крови, замыкает кольцо обратной связи, и в этом месте моделирующая процедура вновь превращается в цифровую. Следовательно, в этом пункте аналогия между живыми организмами и вычислительными машинами явно не полная. Живые организмы являются очень сложными — частично цифровыми, а частично моделирующими — механизмами. Вычислительные же машины, по крайней мере, в том виде, какой они имели до настоящего времени (и из которого я исхожу в настоящем изложении), являются чисто цифровыми. Мне приходится поэтому просить вас принять следующее весьма сильное упрощение системы. Хотя я прекрасно отдаю себе отчет в том, что в живом организме имеются компоненты, работающие по принципу моделирования, и отрицать их значение было бы нелепо, тем не менее для простоты рассмотрения я опущу эту часть вопроса и буду рассматривать живые организмы так, как если бы они были чисто цифровыми автоматами.

Смешанный характер каждого элемента

В добавление к сказанному можно привести аргументы в пользу того, что даже нейрон не является строго цифровым органом. Такие соображения высказывались неоднократно и с большой настойчивостью. Тщательный анализ фактов показывает, что в них, несомненно, содержится большая доля истины. Соображения, о которых идет речь, заключаются в том, что полностью развитый нервный импульс, которому можно приписать характер процесса, протекающего по принципу «все или ничего», является не элементарным, а весьма сложным явлением. Такой импульс представляет собой вырожденное состояние сложного электрохимического комплекса, каковым является нейрон, который — если провести, полный анализ его функционирования — следует рассматривать как моделирующую машину. Действительно, можно возбудить нейрон так, чтобы не произошло скачка, вызывающего возникновение нервного импульса. Рассматривая эту область «подпорогового возбуждения», мы видим, что вначале (т. е. для самых слабых раздражений) реакции пропорциональны раздражениям, а затем (при более сильных, но все еще подпороговых уровнях раздражения) реакции начинают подчиняться более сложному нелинейному закону, оставаясь тем не менее непрерывными переменными и не приобретая скачкообразного характера. Кроме того, следует учесть, что имеются и другие сложные явления как внутри подпороговой области, так и вне ее: усталость, суммация, некоторые виды автоколебаний.

Несмотря на правильность этих замечаний, следует учитывать, что они, быть может, предъявляют чрезмерно строгие требования к органу, работающему по принципу «все или ничего». Электромагнитное реле и электронная лампа, при правильном их использовании, являются, несомненно, органами, работающими по принципу «все или ничего». В самом деле, электромагнитное реле и электронная лампа служат прототипом такого рода органов. И тем не менее в действительности оба являются сложными моделирующими

механизмами, которые при надлежаще подобранном возбуждении реагируют непрерывно (линейно или нелинейно), обнаруживая явления типа «скачка», или «все или ничего», лишь при весьма частных условиях работы. Различие между действием этих механизмов и описанным выше поведением нейронов невелико. Выразим это несколько иначе. Ни один из этих механизмов не является органом, действующим исключительно по принципу «все или ничего» (наш опыт в технике и физиологии недостаточен для того, чтобы утверждать, что существуют органы, работающие исключительно по этому принципу), однако это несущественно. Под органом типа «все или ничего» мы понимаем орган, который удовлетворяет двум следующим условиям. Во-первых, он является механизмом, который при некоторых подходящих условиях действует по принципу «все или ничего». Во-вторых, эти условия являются условиями его нормального использования; в функциональном отношении они представляют собой нормальное состояние того большого организма, частью которого является этот орган. Таким образом, важно не то, удовлетворяет ли орган обязательно и при всех условиях принципу «все или ничего» — вероятно, так никогда и не бывает,— а то, являются ли его функции в надлежащей обстановке в основном функциями типа «все или ничего» и выступает ли он в качестве органа, предназначенного в основном для работы в этом режиме. Я отдаю себе отчет в том, что это определение привносит весьма нежелательные критерии «надлежащей» обстановки и «предназначения». Однако я не вижу, каким образом можно избежать их употребления или не учитывать того, что их применение основано просто на здравом смысле. В соответствии со сказанным выше я в дальнейшем буду пользоваться в качестве рабочей гипотезы допущением, что нейрон является цифровым органом типа «все или ничего». Я сознаю, что последнее слово в этом вопросе еще не сказано; однако надеюсь, что приведенные выше соображения, касающиеся ограниченности этой рабочей гипотезы, а также оснований ее применения, успокоят вас. Моя цель — лишь упростить рассуждения, и я не пытаюсь предрешать ни один из существенно важных вопросов, решение которых еще не найдено.

В том же смысле я полагаю, что нейроны допустимо рассматривать как электрические органы. Раздражение нейрона, развитие и протекание его импульса, а также воздействие этого импульса на синапс³⁸ — все это может быть описано электрически. Что же касается химических реакций и других явлений, сопутствующих этому процессу, то они важны для того, чтобы понять внутренний механизм функционирования нервной клетки. Быть может, они даже более важны, чем электрические явления. Однако вряд ли они необходимы для описания нейрона как «черного ящика» — органа типа «все или ничего». Кроме того, в этом случае ситуация ничуть не хуже, чем, скажем, в случае электронной лампы. В электронной лампе чисто электрические явления тоже сопровождаются многочисленными другими явлениями, относящимися к области физики твердого тела, термодинамики, механики. Все они важны для понимания устройства электронной лампы, но их лучше исключить из рассмотрения, если последнюю рассматривать как «черный ящик», задаваемый схематическим описанием.

Понятие о переключательном, или релейном, органе

Нейрон и электронная лампа, рассматриваемые с изложенных выше точек зрения, служат двумя примерами того, что принято обозначать терминами «переключательный орган» или «релейное устройство». (Разумеется, электромеханическое реле является

³⁸ *Синапсы* — места контакта нервных клеток (нейронов) друг с другом; представляют собой промежуточные слои между (а) концевыми окончаниями разветвлений нервного волокна, отходящего от одной нервной клетки, и (b) нервным волокном или телом другой нервной клетки обеспечивающие передачу нервного импульса от первой клетки ко второй клетке. Процесс возбуждения проходит через синапс только в одном направлении. Большинство нервных клеток через синапсы связаны каждая с концевыми образованиями волокон целого ряда других нервных клеток. (Стр. 73.)

другим примером.) Такое устройство определяют как «черный ящик», который в ответ на определенные стимулы или комбинацию стимулов дает энергетически независимую от них реакцию. Это означает, что энергия реакции предполагается достаточной для того, чтобы вызвать несколько стимулов того же рода, что и тот стимул, который вызвал ее. Следовательно, энергия реакции не может быть, получена от первоначального стимула. Она должна исходить от иного, независимого, источника энергии. Стимул лишь направляет и регулирует поток энергии от этого источника.

Таким источником в случае нейрона является его общий метаболизм³⁹. В случае электронной лампы это энергия, которая поддерживает разность потенциалов между катодом и анодом (независимо от того, находится ли лампа в состоянии проводимости или заперта), и, в меньшей степени, тепловая энергия, удерживающая «горячие электроны» вне катода. В случае электромеханического реле это генератор, создающий ток в той цепи, которая замыкается и размыкается с помощью реле.

Основными переключательными органами живых организмов являются — по крайней мере в той степени, в которой они здесь рассматриваются,— нервные клетки, нейроны. Основными переключательными органами вычислительных машин современного типа служат электронные лампы; в более старых машинах переключательные органы — полностью или частично — представляли собой электромеханические реле. Весьма возможно, что вычислительные машины не всегда будут агрегатами, состоящими преимущественно из переключательных органов, однако новый этап развития в этой области пока принадлежит далекому будущему. Шаг вперед, которого можно ожидать в более близкое время, будет состоять, по-видимому, в том, что электронные лампы перестанут использоваться в качестве переключательных органов в вычислительных машинах и будут заменены какими-либо другими элементами. Но и это, вероятно, произойдет не ранее чем через несколько лет⁴⁰. Поэтому я буду рассматривать эти машины исключительно как агрегаты электронных ламп, играющих роль переключательных органов.

Сравнение размеров больших вычислительных машин и живых организмов

Известны две очень большие действующие электронно-ламповые вычислительные машины. Каждая из них содержит около 20 000 переключательных органов. Одна из них — чисто электронноламповая (она принадлежит Баллистической научно-исследовательской лаборатории Управления артиллерийско-технического снабжения армии США и находится в Абердине, штат Мэриленд, обозначение ЭНИАК), другая машина — смешанного типа: она содержит и электронные лампы, и электромеханические реле (она принадлежит компании ИБМ и находится в Нью-Йорке; обозначение ССЭК)⁴¹.

