

А. ТЬЮРИНГ

# МОЖЕТ ЛИ МАШИНА МЫСЛИТЬ?

С приложением статьи  
ДЖ. фон НЕЙМАНА  
ОБЩАЯ И ЛОГИЧЕСКАЯ  
ТЕОРИЯ АВТОМАТОВ

Перевод с английского  
*Ю. А. Данилова*  
Редакция и предисловие  
проф. *С. А. Яновской*



ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО  
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ  
МОСКВА 1960

---

*Алан М. Тьюринг*  
МОЖЕТ ЛИ МАШИНА МЫСЛИТЬ?  
Редактор *Б. В. Бирюков*

Техн. редактор *И. Ш. Аксельрод*      Корректор *Т. С. Плетнева*

---

Сдано в набор 3/VIII 1960 г. Подписано к печати 11/X 1960 г. Бумага  $84 \times 108^{1/32}$ .  
Физ. печ. л. 3,5. Условн. печ. л. 5,74. Уч.-изд. л. 6,24. Тираж 50 000 экз.  
Т-12850. Цена книги 1 р. 85 к. С 1/I 1961 г. цена 19 к. Заказ № 821.

---

Государственное издательство физико-математической литературы.  
Москва, В-71, Ленинский проспект, 15

---

Первая Образцовая типография имени А. А. Жданова  
Московского городского совнархоза. Москва, Ж-54, Валуевая, 28.

## Содержание

ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ .....	3
<i>А. М. ТЬЮРИНГ</i> МОЖЕТ ЛИ МАШИНА МЫСЛИТЬ? .....	12
I. ИГРА В ИМИТАЦИЮ .....	12
II. КРИТИКА НОВОЙ ПОСТАНОВКИ ПРОБЛЕМЫ .....	13
III. МАШИНЫ, ПРИВЛЕКАЕМЫЕ К ИГРЕ.....	14
IV. ЦИФРОВЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ .....	15
V. УНИВЕРСАЛЬНОСТЬ ЦИФРОВЫХ ВЫЧИСЛИТЕЛЬНЫХ МАШИН .....	18
VI. ПРОТИВОПОЛОЖНЫЕ ТОЧКИ ЗРЕНИЯ ПО ОСНОВНОМУ ВОПРОСУ .....	20
VII. ОБУЧАЮЩИЕСЯ МАШИНЫ.....	32
<i>ДЖ. фон НЕЙМАН</i> ОБЩАЯ И ЛОГИЧЕСКАЯ ТЕОРИЯ АВТОМАТОВ .....	39
I. ПРЕДВАРИТЕЛЬНЫЕ СООБРАЖЕНИЯ.....	39
II. НЕКОТОРЫЕ ЧЕРТЫ ВЫЧИСЛИТЕЛЬНЫХ МАШИН.....	40
III. СРАВНЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН С ЖИВЫМИ ОРГАНИЗМАМИ	45
IV. БУДУЩАЯ ЛОГИЧЕСКАЯ ТЕОРИЯ АВТОМАТОВ .....	51
V. ПРИНЦИПЫ ЦИФРИЗАЦИИ.....	55
VI. ФОРМАЛЬНЫЕ НЕРВНЫЕ СЕТИ .....	57
VII. ПОНЯТИЕ СЛОЖНОСТИ. САМОВОСПРОИЗВЕДЕНИЕ .....	59
ИМЕННОЙ УКАЗАТЕЛЬ.....	66

## ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ

1. В этой книге помещены переводы двух статей: А. М. Тьюринга и Дж. Неймана, опубликованных первая в 1950, вторая в 1951 году. Несмотря на столь значительный срок, прошедший со времени опубликования этих статей, они не утратили свежести и интереса даже для специалистов в области теории автоматов и кибернетики, знакомых с более поздними работами на темы, затронутые в этих статьях. Но статьи эти рассчитаны не на специалистов. Они написаны популярно, без всякого математического аппарата и доступны широкому кругу читателей.

Имена авторов обеих статей хорошо известны каждому математику: имя Тьюринга, прежде всего, в связи с введенным им в 1937 году понятием «машины Тьюринга», являющимся, по существу, одним из наиболее естественных и удобных уточнений понятий *алгоритма* или *вычислимой функции*, имя Дж. Неймана — как имя одного из ведущих математиков XX века, сыгравшего, в частности, важнейшую роль в создании новой науки — теории автоматов. На русском языке имеются уже две работы Дж. Неймана по теории автоматов: «Вероятностная логика и синтез надежных организмов из ненадежных компонент» (сб. «Автоматы», ИЛ, М., 1956, стр. 68–139) и «Вычислительная машина и мозг» («Кибернетический сборник», вып. 1, ИЛ, М., 1960, стр. 11–60). Настоящая статья Неймана носит, однако, гораздо более общий и популярный характер и может служить хорошим введением в только что названные работы. Она представляет собой доклад, прочитанный 20 сентября 1948 года в Калифорнийском технологическом институте на симпозиуме по «механизмам мозга в поведении», в котором принял участие ряд видных психологов, физиологов, и психиатров (Маккаллок, Лорент де Но, Лэшли, Джефрис, Брозин и др.).

2. Обе статьи посвящены вопросам, самое возникновение которых обусловлено созданием новой техники, построением машин, решающих задачи, которые обычно решаются людьми с помощью рассуждения; машин, могущих — в зависимости от результатов своей деятельности — изменять свою программу; машин, которым можно «поручить» составление новых программ, более того, синтез новых — даже более сложных — машин. Естественно, что создание таких машин требует пересмотра самого понятия «машины», более глубокого анализа процессов, связанных с умственной деятельностью человека, и ответа на вопрос о том, что из этой деятельности и как именно может имитироваться машиной. Моделирование умственной деятельности человека автоматами позволяет, с другой стороны, осветить такие черты в структуре человеческого мозга и нервной системы, к которым раньше вообще не было доступа.

Именно эта, вторая, сторона вопроса, связанная со структурой человеческого мозга и нервных сетей, и интересует прежде всего авторов обеих статей. Вопрос «могут ли машины мыслить?» занимает их отнюдь не с точки зрения построения фантастического общества, в котором машины будут соревноваться с людьми. Он занимает их прежде всего потому, что с помощью правильной его постановки они надеются выяснить качественное различие между физическими и умственными возможностями человека, провести четкое разграничение между ними (см. стр. 20). Не машина, а человек, его мозг и умственные возможности находятся в центре внимания обоих авторов. И мы можем это только приветствовать, поскольку и в самой науке, как и во всех других областях человеческой жизни и деятельности, машина должна служить человеку.

Несколько замечаний, относящихся, главным образом, к статье Тьюринга, нам все же хотелось бы сделать, хотя, быть может, они будут понятны читателю лишь по прочтении текста статей.

3. Прежде всего, нужно иметь в виду, что, говоря о «машине», Тьюринг подразумевает обычно не ту машину, с которой имеют дело техники. «Машина» у Тьюринга — это в большинстве случаев некоторая абстракция, эквивалентная — что, впрочем, было уже упомянуто — понятиям *алгоритма* или *вычислимой функции*. И вопрос о том, может ли машина мыслить, равносильен у него, по существу, вопросу о том, носит ли мышление человека алгоритмический характер, нельзя ли представить его себе в виде некоторого универсального алгоритма, универсальной машины Тьюринга<sup>1</sup>, способной «подражать» любой другой машине Тьюринга. Почему же в таком случае Тьюринг не поставил вопрос именно так? Почему он предпочел задавать его в форме вопроса, который он уточнял с помощью некоторой «игры в имитацию»?

На наш взгляд, связанную с этим ситуацию можно представить себе так. Прежде всего, ясно, что мышление человека происходит, конечно, по определенным законам, которые люди еще со времен Аристотеля и древних Китая и Индии пытались сформулировать в виде правил логики. Естественно спросить себя: нельзя ли с помощью этих правил составить список предписаний, схему алгоритма, или программу, следуя которой «машина» будет имитировать умственную деятельность человека? Но правила логики основаны не только на тех законах, по которым происходит действительное мышление у людей. Даже там, где это возможно (и могло бы облегчить их работу), люди обычно не выводят логические следствия путем, например, приведения конъюнкции посылок к совершенной конъюнктивной нормальной форме или к «силлогистическому многочлену», как это делается в некоторых современных логических исчислениях. В практике своего повседневного мышления они не пользуются и правилами силлогистики Аристотеля, хотя последние были установлены уже более двух тысяч лет тому назад. Они мыслят в действительности не формально, а содержательно. Существование логических алгоритмов (соответствующие этим алгоритмам логические устройства уже сейчас «встраиваются» в «думающие» машины) нельзя поэтому рассматривать непосредственно как свидетельствующее в пользу алгоритмического характера человеческого мышления. На попытке Тьюринга опровергнуть возражение своих противников, опирающихся на неформальный характер мышления человека, мы еще остановимся ниже. Пока заметим только, что вопрос о том, можно ли представить себе мышление в виде некоторого алгоритма, хотя бы универсального, нельзя, таким образом, решить, ссылаясь непосредственно на алгоритмический характер правил логики. По-видимому, именно поэтому Тьюринг и предпочитает ставить его с помощью «игры в имитацию».