³⁹ *Метаболизм* — обмен веществ. (Стр. 74.)

⁴⁰ В настоящее время все более широкое применение в качестве переключательных органов цифровых вычислительных машин получают полупроводниковые элементы (полупроводниковые диоды, полупроводниковые триоды — транзисторы и пр.) и магнитные элементы; на этих элементах построен целый ряд современных вычислительных машин. Ведутся интенсивные исследования с целью расширения круга устройств, которые можно применять в вычислительных машинах для построения переключательных органов. Так, разработаны новые виды магнитных элементов, например магнитные сердечники с управляемым магнитным потоком (так называемые трансфлюксоры), созданы новые полупроводниковые приборы (например, так называемые спазисторы); появились также переключательные элементы, базирующиеся на новых физических принципах, например криотроны, работа которых в режиме «все или ничего» основана на явлении сверхпроводимости. См. А. И. К и т о в и Н. А. К р и н и ц к и й, *Электронные цифровые машины и программирование*, Физматгиз, М., 1958. (Стр. 74.)

⁴¹ *ЭНИАК* — англ. ENIAC, сокращение от «Electronic Numerical Integrator and Calculator» («Электронный цифровой интегратор и вычислитель»). Описание машины имеется в книге: D. R. H a r t r e e, *Calculating Instruments and Machines*, Cambridge, 1950, chapter 7, §7.4–7.6 and 7.8.

По своим размерам эти машины, вероятно, намного больше электронноламповых вычислительных машин, которые появятся в ближайшие годы. По-видимому, машины, которые будут строиться в ближайшее время, будут иметь от 2000 до 6000 переключательных органов каждая. (Такое уменьшение размеров объясняется изменением в нашем подходе к устройству машинной «памяти», которое я здесь не рассматриваю.) Возможно, что в дальнейшем размеры машин снова возрастут, однако на современном уровне техники и теоретического знания *<philosophy>* количество переключательных органов вряд ли превзойдет 10 000 (или, может быть, величину порядка 10 000). Таким образом, величина 10^4 дает правильное представление о порядке числа переключательных органов вычислительной машины.

В противоположность этому — как это следует из данных, полученных различными путями,— число нейронов центральной нервной системы имеет порядок 10^{10} . Я не знаю, насколько правильна эта цифра, но, по-видимому, величина показателя степени отклоняется от действительной в ту или иную сторону не более чем на единицу. Таким образом, бросается в глаза то обстоятельство, что центральная нервная система, по меньшей мере, в миллион раз сложнее *<larger>*, чем самый сложный искусственный автомат, о котором мы можем говорить в настоящее время. Весьма интересно выяснить, чем это объясняется и какие вопросы принципиального характера с этим связаны. Мне кажется, что здесь действительно имеется ряд четких принципиальных проблем.

Существенно важные отношения размеров элементов

Совершенно очевидно, что электронная лампа является гигантом по сравнению с нервной клеткой. Ее физический объем и потребляемая ею энергия приблизительно в миллиард раз больше, чем у нервной клетки. (Разумеется, тут невозможно привести вполне определенные цифры, однако те, которые приведены выше, достаточно хорошо обрисовывают ситуацию.) Это компенсируется другими факторами. В областях техники, отличных от области вычислительных машин, электронные лампы можно заставить работать на чрезвычайно высоких скоростях, однако здесь мы не будем касаться этих областей их применения. В вычислительных машинах максимум скорости гораздо ниже, но все же заслуживает уважения. При современном состоянии техники его можно считать равным в среднем одному миллиону реакций *<actuations>* в секунду. Реакция *<response>* нервной клетки развивается гораздо медленнее, вероятно, она длится $\frac{1}{2000}$ секунды, и — что действительно существенно — минимальное время, необходимое для перехода от возбужденного состояния к полному восстановлению, когда клетка может испытывать повторное возбуждение, еще больше — в лучшем случае оно приблизительно равно $\frac{1}{200}$ секунды. Это приводит к отношению 1 : 5000, которое, возможно, в какой-то мере завышено в пользу электронной лампы, так как электронная лампа при ее использовании в качестве переключательного органа, рассчитанного на 1 000 000 операций в секунду, практически не работает на все 100% в этом режиме. Поэтому такое отношение, как 1 : 2000, вероятно, будет более подходящим. Таким образом, электронная лампа, обладая размерами приблизительно в миллиард раз большими, чем нейрон, превосходит последний в работе в 1000 раз (или несколько более). В силу этого с известным основанием можно говорить о том, что ее эффективность меньше эффективности нейрона примерно в миллион раз.

ССЭК — англ. SSEC, сокращение от «Selective Sequence Electronic Calculator» («Электронная вычислительная машина с выбором последовательности действий»). Описание машины можно найти в указанной выше книге Хартри, гл. 7, § 7.7.

ИБМ (англ. IBM, сокращение от «International Business Machines Corporation») — одна из старейших и крупнейших зарубежных фирм по производству математических инструментов и вычислительных машин (США). (Стр. 75.)

Важным во всех отношениях является тот факт, что нейрон по своим размерам гораздо меньше электронной лампы. Как указывалось выше, электронная лампа приблизительно в миллиард раз больше. Чем это обусловлено?

Причины различия в размерах электронной лампы и нейрона

Источник этого расхождения лежит в основном органе управления, или, точнее, в различии между управляющими устройствами электронной лампы и нейрона. В электронной лампе основной областью управления является пространство между катодом (где зарождаются активные агенты — электроны) и сеткой (которая управляет электронным потоком). Это пространство имеет приблизительно один миллиметр глубины. В нейроне ему соответствует стенка нервной клетки — «мембрана», толщина которой равна примерно одному микрону ($1/1000$ миллиметра) или несколько меньше. Следовательно, на этом этапе отношение линейных размеров управляющих устройств электронной лампы и нейрона составляет приблизительно 1 : 1000. В этом, между прочим, и заключается основное различие. Электрические поля, действующие в пространстве управления, почти одинаковы для электронной лампы и для нейрона. Разности потенциалов, обеспечивающие надежную работу этих органов, равны десяткам вольт в одном случае и десяткам милливольт в другом. Их отношение снова равно 1 : 1000, и, следовательно, градиенты разностей потенциалов (напряженности полей) примерно равны. Далее, отношение линейных размеров 1 : 1000 соответствует отношению объемов 1 : 1 000 000 000. Таким образом, коэффициент различия, равный 10^9 для объемов, соответствует, как и должно быть, коэффициенту различия, равному 10^3 для линейных размеров, т. е. различию между миллиметровой глубиной междуэлектродного пространства электронной лампы и микронной толщиной мембраны нейрона.

Стоит обратить внимание на то (хотя в этом и нет ничего удивительного), что это различие между объектами, каждый из которых является микроскопическим и расположен внутри элементарной компоненты, приводит к поразительному макроскопическому различию между организмами, которые построены на их основе. Это различие между миллиметровым объектом и микронным объектом и обуславливает то, что ЭНИАК весит 30 тонн и потребляет 150 киловатт, тогда как центральная нервная система человека, которая в функциональном отношении в миллион раз сложнее, имеет вес в несколько фунтов и умещается в человеческом черепе.

При оценке веса и размеров машины ЭНИАК мы должны также помнить, что, как указывалось выше, этот гигантский аппарат используется для обработки 20 чисел в 10 десятичных знаков каждое, то есть 200 десятичных цифр всего, что эквивалентно примерно 700 двоичным цифрам — всего-навсего семистам (одновременно обрабатываемым) элементам информации типа «да-нет»!

Связь этих причин с характером современной техники

Приведенные соображения показывают, что наша техника обработки информации с большой скоростью и при высоких степенях сложности еще очень несовершенна. Устройства, которые мы создаем, просто чудовищны как по своим физическим размерам, так и точки зрения потребляемой ими энергии.

По-видимому, эта слабая сторона нашей техники, по крайней мере отчасти, вызывается теми материалами, которые мы применяем. Современная техника использует металлические компоненты, находящиеся близко друг к другу и разделенные в определенных критических пунктах только вакуумом. Такая комбинация сред обладает особой механической неустойчивостью, что совершенно несвойственно живой природе. Говоря это, я имею в виду тот простой факт, что, если живой организм получает

механическое повреждение, он обнаруживает сильную тенденцию к самовосстановлению. Если же мы трахнем кувалдой по сделанному человеком механизму, то никакой такой восстановительной тенденции не проявится. Если два куска металла находятся рядом, то незначительные колебания и другие механические возмущения, всегда существующие в окружающей среде, создают опасность соприкосновения этих кусков металла. Если последние имеют различные электрические потенциалы, то в результате получающегося в этом случае короткого замыкания они могут спаяться, и контакт станет постоянным. Тогда произойдет полное и окончательное разрушение всей структуры. Когда же мы повреждаем мембрану нервной клетки, ничего подобного не случается. Напротив, мембрана, как правило, быстро восстанавливается.

Именно эти отрицательные механические свойства наших материалов препятствуют дальнейшему сокращению размеров искусственных автоматов. Механическая неустойчивость и другие аналогичные явления делают поведение употребляемых нами компонент не вполне надежным даже при их современных размерах. Таким образом, именно то обстоятельство, что применяемые нами материалы уступают по качеству материалам, которыми пользуется природа, не позволяет нам достигнуть той высокой сложности организации — при малых размерах,— какая достигнута естественными организмами.

IV. БУДУЩАЯ ЛОГИЧЕСКАЯ ТЕОРИЯ АВТОМАТОВ

Мы подчеркивали, насколько ограничена возможность усложнения искусственных автоматов — усложнения, которого можно достигнуть без крайних трудностей и при котором тем не менее можно ожидать, что автомат будет работать надежно. Уже были названы две причины, ставящие предел усложнению в этом смысле: это большие габариты составных элементов и ограниченная надежность их работы; обе причины обусловлены тем, что мы используем материалы, которые, хотя и вполне удовлетворительны в более простых случаях, все же малопригодны и уступают по качеству естественным в такой сложной области, как автоматы. Существует и третий важный ограничивающий фактор, и теперь нам следует обратить на него внимание. Этот фактор — интеллектуального, а не физического порядка.

Ограничения, вытекающие из отсутствия логической теории автоматов

Мы весьма далеки от того, чтобы располагать теорией автоматов, которая заслуживала бы этого названия,— надлежащей логико-математической теорией автоматов. На сегодняшний день имеется хорошо разработанная система формальной логики, в частности логики, применяемой к математике. Эта дисциплина имеет много сильных сторон, но ей присущи и некоторые серьезные недостатки. Нет смысла распространяться о ее сильных сторонах, значение которых, разумеется, я несколько не собираюсь умалять. Что же касается ее слабых сторон, то по этому поводу можно сказать следующее. Всякий, кто работал в области формальной логики, подтвердит, что она, рассматриваемая с точки зрения применяемого в ней математического аппарата, является одной из наиболее трудных областей математики. Причина этого состоит в том, что формальная логика имеет дело с жесткими понятиями типа «все или ничего» и весьма мало соприкасается со связанными с непрерывностью понятиями действительного или комплексного числа, т. е. с математическим анализом. А ведь именно анализ обладает наиболее развитым математическим аппаратом и является наиболее разработанной областью математики. Таким образом, формальная логика в силу самого существа своего

подхода отрезана от наиболее разработанных частей математики и попадает в ту ее область, которая представляет наибольшие трудности, — в область комбинаторики.

Рассмотренная выше теория цифровых автоматов, автоматов типа «все или ничего», является, несомненно, главой формальной логики. Поэтому может показаться, что этой теории придется разделить с формальной логикой то непривлекательное свойство, о котором мы только что говорили. С математической точки зрения она должна как будто быть скорее комбинаторной, чем аналитической теорией.

Возможные характеристики логической теории автоматов

По моему мнению, на самом деле этого не случится. Очевидно, что при изучении работы автоматов необходимо обращать внимание на одно обстоятельство, которое раньше никогда не давало о себе знать в формальной логике.

Во всей современной логике единственно важным является вопрос, можно или нельзя получить результат в конечное число элементарных шагов. С другой стороны, число шагов, которое для этого требуется, в формальной логике почти никогда не рассматривается. Любая конечная последовательность правильных шагов принципиально так же хороша, как и любая другая. Не играет никакой роли, каково это число: мало оно или велико или, быть может, столь велико, что соответствующую последовательность шагов нельзя выполнить в течение человеческой жизни или в течение предположительного времени существования звезд. Но когда мы имеем дело с автоматами, этот подход должен быть значительно изменен. Суть дела в том, что в случае автомата играет роль не только то, может ли он вообще дать определенный результат в конечное число шагов, но и вопрос о том, сколько потребуется таких шагов. Это объясняется двумя причинами. Во-первых, мы строим автоматы для того, чтобы иметь возможность получать некоторые результаты в течение определенных, наперед указанных отрезков времени или по крайней мере в течение таких отрезков времени, порядок которых указан заранее. Во-вторых, вероятность ошибки в компонентах автомата, используемых в любой индивидуальной операции, хотя и мала, но все же отлична от нуля. Если цепь операций достаточно длинна, то суммарный эффект вероятностей отдельных ошибок может (при отсутствии контроля) достигнуть порядка единицы, вследствие чего полученный результат становится практически полностью ненадежным. Хотя вероятности, которые мы встречаем в этом случае, очень малы, они все же не слишком далеки от того, что имеется в обычном техническом опыте. Нетрудно подсчитать, что быстродействующая вычислительная машина при решении типичной задачи в среднем должна выполнить около 10^{12} индивидуальных операций. Следовательно, вероятность ошибки в отдельной операции, которую можно еще допустить, должна быть достаточно малой по сравнению с 10^{-12} . Замечу, что в настоящее время признают удовлетворительными электромеханические (телефонные) реле, если вероятность ошибки в отдельной операции имеет порядок 10^{-8} ; они считаются просто превосходными, если этот порядок равен 10^{-9} . Таким образом, степени надежности, которые требуются в быстродействующих вычислительных машинах, выше, но не безгранично выше степеней надежности, прочно вошедших уже в некоторые области технической практики. И тем не менее трудно ожидать, чтобы действительно достижимые степени надежности могли выйти далеко за пределы только что упомянутых минимальных требований. В силу этого несомненно потребуются исчерпывающие исследования и нетривиальная теория.

Таким образом, логика автоматов будет отличаться от существующей системы формальной логики в двух отношениях:

1. В ней должна учитываться действительная длина «цепей рассуждения», т. е. цепей операций.
2. Операции логики (силлогизмы, конъюнкции, дизъюнкции, отрицания и т. д., то есть — в обычной для автоматов терминологии — различные формы вентильных схем,

собираемых схем, схем совпадения, несовпадения и т. д.)⁴² в этой теории должны рассматриваться как процедуры, допускающие погрешности хотя и с малыми, но все же отличными от нуля вероятностями. Все это должно привести к теориям, имеющим не столь явно выраженный характер теорий, основанных на принципе «все или ничего», какой являлась и является формальная логика. По своему характеру эти теории станут значительно менее комбинаторными и значительно более аналитическими. В самом деле, имеются многочисленные признаки, дающие основание полагать, что эта новая концепция формальной логики будет все более сближаться с другой дисциплиной, в прошлом мало связанной с логикой. Эта дисциплина — термодинамика, главным образом в том виде, который она приняла после Больцмана. Термодинамика является той частью теоретической физики, которая в некоторых из своих аспектов наиболее близка теории обработки и измерения информации. Ее средства, конечно, являются в гораздо большей степени аналитическими, нежели комбинаторными, что вновь подтверждает точку зрения, которую я пытался изложить выше. Однако более подробное рассмотрение этого вопроса завело бы меня слишком далеко.

Все это еще раз подчеркивает сделанный ранее вывод о том, что необходима детальная математическая теория автоматов и информации, которой в большей степени были бы присущи аналитические методы. В настоящее время мы располагаем только зачатками такой теории. Когда мы оценивали искусственные автоматы, о которых я говорил выше, отличающиеся ограниченной сложностью, мы могли действовать грубо эмпирически, не опираясь на такую теорию. Имеются все основания полагать, что в случае более сложных автоматов такой путь окажется невозможным.