4. Цель такой постановки должна была состоять в том, чтобы сделать вопрос «может ли машина мыслить?» осмысленным, уточнить его так, чтобы на него можно было ответить «да» или «нет» и чтобы верность (или неверность) этого ответа можно было проверить. Аналогично тому, как Тьюринг уточнил понятие алгоритма, или вычислимой функции, с помощью своей «машины Тьюринга» так, что появилась возможность доказывать несуществование алгоритма (с полным основанием отвечать «нет» на некоторые вопросы о том, может ли быть найден искомый алгоритм), он попытался уточнить вопрос «может ли машина мыслить?», прибегнув с этой целью к «игре в имитацию», с правил которой и начал поэтому свою статью.

Однако, в то время как уточнение понятия алгоритма с помощью «машины Тьюринга» носило весьма убедительный характер [все известные в математике алгоритмы было нетрудно представить в виде соответствующей «машины Тьюринга»; все другие уточнения понятия алгоритма, также представлявшиеся достаточно хорошо соответствующими интуитивному пониманию алгоритма математиками, всегда

---

<sup>1</sup> Из того, что такая программа может существовать именно для «игры в имитацию» (происходящей, к тому же, в некоторой конкретной обстановке), не следует еще, на, наш взгляд, возможность «запрограммировать» вообще человеческое мышление — пусть даже и ограниченное только его проявлениями в поведении человека.

оказывались эквивалентными уточнению с помощью «машины Тьюринга»], предложенное Тьюрингом уточнение вопроса «может ли машина мыслить?» с помощью «игры в имитацию» отнюдь не представляется столь убедительным. Можно ли заложить в машину *A* (играющую роль мужчины) такую программу, что ее шансы обмануть человека *C* при продолжительности игры в пять минут будут не меньше 30%? [*C* задает вопросы; по ответам *A* и его партнера *B* (женщина *B*, в отличие от *A*, должна помогать *C*) *C* должен угадать, кто из них мужчина и кто женщина]. Если роль *B* также играет машина, а роль *A* играет универсальная машина, способная «подражать» любой машине (при условии, что ей «известна» программа этой машины), то отличить (по ответам) *A* от *B* (как «мужчину» от «женщины») невозможно, т. е. шансы *A* обмануть *C* заведомо не меньше 50%. Но уже в случае, когда программа машины *B* «неизвестна» машине *A*, тем более, следовательно, в случае, когда роль *B* играет человек (сам вопрос о возможности «запрограммировать» мышление которого лишь должен еще быть решен с помощью «игры в имитацию»), машина *A* будет лишена возможности воспользоваться тактикой «подражания», чтобы обмануть *C*. В другом месте (см. стр. 47) Тьюринг особо отмечает то обстоятельство, что если мы имеем дело с машиной, программа которой нам неизвестна, то открыть по «поведению» этой машины ее программу исключительно трудно. На роль отсутствия информации о структуре «машины»-партнера, как особенно затрудняющего работу по исследованию мышления средствами теории автоматов («нервных сетей»), обращает внимание Маккаллок (в дискуссии по публикуемому здесь докладу Дж. Неймана). «Я признаюсь,— говорил Маккаллок,— что нет ничего, в чем бы я завидовал д-ру фон Нейману более чем в том, что машины, с которыми он должен совладать, отличаются тем, что он с самого начала имеет схематический план (чертеж) машины, из которого видно, для чего предназначается машина и как именно она будет делать это. К несчастью для нас, в биологических науках — по крайней мере в психиатрии — мы имеем дело с чуждой или даже враждебной машиной. Мы не знаем точно, что может делать машина, и заведомо не имеем ее плана».

В этой связи не приходится удивляться тому, что и самому Тьюрингу предложенное им уточнение вопроса о том, может ли машина мыслить, не представляется достаточно убедительным и что вместо того, чтобы начать с аргументов, подтверждающих правомерность его предложения решить вопрос с помощью «игры в имитацию», он предпочитает заняться опровержением аргументов своих противников, считающих заранее — до всякого уточнения вопросов о том, что такое машина и может ли машина мыслить,— что машина вообще мыслить не может. Несмотря на остроумие, проявленное им в этом, Тьюринг не скрывает от читателя, что он, по существу, и сам понимает недостаточность своей аргументации. «Читатель, вероятно, уже почувствовал,— пишет он,— что у меня нет особенно убедительных аргументов позитивного характера в пользу своей собственной точки зрения. Если бы у меня были такие аргументы, я не стал бы так мучиться, разбирая ошибки, содержащиеся в мнениях, противоположных моему собственному» (стр. 49). И хотя он заканчивает статью приведением некоторых позитивных доводов (относящихся к «обучающейся» машине, о которой ниже еще будет идти речь), основной вывод, к которому он приходит, состоит все же в том, что вопрос еще не созрел для того, чтобы на него уже можно было ответить. «Единственно убедительное доказательство, которое могло бы подтвердить правильность нашей точки зрения,— пишет он,— ...состоит в том, чтобы подождать до конца нашего столетия и провести описанный эксперимент» (стр. 50–51).

5. Какой же смысл может иметь в таком случае тот вопрос, который ставит Тьюринг в своей статье? Чтобы сформулировать свой ответ, начну издалека — с критики некоторых аргументов Тьюринга, которые он противопоставляет своим противникам.

Основной довод, с помощью которого Тьюринг фактически пытается обосновать тезис об алгоритмическом характере человеческого мышления, содержится в его ответе на

возражение 8) («с точки зрения неформальности поведения человека»). Как и в других случаях, единственной целью Тьюринга здесь как будто является показать, что утверждения его противников лишены на самом деле доказательной силы. В данном случае он пытается это сделать, отыскав логическую ошибку в некотором доказательстве несуществования полного списка фиксированных правил («программы»), определяющих заранее все поведение человека (см. стр. 46–47). С этой целью Тьюринг предлагает прежде всего отделить «правила действия», т. е. такие предписания, как «остановитесь, если увидите красный свет»,— которые не существуют независимо от людей, устанавливающих их, и которые поэтому всегда могут быть изменены, отменены или, наоборот, введены, почему их и нельзя «запрограммировать» полностью заранее,— от «законов поведения», под которыми он понимает «естественные законы, управляющие жизнью человека» и совершенно независимые от его, человеческого, произвола. Такие законы человек может только открывать, а не создавать. Они зафиксированы в природе заранее, а в таком случае, по мнению Тьюринга, утверждение о том, что не существует их полного списка, не звучит уже столь убедительно. Ведь «в отсутствии законов поведения, которые в своей совокупности полностью определяли бы нашу жизнь, нельзя убедиться столь же легко, как в отсутствии законченного списка правил действия» (стр. 47). По мнению Тьюринга, нельзя ссылаться и на трудности, связанные с составлением п о л н о г о списка естественных законов, управляющих жизнью человека, поскольку в действительности с такими же трудностями мы можем встретиться и когда имеем дело с машиной, программа которой нам неизвестна, хотя заведомо существует. И Тьюринг тут же признается в том, что лично он убежден даже «не только в том, что быть управляемым законами поведения — значит быть некоторым родом машины (не обязательно машиной с дискретными состояниями), но что и, наоборот, быть такой машиной означает быть управляемым законами доведения» (там же).