Как влияет отсутствие логической теории автоматов на процедуру обращения с ошибками

То, что мы не располагаем логической теорией автоматов, является последним и весьма важным ограничивающим фактором. Трудно поверить, чтобы мы, не располагая весьма тонкой и развитой теорией автоматов и информации, могли в будущем создавать автоматы намного сложнее тех, которые имеются у нас теперь. Тем более это немыслимо в отношении автоматов, обладающих той чудовищной степенью сложности, какая присуща центральной нервной системе человека.

Это отставание в области теории несомненно мешает нам продвинуться вперед существенно дальше того, где мы находимся сейчас.

Простым проявлением этого фактора является наш подход к предупреждению ошибок. В живых организмах случаются нарушения работы компонент. Очевидно, что организм располагает средствами для того, чтобы выявить и обезвредить эти нарушения. Легко подсчитать, что число действий нерва, происходящих в течение жизни нормальной продолжительности, имеет порядок 10^{20} . Ясно, что в продолжение этой цепи событий не встречается таких нарушений, вредные последствия которых не могут быть компенсированы самим организмом без какого бы то ни было существенного

⁴² *Операции логики (силлогизмы, конъюнкции, ...)*. Речь идет об операциях так называемой классической логики суждений — части логики, в которой суждения рассматриваются только с точки зрения их свойства быть либо истинными, либо ложными, и притом только одно из двух. Операции логики суждений без труда моделируются в автоматах, для чего, в частности, используются схемы, о которых упомянуто в тексте. Если автомат является электрическим устройством, то истинности суждения соответствует обычно наличие импульса, а ложности — его отсутствие. Об операциях логики суждений см., например: П. С. Н о в и к о в , *Элементы математической логики*, Физматгиз, М., 1959, гл. 1; А. Т а р с к и й , *Введение в логику и методологию дедуктивных наук*, перев. с англ., ИЛ, М., 1948, гл. 2. О моделировании операций логики суждений на машинах см., например: Н. А. А р х а н г е л ь с к и й и Б. И. З а й ц е в , *Автоматические цифровые машины*, Физматгиз, М., 1958, стр. 43–51; А. И. К и т о в и Н. А. К р и н и ц к и й , *Электронные цифровые машины и программирование*, Физматгиз, М., 1959, гл. 2; И. А. П о л е т а е в , *Сигнал*, Изд. «Советское радио», М., 1958, гл. 7. (Стр. 81.)

вмешательства со стороны. Поэтому система организма должна содержать необходимые приспособления для диагностики ошибок, когда последние возникают, для перестройки организма, имеющей целью свести к минимуму эффект этих ошибок, и, наконец, для исправления или полного выключения вышедших из строя компонент. Наш *modus procedendi* в отношении нарушений в работе искусственных автоматов совершенно другой. Существующая практика, которой единодушно придерживаются все специалисты в этой области, состоит приблизительно в следующем. Мы используем все средства (математической или автоматической проверки) для того, чтобы выявить каждую ошибку, поскольку она возникает в работе автомата. После того, как ошибка выявлена, мы стараемся как можно быстрее изолировать компоненту, вызвавшую ошибку. Частично это можно сделать автоматически, но в любом случае значительную часть процедуры выявления ошибок приходится проводить посредством вмешательства извне. Как только неисправная компонента найдена, она немедленно или исправляется, или заменяется новой, исправной компонентой.

Отметим различие между этими двумя подходами. Основной принцип, которого придерживается природа в своем подходе к нарушениям в функционировании живых организмов, состоит в том, чтобы сделать эффект этих нарушений как можно менее заметным, с тем чтобы организм мог осуществлять необходимые коррективы без всякой спешки, так сказать «на досуге» (если в этом вообще возникает необходимость). Напротив, наш подход к искусственным автоматам предполагает немедленное выявление ошибок, как только они возникают. Поэтому мы стараемся так устроить автоматы, чтобы ошибки были как можно заметнее и чтобы необходимое в этих случаях вмешательство и коррективы можно было осуществлять без промедления. Иначе говоря, устройство живых организмов подчинено принципу: сделать ошибки настолько незаметными и безвредными, насколько это вообще возможно. Что же касается искусственных автоматов, то их проектируют так, чтобы ошибки в их работе как можно более резко проявляли себя своими нежелательными последствиями. Нетрудно найти естественное объяснение такого различия. Природа устроила организмы так, что они сохраняют способность функционирования даже после того, как в них возникли повреждения. Они могут действовать, несмотря на повреждения; при этом после появления повреждений в них проявляется тенденция к их самоустранению. Конечно, можно построить такой искусственный автомат, который мог бы нормально действовать несмотря на некоторые неисправности (число которых, а также области, в которых они допустимы, было бы заранее ограничено). Однако даже и в этом случае каждая неисправность таила бы в себе известную опасность того, что в машине уже начался общий разрушительный процесс. Отсюда возникает необходимость немедленного вмешательства, ибо машина, начавшая работать неправильно, весьма редко обнаруживает тенденцию к самовосстановлению и более вероятно то, что ее функционирование будет все более и более ухудшаться. Все сказанное еще раз подтверждает только одно. В области искусственных автоматов мы, по-видимому, движемся вслепую в гораздо большей степени, чем природа в отношении живых организмов. Мы «запуганы» (что, впрочем, вполне естественно, по крайней мере в настоящее время) страхом перед возможностью даже отдельной изолированной ошибки и перед тем нарушением в работе, которое за ней скрывается. Ясно, что наше поведение — это сверхосторожность, порождаемая невежеством.

Принцип единственной ошибки

Кроме того, почти вся наша методика выявления причин ошибок основана на предположении, что в каждом данном случае, когда налицо ошибка, в машине имеется только одна неисправная компонента. Тогда последовательное разделение машины на части позволяет определить, какой ее отдел содержит неисправную компоненту. Всякий раз, когда существует вероятность того, что в машине имеется несколько неисправностей,

этот — достаточно мощный — дихотомический метод⁴³ раскрытия причин нарушений в работе машин теряет свою силу, и установление причин ошибок становится весьма безнадежным предприятием. То, что мы настойчиво стремимся свести число подлежащих выявлению неисправностей к единице (или во всяком случае сделать это число как можно меньше), еще раз иллюстрирует наше невежество в этой области. Именно это является одной из главных причин того, что приходится устраивать так, чтобы ошибки были как можно более заметными, с тем чтобы опознание ошибок и установление их причин происходило как можно быстрее после их появления — прежде, чем появятся другие ошибки.

V. ПРИНЦИПЫ ЦИФРИЗАЦИИ

Цифризация непрерывных величин: метод цифрового представления и метод счета

Рассмотрим цифровую часть живого организма, а именно нервную систему. По-видимому, действительно имеется основание считать, что нервная система является цифровым механизмом и передает сообщения, состоящие из сигналов типа «все или ничего». (Ср. то, что было сказано нами выше на стр. 71.) Иными словами, каждый элементарный сигнал, каждый импульс просто или имеется в наличии, или его нет (без каких-либо нюансов). Особенно яркой иллюстрацией этого служат те случаи, когда задача, которую приходится решать нервной системе, носит противоположный характер, т. е. когда нервная система должна на самом деле служить для передачи некоторой непрерывной величины. Характерным примером в этом отношении является поведение нерва, передающего данные о величине кровяного давления.

Допустим, например, что нужно передать информацию о величине давления (являющегося, очевидно, непрерывной величиной). Хорошо известно, как это дело осуществляется. Соответствующий нерв передает лишь импульсы типа «все или ничего». Спрашивается: как же с помощью этих импульсов, т. е. цифр, нерв выражает непрерывное численное значение давления? Иными словами, как он кодирует непрерывную величину с помощью цифрового обозначения? Разумеется, он осуществляет это, не представляя рассматриваемую величину в виде числа, выраженного в десятичной (или двоичной, или любой другой позиционной) системе счисления в обычном смысле. Процесс, который, по-видимому, здесь имеет место, состоит в том, что нерв передает импульсы с переменной частотой, которая в определенных пределах пропорциональна рассматриваемой непрерывной величине, а в общем случае является ее монотонной функцией. Поэтому механизм, с помощью которого достигается это «кодирование», в сущности есть частотно-модулирующая система.