Поскольку смысл фигурирующих в этом утверждении терминов (например, «некоторый род машины») не уточнен, с ним трудно спорить (можно ведь, например, *по определению*, просто отождествить «машину» с «объектом, управляемым естественными законами»). Тем не менее все это рассуждение Тьюринга отнюдь не убедительно. Ибо, прежде всего, нельзя отделить полностью нормативные правила действия от естественных законов, управляющих жизнью человека, в том числе и его способами установления нормативных правил действия. Ведь даже в грамматике — где заведомо есть и те, и другие: а) нормативные соглашения о том, например, как писать: *по-видимому* или *повидимому*, *безыдейный* или *безидейный*, и б) описания структуры языка, сложившегося независимо от правил, устанавливаемых грамматиками,— нормативные соглашения вводятся обычно со ссылками на естественные свойства языка. Не случайно и в логике пользуются успехом так называемые «естественные», или «натуральные», исчисления, формальные правила которых строят, ориентируясь на содержательные приемы обычного человеческого мышления.

Однако, и наоборот, естественные законы, существующие независимо от произвола человека, открываются людьми, в их (человеческой) *практической* деятельности, с помощью методов абстракции, также предполагающих некоторую активность человека, некоторую обработку исследуемого материала; и в свою очередь естественные законы используются людьми как руководства к действию, т. е. как, в некотором смысле, нормативные предписания.

Далее,— Тьюринг совершенно прав, конечно, утверждая, что поведение человека управляется объективными законами, независимыми от его произвола. Однако из этого отнюдь не следует существование *законченного* списка *раз навсегда фиксированных* законов поведения. Тезис о диалектическом характере познания человека, из которого следует, в частности, что во всяком таком (закшченном)списке может найти выражение только некоторый моментальный снимок с. вечно движущегося,

развивающегося человеческого мышления, содержится в следующих замечательных словах В. И. Ленина:

«Мы не можем представить, выразить, смерить, изобразить движения, не прервав непрерывного, не упростив, у грубив, не разделив, не омертвив живого. Изображение движения мыслью есть всегда огрубление, омертвление,— и не только мыслью, но и ощущением, и не только движения, но и *всякого* понятия.

И в этом *суть* диалектики. *Эту-то суть* и выражает формула: единство, тождество противоположностей». «Закон берет спокойное,— и поэтому закон, всякий закон, узок, неполон, приближителен».

Согласно этому тезису — анализ содержания которого выходит далеко за рамки настоящего предисловия,— лишь с помощью инвариантного (спокойного) науке удастся познавать движение, с помощью дискретного познавать непрерывное, с помощью формализации выявлять содержание, с помощью активного вмешательства человека: огрубления, упрощения, абстрагирования, — познавать объективные законы, независимые от человека, познающего их. В частности, из него следует непосредственно, что самое познание есть *процесс*, каждая ступень в котором — в том числе и каждый список естественных законов — носит лишь приближенный, неполный характер. С помощью такого списка и поведение человека можно, конечно, имитировать посредством машины, и возможности расширения такой имитации вряд ли можно заранее ограничить. Однако из этого отнюдь не следует существование идеальной машины, программа которой представляет собой *полный список раз навсегда* фиксированных естественных законов человеческого мышления.

С только что приведенным тезисом диалектического материализма связаны и другие положения, в том числе уже упоминавшийся нами тезис, согласно которому «всякий закон есть руководство к действию». Во всяком законе есть очень существенный для практики нормативный элемент. Сама попытка отделить «естественные законы, управляющие жизнью человека», от «правил действия», которыми он руководствуется,— в известном смысле вполне оправданная, — также основана на некотором упрощении и огрублении и носит поэтому лишь приблизительный, узкий, неполный характер.

Далее, нужно отметить, что отражая спокойное, сохраняющееся в изменении, так сказать, на одной ступени, естественные законы сохраняют также и постоянную возможность изменения этой ступени — их дальнейшего расширения, уточнения и развития. Последнее происходит при этом в соответствии с требованиями новой обстановки, с новыми обстоятельствами места и времени — с еще одним диалектико-материалистическим принципом, иными словами: *принципом конкретности истины*. Лишь соблюдение требований этого принципа (хотя и носящих неформализуемый полностью характер) и позволяет нам, беря спокойное в явлении, омертвляя и огрубляя его, выявлять объективные законы, управляющие явлениями независимо от нашего произвола, и притом выявлять эти законы так, что задача их проверки практикой становится вполне осмысленной: дает однозначный ответ «да» или «нет» на вопрос о верности их (в данных условиях).

В свете этих положений и заслуживает, на наш взгляд, внимания тьюринговская постановка вопроса. Начать с того, что в этом свете представляется отнюдь не случайным, что поиски нужной ему программы для машины А, участвующей в «игре в имитацию», — в предположении существования такой программы — Тьюринг предлагает осуществить в виде *процесса* поисков последовательных приближений к этой программе, аналогичного процессу обучения ребенка. «Пытаясь имитировать ум взрослого человека,— пишет Тьюринг,— мы вынуждены много размышлять о том процессе, в результате которого человеческий интеллект достиг своего нынешнего состояния» (стр. 51). Задача при этом должна решаться с помощью расчленения ее на две части: «на задачу построить «программу-ребенка» и задачу осуществить процесс воспитания. Обе эти части тесно связаны друг с другом. Вряд ли нам удастся получить хорошую «машину-ребенка» с

первой же попытки. Надо провести эксперименты по обучению какой-либо из машин такого рода и выяснить, как она поддается научению. Затем провести тот же эксперимент с другой машиной и установить, какая из двух машин лучше» (стр. 52).

Тьюринг подробно описывает в дальнейшем весь процесс «обучения» машины, особо останавливаясь на вопросе о том, как увязать это изменение программы с программой, понимаемой как полный список раз навсегда фиксированных правил. «Некоторым читателям мысль об обучающейся машине может показаться парадоксальной,— пишет он. — Как могут меняться правила, по которым машина производит операции? Ведь правила должны полностью описывать поведение машины независимо от того, какова была ее предыстория и какие изменения она претерпела. Таким образом, правила должны быть абсолютно инвариантными относительно времени» (стр. 56). Насколько можно его понять, предлагаемый им выход из этого затруднения состоит в том, чтобы отделить, так сказать, *метанправила*, согласно которым происходит процесс «обучения» машины, от правил, изменяющихся в этом процессе, которые носят «преходящий» характер. Такое разделение имеет, конечно, смысл. Особенно в предположении, из которого исходит Тьюринг. Однако и в этом предположении остается неясным, должен ли избранный нами процесс «обучения» машины обязательно вести к цели — к той самой, пока еще неизвестной нам, программе машины *A*, существование которой было предположено. Да, по существу, Тьюринг и не намечает какой-либо раз навсегда фиксированной программы «обучения» машины. В действительности и эта «мета»-программа может у него, конечно, изменяться (например, при замене одной «обучающейся» машины другой — лучшей, о чем шла речь выше). Причем изменение это должно происходить на основании указаний практики — этого материалистического критерия истины, достаточно «неопределенного», как указывает Ленин<sup>2</sup>, для того, чтобы «не позволять знаниям человека превратиться в «абсолют»» (например, в полный список раз навсегда фиксированных естественных законов человеческого мышления), и в то же время настолько определенного, чтобы «вести беспощадную борьбу со всеми разновидностями идеализма и агностицизма».

Тьюринг не агностик и не идеалист. Наоборот, он убежден в познаваемости объективных законов, управляющих поведением человека. И хотя его постановка вопроса имеет смысл лишь в случае, если вопрос допускает *положительный* ответ (из постановки задачи Тьюрингом отнюдь не ясно, как мог бы быть получен *отрицательный* ответ на вопрос «может ли машина мыслить?»), хотя в его аргументации есть, с нашей точки зрения, заведомо слабые места, значение поставленных им вопросов нельзя недооценивать. Действительно, представим себе, что в результате процесса «обучения» машины (последовательного уточнения ее программы) нам действительно удастся построить машину, которая будет «играть в имитацию» не хуже человека *A*. Разве это не будет означать, что нам удалось (на этом пути) постигнуть некоторые естественные законы, управляющие мышлением человека? И вряд ли кто-нибудь будет отрицать, что такой результат следует расценивать как подлинное достижение науки (хотя и отнюдь не означающее, с нашей точки зрения, будто нам действительно удалось составить полный список естественных законов человеческого мышления). Вообще, область естественных законов, управляющих мышлением человека, пока еще столь мало исследована, что всякая попытка — пусть даже и слишком упростив вопрос — поставить его так, чтобы можно было на практике получить положительный или отрицательный ответ на него, приобретает интерес. И может быть, в этой связи имеет смысл привести еще одни слова Маккаллока, сказанные им по поводу того же доклада Дж. Неймана:

«С моей точки зрения, то, в чем мы нуждаемся здесь прежде и больше всего, это даже не корректная теория, но просто некоторая теория, начав с которой мы могли бы рассчитывать задать вопрос, на который мы получим ответ, хотя бы и означающий

---

<sup>2</sup> См. «Материализм и эмпириокритицизм», гл. 2, § 6.



только, что наше понятие было полностью ошибочным. В большинстве случаев нам обычно даже не удается задать вопрос в такой форме, чтобы он мог получить ответ».

6. Как уже было отмечено, статья Неймана представляет собой общее методологическое введение в развитую им позднее теорию автоматов. Если Тьюринга интересует, по существу, только вопрос о том, как сделать осмысленной самую постановку задачи о моделировании машиной естественного мышления человека, то Неймана занимает прежде всего сравнение вычислительных машин с живыми организмами и выяснение различий между ними, особенно различий, связанных с несравненно большей надежностью последних. Ведь если мы вскроем сущность этих различий, то, с одной стороны, поймем лучше структуру живого организма, а с другой — будем, например, знать, чем нужно руководствоваться в наших попытках построения более надежных автоматов. И Нейман находит два обстоятельства, объясняющие «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» (стр. 78).

Первое из них состоит в том, что «применяемые нами материалы уступают по качеству материалам, которыми пользуется природа», что и «не позволяет нам достигнуть той высокой сложности организации — при малых размерах,— какая достигнута естественными организмами» (там же). Второе относится уже к области логики.

Однако и с первым связаны некоторые вопросы не только естественнонаучного характера (в том числе и биологические, конечно), но и относящиеся к общелогическим проблемам. Действительно, что следует понимать, например, под высокой «сложностью» (чего-либо), на которую, как отмечает Дж. Нейман (см. стр. 92), нам фактически приходится ссылаться во всякой попытке продвинуться вперед в теории автоматов? Ведь, в частности, и из приведенной только что цитаты со стр. 78 ясно, что «понятие сложности, несмотря на его *prima facie* [на первый взгляд] количественный характер, может в действительности выражать нечто качественное — иметь принципиальное значение» (стр. 92). С другой стороны, хотя все мы интуитивно понимаем, что такое «сложность» (чего-либо), но «это понятие и его предполагаемые свойства никогда не были четко сформулированы» (там же). И Нейман пытается выяснить некоторые стороны понятия сложности, обращаясь к объяснению того, каким образом автомат — относительно которого следует ожидать, что когда он выполняет какие-нибудь операции, эти операции должны быть менее сложными, чем сам автомат,— может тем не менее строить другие автоматы той же или даже большей сложности, чем он сам. Решение этого вопроса — связанное и с попыткой объяснения аналогичных явлений эволюции, происходящей в живых организмах,— Нейман получает путем включения в логику категории *времени* (см. стр. 100). В сущности, здесь мы имеем дело с тем самым явлением, которое состоит в разрешении антиномий типа «порочного круга» посредством иерархии типов или какого-нибудь другого аналога изменения во времени. Действительно, если, предположив, что некоторое суждение  $A$  истинно, мы приходим к заключению, что  $A$  ложно, предположив же, что  $A$  ложно, приходим к заключению, что  $A$  истинно, то перед нами противоречие (антиномия): как  $A$ , так и не- $A$  оба должны быть верны одновременно. Но если высказывание  $A$  содержит параметр  $t$  (время), от значений которого зависит истинность или ложность высказывания  $A$ , то из того, что истинность  $A$  в момент  $t_1$  влечет ложность  $A$  в момент  $t_2$ , ложность же  $A$  в момент  $t_2$  влечет истинность  $A$  в момент  $t_3$ , никакого противоречия не получается. На общем вопросе о диалектическом характере самого изменения во времени — вообще всякого движения — мы здесь останавливаться не будем. Его связь с противоречиями типа «порочного круга», встречающимися, например, в теории множеств, заслуживает — в свете приведенных выше положений диалектического материализма — особого рассмотрения. Заметим только, что создание

теории автоматов не случайно должно было привести Дж. Неймана к включению *времени* в логику, моделируемую автоматами. «Имеется одно важное различие между обычной логикой и представляющими ее автоматами,— пишет Дж. Нейман в другом месте.— В логике время нигде не появляется, тогда как в каждой электрической схеме или нервной системе имеется некоторое запаздывание между сигналом входа и ответом на выходе. Работа такой реальной системы всегда связана с определенной последовательностью во времени. Это никак не является недостатком. Это предотвращает, например, появление более или менее явных порочных кругов различного вида (связанных с «неконструктивностью», «импредицибельностью» и т. п.), которые представляют собой основной класс опасностей в современных логических системах».

В настоящей статье Дж. Неймана вопросам логики в связи с теорией автоматов посвящен специальный раздел («IV. Будущая логическая теория автоматов»). Здесь автор обращает особое внимание на то, что абстракция потенциальной осуществимости<sup>3</sup>, в оправданности которой — в отличие от абстракции актуальной бесконечности — математики и логики, по существу, никогда не сомневались, оказывается в применении к теории автоматов уже неправомерной. [Заметим, что абстракция потенциальной осуществимости играет важнейшую роль в теории алгоритмов, в том числе в определении алгоритма с помощью «машины Тьюринга», которая именно поэтому является не машиной, а лишь некоторым абстрактным понятием.] «Во всей современной логике единственно важным,— пишет Нейман, — 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» (стр. 80).

О значении, которое может иметь в теории автоматов отказ от абстракции потенциальной осуществимости, можно судить уже по второму из тех двух обстоятельств, которые были упомянуты выше в связи с различиями в надежности между живым организмом и автоматом. Если пользоваться абстракцией потенциальной осуществимости, то совершенно безразлично, как представлять натуральное число: по методу простого счета (в виде последовательности палочек, например), с помощью цифр в некоторой позиционной системе или даже в записи, аналогичной  $67^{(257729)}$  [пример Бернайса].

Вряд ли многим из математиков придет в голову сомневаться в том, что от любой из этих записей можно перейти к любой другой, хотя бы для представления этого как возможного потребовалось даже отвлечься от ограниченности размеров такой планеты, как Земля. Между тем в теории автоматов различие в способах представления числа может оказаться, как отмечает Нейман, весьма существенным. Автомат представляет число цифровым (позиционным) способом, и хотя это гораздо более экономично в обозначениях, чем метод простого счета, на котором основана работа нерва, проводящего последовательно один импульс за другим, но зато «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

---

<sup>3</sup> Абстракция потенциальной осуществимости состоит, по А. А. Маркову (см. А. А. М а р к о в , Теория алгоритмов, Тр. матем. ин-та им. В. А. Стеклова, т. XLII, М.—Л., 1954, стр. 15), «в отвлечении от реальных границ наших конструктивных возможностей, обусловленных ограниченностью нашей жизни в пространстве и во времени». В применении к натуральным числам эта абстракция позволяет нам рассуждать о сколь угодно больших натуральных числах как об осуществимых.

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» (стр. 86–87).

С точки зрения логики интересно также, что отказ от абстракции потенциальной осуществимости Дж. Нейман соединяет с требованием сделать логический аппарат гораздо менее комбинаторным (комбинаторный аппарат имеет дело с жесткими понятиями типа «да или нет», «все или ничего») и гораздо более близким к математическому анализу, имеющему дело с непрерывностью и широко пользующемуся абстракцией актуальной бесконечности. В этом свете, вероятно, стоит отметить попытку А. С. Есенина-Вольпина обосновать правомерность абстракции актуальной бесконечности с помощью отказа от абстракции потенциальной осуществимости.

Статья Дж. Неймана, таким образом, представляет интерес не только в специальной области теории автоматов: она имеет прямое отношение и к вопросам логики и оснований математики. И хотя в этой области все еще много спорных вопросов — основная борьба здесь по-прежнему идет вокруг вопросов теории познания, между материалистическим и идеалистическим подходом к вопросу о соотношении мышления и бытия,— но обе публикуемые нами статьи — и Тьюринга, и Неймана — являются еще одним подтверждением того, насколько прав был Ленин, утверждая, что подлинный путь развития науки, при всех зигзагах и колебаниях, характерных для буржуазных ученых, всегда есть путь диалектического материализма.