Известны подробности этого процесса. Нерв имеет конечное время восстановления. Иначе говоря, после того как прошел один импульс, время, которое должно истечь прежде, чем нерв снова сможет прийти в состояние возбуждения, отлично от нуля и зависит от силы следующего (уже действующего на нейрон) раздражения. Таким образом, если на нерв воздействует непрерывно действующий раздражитель (т. е. раздражитель, действие которого во времени распределяется равномерно; таковым, например, является рассматриваемое нами кровяное давление), то нерв будет реагировать периодически, причем время, протекающее между двумя последовательными состояниями возбуждения, есть упомянутое выше время восстановления нерва, которое является функцией силы постоянного раздражения (в нашем случае — давления). Так, при высоком давлении нерв

⁴³ *Дихотомический метод* (от греч. *dicha* — на две части и *tome* — сечение) — метод, основанный на разделении — мысленном или реальном — объекта исследования на две части. (Стр. 84.)

может давать ответ на раздражение по меньшей мере через каждые 8 миллисекунд, т. е. проводить 125 импульсов в секунду, тогда как под действием меньшего давления он будет реагировать только через каждые 14 миллисекунд, то есть передавать 71 импульс в секунду. Совершенно ясно, что это — поведение настоящего органа типа «да или нет», органа цифрового типа. Однако чрезвычайно поучительно, что работа нерва основана на использовании «счета», а не «десятичного (двоичного, троичного и т. д.) представления» величин.

Сопоставление обоих методов. Предпочтительное использование живыми организмами метода счета

Сравним достоинства и недостатки этих двух методов. Конечно, метод счета менее эффективен, чем метод цифрового разложения. Чтобы посредством счета выразить число порядка одного миллиона (т. е. физическую величину, состоящую из миллиона отличных друг от друга составных элементов), нужно передать миллион импульсов. Чтобы выразить число того же порядка с помощью цифрового представления, потребуется 6 или 7 десятичных цифр, т. е. около 20 двоичных. Следовательно, в этом случае потребуется только 20 импульсов. Таким образом, наш метод цифрового представления является гораздо более экономичным в обозначениях, чем избранный природой метод счета. С другой стороны, метод счета весьма надежен и предохраняет от ошибок. Если вы, выражая число порядка одного миллиона посредством простого счета, пропустите по ошибке один необходимый шаг, результат изменится лишь несущественно. Если же вы выразите то же число с помощью цифрового представления (в десятичной или двоичной системе), то одна-единственная ошибка в одной-единственной цифре может испортить весь результат. Таким образом, отрицательные стороны наших вычислительных машин вновь проявляются в нашей системе представления чисел с помощью цифр; фактически совершенно очевидно, что они глубоко связаны с этой системой и отчасти являются ее следствием. С другой стороны, высокая устойчивость, а также способность устранять ошибки и нарушения в своем функционировании, характеризующие естественные организмы, находят свое отражение в методе счета, которым, по-видимому, они пользуются в этом случае. Все сказанное отражает общее правило. Можно в большой мере обезопасить себя от ошибок, понизив эффективность обозначений, или, точнее говоря, допустив избыточность в обозначениях. Очевидно, что простейший способ добиться надежности за счет избыточности в обозначениях состоит в том, чтобы использовать *per se* совершенно ненадежный метод цифрового представления чисел, но каждое сообщение, выраженное его средствами, повторять по нескольку раз. В рассматриваемом случае природа, очевидно, избрала систему, еще более избыточную в обозначениях и еще более надежную в работе.

Разумеется, следует допустить, что имеются и другие причины, обуславливающие то, что нервная система использует метод счета, а не метод цифрового представления. Кодирование и декодирование происходит гораздо проще в первом случае, чем во втором. Однако справедливо и то, что природа, по-видимому, стремится и может идти гораздо дальше в направлении усложнения, чем идем мы, или, вернее, чем мы можем позволить себе идти. Поэтому можно сомневаться в том, что если бы единственным недостатком системы цифрового представления была ее большая логическая сложность, природа отвергла бы ее единственно по этой причине. Тем не менее справедливо и то, что мы нигде не находим указаний относительно того, что цифровое представление действительно используется в естественных организмах. Трудно сказать, в какой мере это наблюдение «окончательно». Во всяком случае оно заслуживает внимания, и его следует учесть в дальнейших исследованиях деятельности нервной системы.

VI. ФОРМАЛЬНЫЕ НЕРВНЫЕ СЕТИ

Теория формальных нервных сетей Маккаллока–Питтса

Можно было бы еще многое сказать об этих вещах с точки зрения логики и структуры автоматов и живых организмов, но я не буду пытаться сделать это здесь. Вместо этого я перейду к рассмотрению того, что, по-видимому, является наиболее важным результатом, который до сих пор удалось получить с помощью аксиоматического метода. Я имею в виду замечательные теоремы *Маккаллока* и *Питтса* о связи между логикой и нервными сетями.

Как уже говорилось выше, в своих рассуждениях я придерживаюсь строго аксиоматической точки зрения. Поэтому я буду рассматривать нейрон как «черный ящик», имеющий определенное число входов, на которые подаются импульсы, и выход, который отдает импульсы. В целях большей конкретности изложения допустим, что входные связи каждого нейрона могут быть двух типов: возбуждающего и тормозящего. Сами «черные ящики» также могут быть двух типов: с порогом 1 и с порогом 2. Эти понятия связаны между собой и удовлетворяют следующим определениям. Чтобы возбудить такого рода орган, необходимо, чтобы он получил одновременно по крайней мере столько импульсов на своих входах возбуждающего типа, сколько соответствует его порогу, и не получил импульса ни на одном из своих входов тормозящего типа. Возбужденный указанным способом, нейрон после определенного интервала времени — запаздывания (величина которого, по предположению, всегда одинакова и может быть использована для определения единицы времени) — дает выходной импульс. С помощью соответствующих связей этот импульс можно передать на любое число входов других нейронов (так же, как на любой из собственных входов данного нейрона), и он вызовет в каждом из них входной импульс того же типа, как те, которые были описаны выше.

Разумеется, следует иметь в виду, что все это — чрезвычайно сильное упрощение истинной картины функционирования нейронов. Я уже рассматривал характер, ограниченность и преимущества аксиоматического метода (см. стр. [60–61](#) и [67–68](#)). Все, что там было сказано по этому поводу, применимо и в настоящем случае, и дальнейшее изложение следует понимать именно в этом смысле.

Маккаллок и Питтс использовали эти элементы для построения сложных схем, которые можно назвать «формальными нервными сетями». Система такого рода строится из произвольного числа таких элементов, входы и выходы которых надлежащим образом соединяются между собой с произвольной степенью сложности. «Функционирование» такой сети может быть определено посредством выделения некоторых из входов всей системы и некоторых из ее выходов и описания того, какого рода стимулы, воздействующие на выделенные входы, вызывают такие-то стимулы на выделенных выходах.

Основной результат теории Маккаллока–Питтса

Важный результат, полученный Маккаллоком и Питтсом, заключается в том, что всякое функционирование в этом смысле, которое вообще может быть определено логически — строго и однозначно — с помощью конечного числа слов, может также быть реализовано с помощью указанной выше формальной нервной сети.

Здесь уместно остановиться и рассмотреть вытекающие отсюда следствия. Часто можно было слышать разговоры о том, что деятельность и функции нервной системы человека настолько сложны, что никакой обычный механизм не может их выполнить. Пытались указать специфические функции, которые, по самой своей природе, налагают

это ограничение. Делались также попытки доказать, что такого рода специфические функции, полностью описанные логически, *per se*⁴⁴ недоступны механической, нервной реализации. Результат Маккаллока–Питтса кладет всему этому конец. Он доказывает, что все, что можно описать исчерпывающим и однозначным образом, все, что можно полностью и однозначно выразить словами, *ipso facto*⁴⁵ реализуемо с помощью соответствующей конечной нервной сети. Так как обратное утверждение очевидно, мы можем сказать, что не существует различия между возможностью описать словесно, полностью и однозначно, действительный или воображаемый способ поведения и возможностью реализовать этот способ поведения посредством конечной формальной нервной сети. Эти два понятия равнообъемны. Принципиальная трудность выражения всякого способа поведения в такой сети возникает только тогда, когда мы не в состоянии дать полного описания рассматриваемого способа поведения.

Следовательно, остаются две следующие проблемы. Во-первых, если какой-нибудь способ поведения может быть выполнен некоторой конечной нервной сетью, то все же остается открытым вопрос, можно ли практически построить эту сеть; именно, можно ли реализовать ее в рамках данного организма, учитывая те физические ограничения, которые наложены на него природой. Во-вторых, возникает вопрос, возможно ли в действительности дать полное и однозначное словесное описание каждого реального способа поведения.