7. Перевод статей представил местами довольно значительные затруднения,— прежде всего, поскольку в нашем распоряжении не всегда имеется (непосредственно) достаточная информация о смысле некоторых имен или терминов, значение которых авторы могли предполагать известным в своем кругу. Что такое «Касабьянка» или «двадцать вопросов»? Чем замечателен «мост через реку Форт»? Кто такая «леди Лавлейс»? На все эти и ряд других, аналогичных, вопросов, знакомство с которыми для читателя существенно для понимания авторского текста, ответ дается в примечаниях, принадлежащих издательскому редактору перевода Б. В. Бирюкову, которому пришлось проделать значительную работу, чтобы собрать нужные сведения. Б. В. Бирюкову же принадлежит и ряд примечаний методологического характера, а также пояснения к некоторым научным терминам или местам текста, для понимания которых может потребоваться и более подробный логический анализ, чем приводимый автором (см., например, примечание 12 на стр. 105–106 об анализе ошибки, связанной с *распределенностью терминов* категорического силлогизма).

С. Яновская

# А. М. ТЬЮРИНГ МОЖЕТ ЛИ МАШИНА МЫСЛИТЬ?<sup>4</sup>

## І. ИГРА В ИМИТАЦИЮ

Я собираюсь рассмотреть вопрос «Могут ли машины мыслить?». Но для этого нужно сначала определить смысл терминов «машина» и «мыслить». Можно было бы построить эти определения так, чтобы они по возможности лучше отражали обычное употребление этих слов, но такой подход таит в себе некоторую опасность. Дело в том, что, если мы будем выяснять значение слов «машина» и «мыслить», исследуя, как эти слова употребляются обычно, нам трудно будет избежать того вывода, что значение этих слов и ответ на вопрос «могут ли машины мыслить?» следует искать путем статистического обследования наподобие анкетного опроса, проводимого институтом Гэллага<sup>5</sup>. Однако это нелепо. Вместо того, чтобы пытаться дать такого рода определения, я заменю наш вопрос другим, который тесно с ним связан и выражается словами с относительно четким смыслом.

Эта новая форма проблемы может быть описана с помощью игры, которую мы назовем «игрой в имитацию». В этой игре участвуют три человека: мужчина (*A*), женщина (*B*) и кто-нибудь задающий вопросы (*C*), которым может быть лицо любого пола. Задающий вопросы отделен от двух других участников игры стенами комнаты, в которой он находится. Цель игры для задающего вопросы состоит в том, чтобы определить, кто из двух других участников игры является мужчиной (*A*), а кто — женщиной (*B*). Он знает их под обозначениями *X* и *Y* и в конце игры говорит либо: «*X* есть *A*, и *Y* есть *B*», либо: «*X* есть *B*, и *Y* есть *A*». Ему разрешается задавать вопросы такого, например, рода:

*C*: « Попрошу *X* сообщить мне длину его (или ее) волос».

Допустим теперь, что в действительности *X* есть *A*. В таком случае *A* и должен давать ответ. Для *A* цель игры состоит в том, чтобы побудить *C* прийти к неверному заключению. Поэтому его ответ может быть, например, таким;

«Мои волосы коротко острижены, а самые длинные пряди имеют около девяти дюймов в длину».

Чтобы задающий вопросы не мог определить по голосу, кто из двух других участников игры мужчина, а кто — женщина, ответы на вопросы следовало бы давать в письменном виде или, еще лучше, печатать на машинке. Идеальным случаем было бы телеграфное сообщение между комнатами, где находятся участники игры. Если же этого сделать нельзя, то ответы и вопросы может передавать какой-нибудь посредник. Цель игры для третьего игрока — женщины (*B*) — состоит в том, чтобы помочь задающему вопросы. Для нее, вероятно, лучшая стратегия — давать правдивые ответы. Она также может делать такие замечания, как: «Женщина — я, не слушайте его!», но этим она ничего не достигнет, так как мужчина тоже может делать подобные замечания.

Поставим теперь вопрос: «Что произойдет, если в этой игре вместо *A* будет участвовать машина?» Будет ли в этом случае задающий вопросы ошибаться столь же часто, как и в игре, где участниками являются только люди? Эти вопросы и заменят наш первоначальный вопрос «могут ли машины мыслить?».

---

<sup>4</sup> Статья впервые опубликована в английском журнале «Mind» в 1950 г. (т. 59, стр. 433–460) под заголовком «Computing Machinery and Intelligence». Перепечатана в 4-м томе «Мира математики» Ньюмана (The World of Mathematics. A small library of the literature of mathematics from A'h-mosé the Scribe to Albert Einstein, presented with commentaries and notes by James R. Newman, Simon and Schuster, New York, v. 4, 1956, p. 2099–2123). В издании Ньюмана статья озаглавлена «Can the Machine Think?». Перевод сделан по тексту издания Ньюмана

<sup>5</sup> Институт Гэллага — Американский институт общественного мнения <American Institute of Public Opinion>. Основан Гэллагом <George Gallup> в 1935 г. Проводит опросы групп населения по специально разработанной методике с целью определения настроения избирателей перед выборами (президента, депутатов конгресса США и др.) и предсказания результатов выборов, а также по другим вопросам. (Стр. 19)

## II. КРИТИКА НОВОЙ ПОСТАНОВКИ ПРОБЛЕМЫ

Подобно тому как мы задаем вопрос: «В чем состоит ответ на проблему в ее новой форме?», можно спросить: «Заслуживает ли рассмотрения проблема в ее новой постановке?» Этот последний вопрос мы рассмотрим, не откладывая дела в долгий ящик, с тем чтобы в последующем уже не возвращаться к нему.

Новая постановка нашей проблемы имеет то преимущество, что позволяет провести четкое разграничение между физическими и умственными возможностями человека. Ни один инженер или химик не претендует на создание материала, который было бы невозможно отличить от человеческой кожи. Такое изобретение, быть может, когда-нибудь и будет сделано. Но даже допустив возможность создания материала, не отличимого от человеческой кожи, мы все же чувствуем, что вряд ли имеет смысл стараться придать «мыслящей машине» большее сходство с человеком, одевая ее в такую искусственную плоть. Форма, которую мы придали проблеме, отражает это обстоятельство в условии, не позволяющем задающему вопросы соприкасаться с другими участниками игры, видеть их или слышать их голоса. Некоторые другие преимущества введенного критерия можно показать, если привести образчики возможных вопросов и ответов. Например:

С: Напишите, пожалуйста, сонет на тему о мосте через реку Форт<sup>6</sup>.

А: Увольте меня от этого. Мне никогда не приходилось писать стихи.

С: Прибавьте 34 957 к 70 764.

А (молчит около 30 секунд, затем дает ответ): 105 621.

С: Вы играете в шахматы?

А: Да.

С: У меня только король на e8 и других фигур нет. У Вас только король на e6 и ладья на h1. Как Вы сыграете?

А (после 15 секунд молчания): Lh8. Mat.

Нам кажется, что метод вопросов и ответов пригоден для того, чтобы охватить почти любую область человеческой деятельности, какую мы захотим ввести в рассмотрение. Мы не желаем ни ставить в вину машине ее неспособность блистать на конкурсах красоты, ни винить человека в том, что он терпит поражение в состязании с самолетом. Условия нашей игры делают эти недостатки несущественными. Отвечающие, если найдут целесообразным, могут хвастать своим обаянием, силой или храбростью, сколько им вздумается, но задающий вопросы не может требовать практических тому доказательств.

Вероятно, нашу игру можно подвергнуть критике на том основании, что в ней преимущества в значительной степени находятся на стороне машины. Если бы человек попытался притвориться машиной, то, очевидно, вид у него был бы весьма жалкий. Он сразу выдал бы себя медлительностью и неточностью при подсчетах. Кроме того, разве машина не может выполнять нечто такое, что следовало бы характеризовать как мышление, но что было бы весьма отлично от того, что делает человек? Это возражение очень веское. Но в ответ на него мы во всяком случае можем сказать, что если можно все-таки осуществить такую машину, которая будет удовлетворительно играть в имитацию, то относительно него особенно беспокоиться не следует.

Можно было бы заметить, что при «игре в имитацию» не исключена возможность того, что простое подражание поведению человека не окажется для машины наилучшей стратегией. Такой случай возможен, но я не думаю, чтобы он привел нас к чему-нибудь существенно новому. Во всяком случае никто не пытался исследовать теорию нашей игры

---

<sup>6</sup> *Мост через реку Форт* — известный мост консольно-арочного типа, в два пролета перекрывающий реку Форт (Шотландия) при впадении ее в залив Ферг-оф-Форт. Сооружен в 1882–1889 гг. и в течение 28 лет держал мировой рекорд длины пролетов (длина каждого пролета — свыше 518 м, длина моста — около 1626 м). (Стр. 21.)

в этом направлении, и мы будем считать, что наилучшая стратегия для машины состоит в том, чтобы давать ответы, которые в соответствующей обстановке дал бы человек.

### III. МАШИНЫ, ПРИВЛЕКАЕМЫЕ К ИГРЕ

Вопрос, поставленный в разделе I, не станет совершенно точным до тех пор, пока мы не укажем, что именно следует понимать под словом «машина». Разумеется, нам бы хотелось, чтобы в игре можно было применять любой вид инженерной техники. Мы склонны также допустить возможность, что инженер или группа инженеров могут построить машину, которая будет работать, но удовлетворительного описания работы которой они не смогут дать, поскольку метод, которым они пользовались, был в основном экспериментальным. Наконец, мы хотели бы исключить из категории машин людей, рожденных обычным образом. Трудно построить определение так, чтобы оно удовлетворяло этим трем условиям. Можно, например, потребовать, чтобы все конструкторы машины были одного пола; в действительности, однако, этого недостаточно, так как, по-видимому, можно вырастить законченный индивидуум из одной-единственной клетки, взятой (например) из кожи человека. Сделать это было бы подвигом биологической техники, заслуживающим самой высокой похвалы, но мы не склонны рассматривать этот случай как «построение мыслящей машины».

Сказанное наводит нас на мысль отказаться от требования, согласно которому в игре следует допускать любой вид техники. Мы еще больше склоняемся к этой мысли в силу того обстоятельства, что наш интерес к «мыслящим машинам» возник благодаря машине особого рода, обычно называемой «электронной вычислительной машиной» или «цифровой вычислительной машиной». Поэтому мы разрешаем принимать участие в нашей игре только цифровым вычислительным машинам.

На первый взгляд это ограничение кажется весьма сильным. Я постараюсь показать, что в действительности дело обстоит не так. Для этого мне придется дать краткий обзор природы и свойств этих вычислительных машин.

Можно также сказать, что отождествление машин с цифровыми вычислительными машинами — равно как и наш критерий «мышления» — должно быть признано совершенно неудовлетворительным, если (вопреки моему убеждению) окажется, что цифровые вычислительные машины не в состоянии хорошо играть в имитацию.

Целый ряд вычислительных машин уже находится в действии, и естественно возникает вопрос: «А почему бы нам, вместо того чтобы сомневаться в правильности наших рассуждений, не поставить эксперимент? Удовлетворить условиям игры было бы нетрудно. В качестве задающих вопросы можно было бы использовать много различных людей, и полученные статистические данные показали бы, как часто задающим вопросы удавалось прийти к правильному заключению».

Коротко на этот вопрос можно ответить так: нас интересует не то, будут ли все цифровые вычислительные машины хорошо играть в имитацию, и не то, будут ли хорошо играть в эту игру те вычислительные машины, которыми мы располагаем в настоящее время; вопрос заключается в том, существуют ли воображаемые<sup>7</sup> вычислительные

---

<sup>7</sup> Очевидно, в смысле: «существуют ли абстрактные», т. е. идеализированные машины, которые строятся теоретически с использованием каких-нибудь абстракций, например абстракции *потенциальной осуществимости*, согласно которой некоторые операции можно повторять сколь угодно большое число раз, некоторые записи могут быть сколь угодно длинными, емкость «памяти» машины может быть неограниченно большой и т. п.

Фактически поставленный здесь вопрос эквивалентен вопросу о том, существует ли (осуществим ли потенциально) алгоритм, решающий некоторую массовую задачу, формулируемую в терминах игры в имитацию. (Вопросам этого рода посвящена брошюра Б. А. Трахтенброта «Алгоритмы и машинное решение задач», М., Физматгиз, 1957.) (Стр. 23)

машины, которые могли бы играть хорошо. Но это только краткий ответ. Ниже мы рассмотрим этот вопрос в несколько ином свете.

#### IV. ЦИФРОВЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ

То, что мы имеем в виду, говоря о цифровых вычислительных машинах, можно пояснить следующим образом. Предполагается, что эти машины могут выполнять любую операцию, которую мог бы выполнить человек-вычислитель. Мы считаем, что вычислитель придерживается определенных, раз навсегда заданных правил и не имеет права ни в чем отступать от них. Мы можем также считать, что эти правила собраны в книге, которая заменяется другой, когда вычислитель приступает к новой работе. У человека-вычислителя имеется также неограниченный запас бумаги, на которой он производит вычисления. Кроме того, он может выполнять операции сложения и умножения с помощью арифмометра — это несущественно.

Если данное выше пояснение принять за определение, то возникает угроза того, что наше рассуждение окажется движущимся в замкнутом круге. Чтобы избежать этой опасности, мы приведем перечень тех средств, с помощью которых достигается требуемый эффект). Можно считать, что цифровая вычислительная машина состоит из трех частей:

- 1) запоминающего устройства,
- 2) исполнительного устройства,
- 3) контролирующего устройства.

Запоминающее устройство — это склад информации. Оно соответствует бумаге, имеющейся у человека-вычислителя, независимо от того, является ли эта бумага той, на которой производятся выкладки, или той, на которой напечатана книга правил. Поскольку человек-вычислитель некоторые расчеты проводит в уме, часть запоминающего устройства машины будет соответствовать памяти вычислителя.

Исполнительное устройство — это часть машины, выполняющая разнообразные индивидуальные операции, из которых состоит вычисление. Характер этих операций изменяется от машины к машине. Обычно можно проделывать весьма громоздкие операции, например: «Умножить 3 540 675 445 на 7 076 345 687», однако на некоторых машинах можно выполнять только очень простые операции, вроде таких: «написать 0».

Мы уже упоминали, что имеющаяся у вычислителя «книга правил» заменяется в машине некоторой частью запоминающего устройства, которая в этом случае называется «таблицей команд». Обязанность контролирующего устройства — следить за тем, чтобы эти команды выполнялись безошибочно и в правильном порядке. Контролирующее устройство сконструировано так, что это происходит непременно.

Информация, хранящаяся в запоминающем устройстве, разбивается на небольшие части, которые распределяются по ячейкам памяти. Например, для некоторых машин такая ячейка может состоять из десяти десятичных цифр. Тем ячейкам, в которых хранится различная информация, в некотором определенном порядке приписывают номера. Типичная команда может гласить:

«Число, хранящееся в ячейке 6809, прибавить к числу, хранящемуся в ячейке 4302, а результат поместить в ту ячейку, где хранилось последнее из чисел».

Нет необходимости говорить о том, что если все это выразить на русском языке, то машина не выполнит такую команду. Более удобно было бы закодировать эту команду в виде, например, числа 6809430217. Здесь 17 говорит о том, какую из различных операций, которые можно выполнять с помощью данной машины, следует проделать с числами, хранящимися в указанных ячейках. В данном случае имеется в виду описанная выше операция, т. е. операция «число... прибавить к числу...». Следует заметить, что сама

команда занимает 10 цифр и, таким образом, заполняет одну ячейку памяти, что весьма удобно. Обычно контролирующее устройство выбирает необходимые команды в том порядке, в котором они расположены, но иногда могут встречаться и такие команды:

«Теперь выполнить команду, хранящуюся в ячейке 5606, и продолжать оттуда»

или же:

«Если ячейка 4505 содержит 0, выполнить команду, хранящуюся в ячейке 6707, в противном случае продолжать идти по порядку».

Команды этих последних типов очень важны, так как они позволяют повторять снова и снова некоторую последовательность операций до тех пор, пока не будет выполнено определенное условие, причем для повторения данной последовательности операций не приходится прибегать к новым командам. Машина просто выполняет вновь и вновь одни и те же команды. Воспользуемся аналогией из повседневной жизни. Допустим, что мама хочет, чтобы Томми по дороге в школу заходил каждое утро к сапожнику для того чтобы справиться, не готовы ли ее туфли. Она может каждое утро снова и снова просить его об этом. Но она может также раз и навсегда повесить в прихожей записку, которую Томми будет видеть, уходя в школу, и которая будет напоминать ему о том, чтобы он зашел за туфлями. Когда Томми принесет туфли от сапожника, мама должна разорвать записку.