Разумеется, первая проблема есть основная задача нейрофизиологии, и здесь я не буду вдаваться в ее подробности. Второй вопрос носит иной характер и имеет интересные логические аспекты.

Осмысление этого результата

Нет сомнения в том, что любую отдельную фазу любой мыслимой *<conceivable>* формы поведения можно «полностью и однозначно» описать с помощью слов. Это описание может быть длинным, однако оно всегда возможно. Отрицать это означает примкнуть к разновидности логического мистицизма, от чего большинство из нас, несомненно, далеки. Имеется, однако, существенное ограничение, состоящее в том, что все сказанное применимо только к каждому элементу поведения, рассматриваемому в отдельности, но далеко не ясно, как все это применять ко всему комплексу поведения в целом. Поясню свою мысль на примере. Нетрудно описать, каким образом организм оказывается в состоянии отождествлять любые два треугольника, образованные прямыми линиями, изображение которых появляется на сетчатке, по их принадлежности к одной и той же категории «треугольников». К этому нетрудно добавить требование, чтобы многочисленные другие объекты, помимо правильно начерченных прямолинейных треугольников, также опознавались в качестве треугольников и относились к их различным классам (треугольники с искривленными сторонами, треугольники с не полностью начерченными сторонами, треугольники, которые выделены только более или менее равномерной штриховкой их внутренней части, и т. д.). Чем более полно мы попытаемся описать все, что можно считать треугольником, тем длиннее будет описание. У нас может появиться смутное и неприятное чувство, что полный перечень не только был бы чрезвычайно длинным, но, больше того, — что он по необходимости должен быть бесконечным. И все же такое описание возможно.

Сказанное выше составляет лишь незначительный фрагмент более общего понятия об отождествлении сходных геометрических объектов. А это в свою очередь есть лишь микроскопическая часть общего понятия аналогии. Никто не взялся бы в пределах практически допустимого числа страниц дать описание и определение общего понятия аналогии, которое доминирует в нашем объяснении зрения. Мы не знаем, потребует ли

⁴⁴ Сами по себе; как таковые (лат.) (Прим. ред.)

⁴⁵ В силу самого факта (лат.). (Прим. ред.)

такое предприятие тысячу, или миллион томов, или какое-либо другое их число, которое практически недостижимо. Но зато очень возможно, что простейший и единственно доступный на практике способ показать, что представляет собой явление зрительного сходства, состоит в описании связей, существующих в зрительном аппарате мозга <*the visual brain*>. Здесь нам придется иметь дело с такими разделами логики, в которых у нас практически нет предшествующего опыта. Степень сложности, с которой мы сталкиваемся в этом случае, далеко выходит за рамки всего того, что нам известно. Мы не имеем права считать, что логические обозначения и методы, применявшиеся ранее, могут быть использованы и в этой области. У нас нет полной уверенности в том, что в этой области реальный объект не может являться простейшим описанием самого себя, т. е. что всякая попытка описать его с помощью обычного словесного или формальнологического метода не приведет к чему-то более сложному, запутанному и трудновыполнимому. В самом деле, некоторые результаты современной логики как будто указывают на то, что подобных явлений следует ожидать, когда мы переходим к рассмотрению действительно сложных сущностей. Поэтому отнюдь не исключено, что поиски точного логического понятия «зрительной аналогии» (т. е. точного словесного описания зрительного сходства) являются напрасными. Весьма возможно, что уже сама схема связей в зрительном аппарате мозга является простейшим логическим выражением (или определением) принципа зрительной аналогии.

Очевидно, что на этом уровне в результате Маккаллока–Питтса не содержится более ничего полезного. В этом отношении он представляет собой лишь новую иллюстрацию той ситуации, которая была обрисована выше. Налицо эквивалентность между законами логики и их осуществлением в нервной сети, и, хотя в более простых случаях с помощью этих законов можно получать более простые формы представления для нервных сетей, весьма возможно, что в особо сложных случаях справедливо обратное.

Все это не меняет моей глубокой уверенности в том, что для понимания высокосложных автоматов, и в частности центральной нервной системы, требуется новая существенно-логическая теория. Тем не менее не исключена возможность того, что в ходе этого процесса логика вынуждена будет претерпеть метаморфозу и превратиться в неврологию в гораздо большей степени, чем неврология — в раздел логики. Проведенный выше анализ показывает, что для развития теории центральной нервной системы одну немаловажную вещь можно сделать уже сейчас: именно можно показать, каковы те направления, которые уведут в сторону от действительных проблем.

VII. ПОНЯТИЕ СЛОЖНОСТИ. САМОВОСПРОИЗВЕДЕНИЕ

Понятие сложности

Изложенные выше соображения показали, что фактор сложности играет важную роль во всякой попытке продвинуться вперед в теории автоматов и что понятие сложности, несмотря на его *prima facie*⁴⁶ количественный характер, может в действительности выражать нечто качественное — иметь принципиальное значение. В оставшейся части своего доклада я подвергну рассмотрению более отдаленное следствие из этого понятия, следствие, которое делает еще более явным один из качественных аспектов этого понятия.

Совершенно очевидно, что в природе существует связь типа «порочного круга», простейшим выражением которой является тот факт, что очень сложные организмы могут воспроизводить себя.

⁴⁶ На первый взгляд (лат.). (Прим. ред.)

Мы вообще склонны неясно подозревать наличие понятия о «сложности»; это понятие и его предполагаемые свойства никогда не были четко сформулированы. Однако мы всегда склоняемся к допущению, что они проявляются следующим образом. Когда автомат выполняет некоторые операции, следует ожидать, что эти операции будут менее сложными, чем сам автомат. В частности, если автомат способен строить другие автоматы, то должно существовать уменьшение сложности при переходе от автомата-строителя к построенному им автомату. Это означает, что если автомат A может произвести автомат B , то автомат A каким-то образом должен содержать полное описание автомата B . Чтобы описание было эффективным, в A , кроме того, должны иметься различные устройства для наблюдения за тем, чтобы это описание соответствующим образом интерпретировалось, а предусматриваемые им строительные операции выполнялись. В этом смысле кажется как будто естественным ожидать известной тенденции к вырождению, т. е. того, что будет наблюдаться некоторое уменьшение сложности по мере того, как одни автоматы будут производить другие.

Хотя это утверждение кажется в какой-то мере правдоподобным, тем не менее оно находится в явном противоречии с весьма очевидными фактами, наблюдаемыми в природе. Организмы воспроизводят себя, т. е. воспроизводят новые организмы, без уменьшения сложности. Кроме того, встречаются продолжительные периоды эволюции, в течение которых сложность даже возрастает. В этом случае, если рассматривать несколько поколений, организмы происходят от других организмов, обладающих меньшей сложностью.

Таким образом, между правдоподобием наших выводов и очевидностью фактов налицо явное несоответствие, если не хуже. Ввиду этого заслуживает, по-видимому, внимания попытка выяснить, нет ли здесь чего-нибудь такого, что можно было бы сформулировать строго.

Не случайно в изложенных выше рассуждениях я прибегал к расплывчатым и неточным формулировкам. Мне кажется, что иначе нельзя было бы создать яркое впечатление о той ситуации, которая сложилась вокруг рассматриваемого вопроса. Теперь я постараюсь быть более точным.

Теория вычислительных автоматов Тьюринга

Английский логик Тьюринг около 12 лет тому назад рассмотрел следующую проблему.

Тьюринг хотел сформулировать общее определение вычислительного автомата. Формальное определение получилось таким.