Читатель должен считать твердо установленным, что цифровые вычислительные машины можно строить на основе тех принципов, о которых мы рассказали выше, и что их действительно строят, придерживаясь этих принципов. Ему должно быть ясно, что цифровые вычислительные машины могут в действительности весьма точно подражать действиям человека-вычислителя.

Разумеется, описанная нами книга правил, которой пользуется вычислитель, является всего лишь удобной фикцией. На самом деле настоящие вычислители помнят, что они должны делать. Если мы хотим построить машину, подражающую действиям человека-вычислителя при выполнении некоторой сложной операции, то следует спросить последнего, как он выполняет эту операцию, и ответ представить в виде таблицы команд.

Составление таблицы команд обычно называют «программированием». «Запрограммировать выполнение машиной операции А» — значит ввести в машину подходящую таблицу команд, следуя которым машина может выполнить операцию А.

Интересной разновидностью цифровых вычислительных машин являются «цифровые вычислительные машины, со случайным элементом». Такие машины имеют команды, содержащие бросание игральной кости или какой-нибудь эквивалентный электронный процесс. Одной из таких команд может быть, например, следующая: «бросить кость и полученное при бросании число поместить в ячейку 1000». Иногда говорят, что такие машины обладают свободой воли (хотя лично я не стал бы употреблять такое выражение). Установить наличие «случайного элемента» в машине путем наблюдений за ее действием обычно оказывается невозможным, так как если сделать, например, выбор команды зависимым от последовательности цифр в десятичном разложении числа  $\pi$ , то результат получится совершенно аналогичный.

Все существующие в действительности цифровые вычислительные машины обладают лишь конечной памятью. Однако теоретически нетрудно представить себе машину с неограниченной памятью. Разумеется, в любое данное время возможно использование только конечной части запоминающего устройства. Точно так же, запоминающее устройство, которое можно физически осуществить, всегда имеет конечные размеры, но мы можем представлять дело так, что по мере надобности к нему пристраиваются все новые и новые части. Такие вычислительные машины представляют особый теоретический интерес, и впредь мы будем их называть машинами с бесконечной емкостью памяти.



Сама идея цифровой вычислительной машины отнюдь не является новой. Чарльз Бэббедж<sup>8</sup>, занимавший с 1828 по 1839 г. Люкасовскую кафедру по математике в Кембридже<sup>9</sup>, разработал проект вычислительного устройства, названного им Аналитической машиной; создание ее, однако, так и не удалось завершить. Хотя у Бэббеджа были все основные идеи, существенные для создания такого механизма, его машина не имела перспектив. Скорость вычислений, которую позволила бы достичь машина Бэббеджа, оказалась бы, разумеется, выше скорости, достигаемой человеком, однако-она была бы почти в 100 раз меньше, чем у той вычислительной машины, которая в настоящее время работает в Манчестере<sup>10</sup>) и которая является одной из самых медленных современных машин. Запоминающее устройство в машине Бэббеджа было задумано как чисто механическое, с использованием карт и зубчатых колес.

То, что Аналитическая машина Бэббеджа была задумана как чисто механический аппарат, помогает нам избавиться от одного предрассудка. Часто придают значение тому обстоятельству, что современные цифровые машины являются электрическими устройствами и что нервная система также является таковым. Но поскольку машина Бэббеджа не была электрическим аппаратом и поскольку в известном смысле все цифровые вычислительные машины эквивалентны, становится ясно, что использование электричества в этом случае не может иметь теоретического значения. Естественно, что там, где требуется быстрая передача сигналов, обычно появляется электричество; поэтому неудивительно, что мы встречаем его в обоих указанных случаях. Для нервной системы химические явления играют, по крайней мере, столь же важную роль, что и электрические. В некоторых же вычислительных машинах запоминающее устройство в основном акустическое. Отсюда ясно, что сходство между нервной системой и цифровыми вычислительными машинами, состоящее в том, что в обоих случаях используется электричество, сводится лишь к весьма поверхностной аналогии. Если мы действительно хотим открыть глубокие связи, нам скорее следует искать сходство в математических моделях функционирования нервной системы и цифровых вычислительных машин.

---

<sup>8</sup> Чарльз Бэббедж <Charles Babbage> (1792–1871) — английский ученый, работавший в области математики, вычислительной техники и механики. Выступил инициатором применения механических устройств для вычисления и печатания математических таблиц. В 1812 г. у Бэббеджа возникла идея разностной вычислительной машины (Difference Engine). Строительство этой машины, которая должна была вычислять любую функцию, заданную ее первыми пятью разностями, началось в 1823 г. на средства английского правительства, однако в 1833 г. работа была прекращена главным образом в связи с финансовыми затруднениями. К этому времени у Бэббеджа возник проект другой, более совершенной машины. Эта машина, которую Бэббедж назвал «Аналитической машиной» <Analytical Engine>, должна была проводить вычислительный процесс, заданный любыми математическими формулами. Бэббедж весь отдался конструированию своей новой машины, однако к моменту его смерти она так и не была закончена. Сын Бэббеджа завершил строительство некоторой части машины и провел успешные опыты по применению ее для вычислений некоторого рода. Подробнее о Ч. Бэббедже и его машинах см.: Э. Б у т и К. Б у т , Автоматические цифровые машины, перев. с англ., Физматгиз, М., 1959, стр. 18–21; D. R. H a r t r e e , Calculating Instruments and Machines, Cambridge, 1950, chapter 9: «Charles Babbage and the Analytical Engine»; см. также примечание 10 (Стр. 27.)

<sup>9</sup> Люкасовская кафедра в Тринити-колледже основана в 1663 г. на средства, пожертвованные Генри Люкасом. Первым люкасовским профессором был учитель Ньютона Барроу, вторым — сам Ньютон. Получение этой кафедры, сохранившейся до нашего времени, считалось всегда большой честью. В настоящее время ее занимает Дирак. (Стр. 27.)

<sup>10</sup> Манчестерская машина была построена в Манчестерском университете (Англия) в конце 40-х годов. Конструирование машины происходило под руководством Вильямса <F. C. Williams> и Килберна <T. Kilburn>. В разработке и отладке машины принимал участие Тьюринг, который с этой целью в 1948 г. был приглашен в Манчестерский университет. Тьюринг занимался математическими вопросами, связанными с Манчестерской машиной, и особенно вопросами программирования (см. B i o g r a p h i c a l M e m o i r s o f F e l l o w s o f t h e R o y a l S o c i e t y , v. 1, London, 1955, p. 254–255). Описание Манчестерской машины см. в кн.: F a s t e r t h a n T h o u g h t . A Symposium on Digital Computing Machines. Ed. by B.V. Bowden, London, 1953, chapter I. (Стр. 27.)

## V. УНИВЕРСАЛЬНОСТЬ ЦИФРОВЫХ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

Рассмотренные в предыдущем разделе цифровые вычислительные машины можно отнести к классу «машин с дискретными состояниями». Так называются машины, работа которых складывается из совершающихся последовательно одна за другой резких смен их состояния. Состояния, о которых идет речь, достаточно отличны друг от друга, для того чтобы можно было пренебречь возможностью принять по ошибке одно из них за другое. Строго говоря, таких машин не существует. В действительности всякое движение непрерывно. Однако имеется много видов машин, которые удобно *считать* машинами с дискретными состояниями.

Например, если рассматривать выключатели осветительной сети, то удобно считать, отвлекаясь от действительного положения дела, что каждый выключатель может быть либо включен, либо выключен. То, что выключатель фактически имеет также и промежуточные состояния, несущественно для наших целей, и мы можем об этом забыть. Приведу пример машины с дискретными состояниями. Рассмотрим колесико, способное через каждую секунду совершать скачкообразный поворот (щелчок) на  $120^\circ$ , но которое можно застопоривать с помощью рычажка, управляемого извне. Пусть, кроме того, когда колесико принимает какое-нибудь определенное положение (одно из трех возможных для него), загорается лампочка. В абстрактном виде эта машина выглядит так. Внутреннее состояние машины (которое задается положением колесика) может быть  $q_1$ ,  $q_2$  или  $q_3$ . На вход машины подается либо сигнал  $i_0$ , либо сигнал  $i_1$  (положения рычажка). Внутреннее состояние в любой момент определено предыдущим состоянием и сигналом на входе согласно следующей таблице:

		Предыдущее состояние		
		$q_1$	$q_2$	$q_3$
Вход	$i_0$	$q_2$	$q_3$	$q_1$
	$i_1$	$q_1$	$q_2$	$q_3$

Сигналы на выходе, единственно видимые извне проявления внутреннего состояния (загорание лампочки), задаются таблицей

Состояние	$q_1$	$q_2$	$q_3$
Выход	$o_0$	$o_0$	$o_1$

Этот пример типичен для машин с дискретными состояниями. Такие машины можно описывать с помощью таблиц при условии, что они обладают конечным числом возможных состояний.