Автомат есть «черный ящик», который мы не описываем подробно, но который обладает следующими свойствами. Он имеет конечное число состояний, которые следует *prima facie* характеризовать, только указав их число (скажем, n) и занумеровав их числами $1, 2, \dots, n$. Работа автомата будет существенно охарактеризована, если указать, каким образом можно вызвать изменение его состояния, т. е. как перевести автомат из состояния i в состояние j . Это изменение состояния потребует некоторого взаимодействия с внешним миром, которое будет стандартизовано следующим образом. Поскольку речь идет о машине, весь внешний мир можно представить себе состоящим из длинной бумажной ленты. Пусть эта лента имеет, например, 1 дюйм в ширину и разделена на ячейки (клетки), имеющие 1 дюйм в длину. В каждой ячейке этой ленты мы можем ставить или не ставить какой-нибудь знак, например точку, причем мы предполагаем, что эту точку можно как ставить, так и стирать. Ячейку, отмеченную точкой, мы будем называть «1», ячейку, не отмеченную точкой, будем называть «0». (Мы могли бы отмечать ячейки, используя большее число знаков, но, как показал Тьюринг, это не играет роли, ибо

не приводит к чему-либо существенно более общему.) При описании положения ленты относительно автомата предполагается, что автомат может непосредственно контролировать одну ячейку ленты и что он обладает способностью передвигать ленту вперед и назад, скажем, на одну клетку за один раз. Чтобы пояснить вышеизложенное, допустим, что автомат находится в состоянии i ($= 1, 2, 3, \dots, n$) и что на ленте он видит знак e ($= 0, 1$); потом он переходит в состояние j ($= 1, 2, 3, \dots, n$) передвигает ленту на p ячеек ($p = 0, +1, -1$; $+1$ означает, что автомат передвинул ленту на одну ячейку вперед, -1 — на одну ячейку назад) и вписывает в новую клетку, которая оказывается в поле его зрения, знак f ($= 0, 1$; «вписывание нуля» означает, что автомат стирает точку; «вписывание единицы» означает, что автомат ставит точку). Задав j, p и f как функции от i и e , мы полностью определим действие такого автомата.

Тьюринг тщательно проанализировал, какие математические процессы могут осуществлять автоматы этого типа. В связи с этим он доказал различные теоремы, касающиеся классической «проблемы разрешимости» логики⁴⁷, но я не буду касаться здесь этого вопроса. Он также ввел и проанализировал понятие «универсального автомата». Эта часть его работы имеет непосредственное отношение к нашей теме. Бесконечные последовательности цифр e ($= 0, 1$) являются одним из основных объектов математического исследования. Рассматриваемые как представления чисел в двоичной системе, они в сущности оказываются эквивалентными понятию действительного числа. Поэтому Тьюринг в своих рассуждениях исходил из таких последовательностей.

Тьюринг исследовал вопрос, какие автоматы могли бы построить ту или иную последовательность. Иначе говоря, если задан закон образования такой последовательности, то спрашивается, какой автомат следует применить для образования последовательности согласно этому закону. При этом под процессом «образования» последовательности понимается следующее. Автомат способен «образовать» некоторую последовательность, если возможно разметить определенный конечный участок ленты таким образом, что если ленту ввести в рассматриваемый автомат, последний выпишет эту последовательность на остальной свободной (и бесконечной) части ленты. Разумеется, этот процесс выписывания бесконечной последовательности никогда не закончится. То, что имеется в виду, когда говорят, что автомат способен выписать на ленте данную бесконечную последовательность,— это лишь то, что, выполняя эту задачу, он будет работать неограниченно долго и при условии, что ему предоставят достаточно времени, выпишет на ленте любую требуемую (разумеется, конечную) часть данной (бесконечной) последовательности. Упомянутый выше конечный участок ленты, размечаемый перед введением ленты в автомат, представляет собой «инструкцию» автомату для решения этой задачи.

Основной результат теории Тьюринга

A priori кажется, что создание «универсального автомата» невозможно. Как может существовать автомат, столь же эффективный, как и любой автомат, который только можно себе представить, в том числе, например, автомат, вдвое больший данного по размерам и сложности?

Тем не менее Тьюринг доказал, что такой автомат возможен. Хотя структура универсального автомата очень сложна, принцип, лежащий в его основе, весьма прост.

⁴⁷ *О проблемах разрешимости* см.: П. С. Н о в и к о в, Элементы математической логики, Физматгиз, М., 1959, гл. I, § 4 и гл. III, § 10; Д. Г и л ь б е р т и В. А к к е р м а н, Основы теоретической логики, перев. с нем., ИЛ, М., 1947, гл. 3, § 12 и гл. 4; В. А. У с п е н с к и й, Алгоритм, «Философская энциклопедия», М., Гос. научное изд. «Советская энциклопедия», том I (печ.). (Стр. 94.)

Тьюринг заметил, что совершенно общее описание произвольного автомата может быть дано (в смысле предыдущего определения) с помощью конечного числа слов. Это описание будет содержать некоторые пустые места — пробелы, которые соответствуют упомянутым выше функциям (функциям j , p , f , которые зависят от i , e), определяющим работу данного автомата. Если на пустые места подставлены соответствующие значения, мы имеем дело с конкретным автоматом. Если же пустые места не заполнены, эта схема представляет собой общее определение автомата в самом широком смысле слова. Так вот: можно описать автомат, обладающий способностью интерпретировать такого рода определение, иначе говоря, такой автомат, который, если ввести в него функции, определяющие в указанном выше смысле работу того или иного конкретного автомата, будет работать так же, как работает последний. Способность выполнять эти действия является не более загадочной, чем способность читать словарь и грамматику и следовать их указаниям относительно использования слов и законов их сочетания. Этот автомат, построенный так, что он может читать описания и имитировать описанный объект, и является универсальным автоматом в смысле Тьюринга. Чтобы он мог дублировать любую операцию, которую может выполнять любой другой автомат, достаточно снабдить его описанием этого автомата и, кроме того, инструкциями, необходимыми последнему для выполнения рассматриваемых операций.

Расширение программы на случай автоматов, которые производят автоматы

Для решения вопроса, который я рассматриваю здесь — проблемы «самовоспроизведения» автоматов,— процедура Тьюринга недостаточна лишь в одном отношении. Его автоматы являются чисто вычислительными машинами. Выдаваемая ими продукция — это участки ленты с нанесенными на ней нулями и единицами. Предметом же нашего рассмотрения является автомат, на выходе которого получается другой автомат. В принципе, однако, нетрудно исследовать это более широкое понятие и вывести из него результат, эквивалентный результату Тьюринга.

Основные определения

Как и в предыдущем случае, здесь тоже весьма важно дать строгое определение того, что следует понимать под автоматом в рамках нашего исследования. Прежде всего необходимо составить полный список тех элементарных частей, которые будут использоваться. Этот список должен содержать не только перечисление всех элементарных частей, но и полный набор сведений о том, как работает каждая элементарная часть в отдельности. Относительно легко составить такой список, т. е. написать каталог «машинных деталей», который был бы достаточно обширен для того, чтобы из них можно было строить множество нужных механизмов, и который удовлетворял бы требованиям аксиоматической строгости, необходимой в рассмотрении этого рода. Этот список даже необязательно делать длинным. Его можно, конечно, сделать произвольно длинным или произвольно коротким. Список получится длинным, если в качестве элементарных частей в него будут включены объекты, которые можно получить в виде комбинаций других элементарных частей. Но список можно сделать и коротким — фактически можно устроить даже так, чтобы в нем была только одна-единственная деталь,— если каждую элементарную часть наделить разнообразными свойствами и функциями. Поэтому любое утверждение относительно числа необходимых

элементарных частей представляет собой некоторый разумный компромисс, при котором ни от одной элементарной части не ожидается ничего слишком сложного и ни одна элементарная часть не предполагается выполняющей несколько явно не связанных друг с другом функций. В этом смысле, как можно показать, достаточно около дюжины элементарных частей. После этого проблему самовоспроизведения автоматов можно сформулировать следующим образом. Можно ли из указанных элементов построить такой агрегат, что, если его поместить в резервуар, где в большом количестве «плавают» все эти элементы, он начнет строить другие агрегаты, каждый из которых в конце концов станет новым автоматом, в точности подобным первоначальному? Оказывается, это возможно, и принцип, на котором эта возможность основана, тесно связан с очерченным ранее принципом Тьюринга.

Основная идея доказательства теоремы о самовоспроизведении

Прежде всего, можно дать полное описание того, что, собственно, является автоматом в рассматриваемом здесь смысле. Это описание должно носить общий характер, т. е. в нем опять-таки должны быть пустые места, пробелы. Эти пробелы предназначаются для заполнения функциями, описывающими фактическую структуру того или иного автомата. Как и раньше, различие между описанием, в котором имеются пустые места, и описанием, в котором нет пустых мест (так как пробелы подобающим образом заполнены), представляет собой различие между общим описанием произвольного автомата и описанием некоторого конкретного автомата. В принципе нетрудно описать следующие автоматы.

а) Автомат A , который отличается тем, что если в него ввести описание любого другого автомата в терминах соответствующим образом подобранных функций, он построит этот автомат. В данном случае описание совсем необязательно должно представлять собой ленту с нанесенными на ней пометками (как это было необходимо для машин Тьюринга), потому что, как правило, мы вряд ли выберем ленту в качестве структурного элемента. Однако совсем нетрудно описать такие комбинации структурных элементов, которые будут обладать всеми свойствами ленты, как устройства для кодирования, содержащего ячейки, в которых можно делать пометки. Описание в этом смысле мы будем называть инструкцией и обозначать буквой J .

«Строительство», или «конструирование», одним автоматом другого следует понимать в том же смысле, что и раньше. Предполагается, что строящий автомат помещен в резервуар, в котором в большом числе «плавают» все элементарные компоненты. В этой среде наш автомат и будет строить новые автоматы. Не следует особенно беспокоиться о том, каким образом фиксированный автомат этого вида окажется в состоянии строить другие автоматы, превосходящие его самого по размерам и сложности. Ибо очевидно, что в этом случае большие размеры и большая сложность автомата, который должен быть построен, найдет свое отражение в, быть может, еще большем увеличении размеров инструкции J , вводимой в автомат-строитель. Как указывалось выше, эти инструкции должны представлять собой агрегаты элементарных частей. В этом смысле можно сказать, что некоторая вещь вызывает процесс, объем и сложность которого определяются объемом и сложностью объекта, который должен быть построен в ходе этого процесса.

В дальнейшем все автоматы, для построения которых использовалась способность автомата A строить другие автоматы, будут разделять с ним это его свойство. Все они будут иметь определенное место для инструкции J , т. е. место, в которое может быть введена такого рода инструкция. Совершенно ясно, что при описании такого автомата (например, с помощью соответствующей инструкции) указание места для инструкции J

(понимаемой в указанном выше смысле) составляет некоторую часть всего описания автомата. Поэтому мы можем без каких-либо дальнейших разъяснений говорить о «вводе данной инструкции J в данный автомат».

б) Автомат B , который может копировать любую введенную в него инструкцию J . Инструкция J есть агрегат элементарных частей в смысле, указанном в а), заменяющий бумажную ленту машины Тьюринга. Указанная особенность автомата B будет использоваться в случае, когда J представляет собой описание другого автомата. Иначе говоря, автомат B является не чем иным, как «копировальной машиной», которая может, просматривая введенную в нее перфорированную ленту, производить другую перфорированную ленту, тождественную первой. Заметим, что и этот автомат может производить объекты, превосходящие его по размерам и сложности. Отметим также, что в этом нет ничего удивительного. Поскольку автомат B может только копировать, то чтобы получить на его выходе некоторый объект, нам следует лишь подать на его вход объект в точности таких же размеров и такой же сложности, что и объект, построения которого мы от него ожидаем.

Сделав эти предварительные замечания, мы можем перейти к решающему этапу нашего доказательства.

с) Соединим автоматы A и B друг с другом и с механизмом управления C , который выполняет следующие действия. Пусть в автомат A введена инструкция J (понимаемая опять-таки в смысле а) и б)). После этого механизм управления C прежде всего заставит A построить автомат, который описан этой инструкцией J . Затем C заставит B скопировать указанную выше инструкцию J и введет копию в автомат, только что построенный A . Наконец, C отделит это устройство от системы $A + B + C$ и «даст ему жизнь» уже как самостоятельному объекту.

д) Обозначим весь агрегат $A + B + C$ через D .

е) Для того чтобы агрегат $D = A + B + C$ мог функционировать, его следует снабдить некоторой инструкцией J , как это описано выше. Как мы уже отмечали, эта инструкция должна быть введена в автомат A . образуем теперь инструкцию J_D , которая описывает этот автомат D , и введем J_D в автомат A , составляющий часть всего агрегата D . Назовем получившийся при этом агрегат E .

Очевидно, что E обладает способностью к самовоспроизведению. Заметим, что никакого порочного круга при этом не возникает. Решающий этап работы агрегата E наступает тогда, когда инструкция J_D , описывающая D , построена (скопирована) и присоединяется к D . При этом автомат D уже существует к тому времени, когда возникает необходимость построить (скопировать) J_D , и построение J_D никак не может изменить его. J_D просто добавляется к D , в результате чего образуется новый агрегат E , подобный первоначальному. Таким образом, существует определенный хронологический и логический порядок, в котором должно происходить образование D и J_D , и этот процесс согласуется с правилами логики.

Осмысление полученного результата и его непосредственных обобщений

Описание автомата E обладает и некоторыми другими интересными сторонами, о которых я не буду говорить здесь слишком подробно. Например, совершенно очевидно, что инструкция J_D в основном выполняет функцию генов. Ясно также, что копирующий механизм B выполняет основной акт воспроизведения — дубликацию генетического материала, что, очевидно, является основной операцией в процессе деления живых клеток. Нетрудно также видеть, что произвольные изменения системы E и в особенности инструкции J_D могут породить некоторые типичные черты, проявляющиеся в живых организмах в связи с мутациями, которые хотя и являются летальными, как правило, тем не менее содержат в себе возможность дальнейшего самовоспроизведения организмов,

уже не вполне тождественных первоначальным. Разумеется, ясно и то, в каком пункте эта аналогия нарушается. По-видимому, природный ген не содержит полного описания объектов, создание которых стимулируется его присутствием. Он содержит, вероятно, только общие указания, общие наброски.

В проведенных выше весьма общих рассуждениях мы не стремились учесть это используемое природой упрощение. Тем не менее ясно, что это упрощение, как и другие подобные ему, имеет само по себе огромное качественное значение. Мы были бы весьма далеки от настоящего понимания процессов природы, если бы не пытались постичь такого рода упрощающие принципы.

Небольшое изменение предыдущей схемы позволяет нам также построить автомат, который сможет воспроизводить самого себя и, кроме того, строить другой автомат. (Если говорить более подробно, автомат такого рода выполняет, по-видимому, одну из самых типичных — если не самую типичную — функцию гена, состоящую в самовоспроизведении плюс производстве — или стимуляции производства — некоторых специфических ферментов. Действительно, для этого достаточно заменить инструкцию J_D инструкцией J_{D+F} , которая описывает автомат D плюс некоторый данный автомат F . Обозначим D вместе с инструкцией J_{D+F} (которая помещена в автомат A , являющийся частью автомата D) через E_F . Очевидно, что автомат E_F обладает вышеописанным свойством. Он будет воспроизводить себя и, кроме того, строить F .

Заметим, что «мутация»⁴⁸, происходящая в автомате E_F , не является летальной, если она имеет место в пределах F — части инструкции J_{D+F} . Если в результате такой мутации F перейдет в F' , это приведет к превращению E_F в $E_{F'}$, т. е. «мутант» все еще будет обладать способностью к самовоспроизведению. Разумеется, это типичный нелетальный мутант.

Все сказанное представляет собой только первые скромные шаги в направлении систематической теории автоматов. Кроме того, эти шаги делаются лишь в одном частном направлении, которое, как я уже указывал выше, должно привести к выработке строгого понятия о «сложности». Эти шаги показывают, что «сложность» на своем низшем уровне является, по-видимому, вырождающейся, т. е. что каждый автомат, который может производить другие автоматы, на этом уровне будет производить только менее сложные автоматы. Существует, однако, некоторый минимальный уровень, начиная с которого эта склонность к вырождению перестает быть всеобщей. Преодоление этого уровня делает возможным создание автоматов, которые воспроизводят себя или даже строят еще более сложные вещи. Тот факт, что сложность, точно так же как и структура организмов, ниже некоторого минимального уровня является вырождающейся, а выше этого уровня может стать самоподдерживающейся и даже расти, несомненно сыграет важную роль во всякой будущей теории рассматриваемого нами предмета.

⁴⁸ *Мутации* (от лат. *mutatio* — перемена) — дискретные изменения наследственных свойств живого организма. Особи, получившие измененные наследственные свойства — результат мутаций, происшедших в организмах родителей, — называются *мутантами*. Мутации, ведущие к резкому понижению жизнеспособности организма-мутанта или его смерти на той или иной стадии индивидуального развития, называются *летальными*. (Стр. 100.)