Очевидно, что при заданном начальном состоянии машины и заданном сигнале на входе всегда возможно предсказать все будущие состояния. Это напоминает точку зрения Лапласа, утверждавшего, что если известны положения и скорости всех частиц во вселенной в некоторый момент времени, то из такого полного описания ее состояния можно предсказать все ее будущие состояния. Однако то предсказание будущего, о котором у нас идет речь, гораздо ближе к практическому осуществлению, чем то, которое имел в виду Лаплас. Система «вселенной как единого целого» такова, что даже очень небольшие отклонения в начальных состояниях могут иметь решающее значение в

последующем. Смещение одного электрона на одну миллиардную долю сантиметра в некоторый момент времени может явиться причиной того, что через год человек будет убит обвалом в горах. Существенной особенностью тех механических систем, которые мы назвали «машинами с дискретными состояниями», является то, что в них это явление не имеет места. Даже если вместо идеализированных машин взять реальные физические машины, то точное (в разумных пределах) знание о состоянии машины в один момент времени позволяет нам с разумной степенью точности предсказать любое число ее состояний в последующем.

Как мы уже упоминали, цифровые вычислительные машины относятся к классу машин с дискретными состояниями. Но число состояний, в которых может находиться такая машина, обычно велико. Например, число состояний машины, работающей в настоящее время в Манчестере, равно приблизительно  $2^{165\,000}$ , т. е. почти  $10^{50\,000}$ . Сравните эту величину с числом состояний описанного выше «щелкающего» колесика. Нетрудно понять, почему число состояний вычислительной машины оказывается столь огромным. В вычислительной машине имеется запоминающее устройство, соответствующее бумаге, которой пользуется человек-вычислитель. Запоминающее устройство должно быть таково, чтобы в нем можно было записать любую комбинацию символов, которая может быть написана на бумаге. Для простоты допустим, что в качестве символов используются только цифры от 0 до 9. Различия в почерках не принимаются во внимание. Допустим, что человек-вычислитель располагает 100 листами бумаги, разграфленными на 50 строк каждый. Строка может вместить 30 цифр. Число состояний в этом случае равно  $10^{100 \times 50 \times 30}$ , т. е.  $10^{150\,000}$ . Это приблизительно равно числу состояний трех Манчестерских машин, взятых вместе. Логарифм числа состояний по основанию 2 обычно называют «емкостью памяти» машины. Например, Манчестерская машина обладает емкостью памяти около 165 000, а машина с колесиком из нашего примера — около 1,6. Если две машины соединены вместе, то емкость памяти объединенной машины представляет собой сумму емкостей памяти составляющих машин. Это позволяет формулировать такие утверждения, как «Манчестерская машина содержит 64 магнитных трека (направляющих приспособлений), каждый емкостью по 2560, восемь электронно-лучевых трубок емкостью по 1280. Число различных запоминающих устройств доходит до 300, что в целом приводит к емкости памяти в 174 380 единиц».

Если задана таблица, соответствующая некоторой машине с дискретными состояниями, то можно предсказать на будущее, что будет делать эта машина. Нет причин, по которым эти вычисления не могли бы выполняться с помощью цифровой вычислительной машины. Если бы с помощью цифровой вычислительной машины можно было достаточно быстро производить вычисления, то ее можно было бы использовать для имитации поведения любой машины с дискретными состояниями. В «игре в имитацию» тогда могли бы участвовать: машина с дискретными состояниями (которая играла бы за *B*) и имитирующая ее цифровая вычислительная машина (в качестве *A*), и задающий вопросы не смог бы отличить их друг от друга. Разумеется, для этого необходимо, чтобы цифровая вычислительная машина имела надлежащую емкость памяти, а также работала достаточно быстро. Кроме того, ее пришлось бы снабжать новой программой для каждой новой машины, которую она должна была бы имитировать.

Именно это особое свойство цифровых вычислительных машин — то, что они могут имитировать любую машину с дискретными состояниями, и имеют в виду, когда говорят, что цифровые вычислительные машины являются *универсальными* машинами. Из того, что имеются машины, обладающие свойством универсальности, вытекает важное следствие: чтобы выполнять различные вычислительные процедуры, нам вовсе не нужно создавать все новые и новые разнообразные машины (если отвлечься от растущих требований к скорости вычислений). Все вычисления могут быть выполнены с помощью одной-единственной цифровой вычислительной машины, если снабжать ее надлежащей программой для каждого случая. В дальнейшем мы увидим в качестве следствия из этого

результата, что все цифровые вычислительные машины в каком-то смысле эквивалентны друг другу.

Теперь мы можем вернуться к вопросу, поднятому нами в конце раздела III. Там мы высказали предположение, что вопрос «могут ли машины мыслить?» можно заменить вопросом «существуют ли воображаемые цифровые вычислительные машины, которые могли бы хорошо играть в имитацию?». Если угодно, мы можем придать этому вопросу видимость еще большей общности и спросить: «Существуют ли машины с дискретными состояниями, которые могли бы хорошо играть в эту игру?» Но в свете того, что цифровые вычислительные машины универсальны, мы видим, что любой из этих вопросов эквивалентен следующему: «Если взять только одну конкретную цифровую вычислительную машину  $C$ , то, спрашивается: справедливо ли утверждение о том, что, изменяя емкость памяти этой машины, увеличивая скорость ее действия и снабжая ее подходящей программой, можно заставить  $C$  удовлетворительно исполнять роль  $A$  в «игре в имитацию» (причем роль  $B$  будет исполнять человек)?»

## VI. ПРОТИВОПОЛОЖНЫЕ ТОЧКИ ЗРЕНИЯ ПО ОСНОВНОМУ ВОПРОСУ

Теперь мы можем считать, что основные понятия нами выяснены, и перейти к рассмотрению вопроса «могут ли машины мыслить?» и его варианта, изложенного в конце предыдущего раздела. Вместе с тем мы не можем отказаться и от первоначальной формы вопроса, так как по поводу равноценности замены одной формы вопроса другой мнения могут расходиться и в любом случае необходимо выслушать то, что было бы сказано в этой связи.

Читателю будет легче разобраться в этой дискуссии, если я сначала разъясню свои собственные убеждения. Рассмотрим сперва более точную форму вопроса. Я уверен, что лет через пятьдесят станет возможным программировать работу машин с емкостью памяти около  $10^9$  так, чтобы они могли играть в имитацию настолько успешно, что шансы среднего человека установить присутствие машины через пять минут после того, как он начнет задавать вопросы, не поднимались бы выше 70%. Первоначальный вопрос «могут ли машины мыслить?» я считаю слишком неосмысленным, чтобы он заслуживал рассмотрения. Тем не менее я убежден, что к концу нашего века употребление слов и мнения, разделяемые большинством образованных людей, изменятся настолько, что можно будет говорить о мыслящих машинах, не боясь, что тебя поймут неправильно. Более того, я считаю вредным скрывать такие убеждения. Широко распространенное представление о том, что ученые с неуклонной последовательностью переходят от одного вполне установленного факта к другому, не менее хорошо установленному факту, не давая увлечь себя никакому непроверенному предположению, в корне ошибочно. Не будет никакого ущерба от того, что мы ясно осознаем, что является доказанным фактом, а что предположением. Догадки очень важны, ибо они подсказывают направления, полезные для исследований.

Теперь я перехожу к рассмотрению мнений, противоположных моему собственному.

### 1) Теологическое возражение

«Мышление есть свойство бессмертной души человека. Бог дал бессмертную душу каждому мужчине и каждой женщине, но не дал души никакому другому животному или машинам. Следовательно, ни животное, ни машина не могут мыслить».

Я не могу согласиться ни с чем из того, что было только что сказано, и попробую возразить, пользуясь теологическими же терминами. Я счел бы данное возражение более убедительным, если бы животные были отнесены в один класс с людьми, ибо, на мой взгляд, между типичным одушевленным и типичным неодушевленным предметами имеется большее различие, чем между человеком и другими животными. Произвольный характер этой ортодоксальной точки зрения станет еще яснее, если мы рассмотрим, в каком свете она может представиться человеку, исповедующему какую-нибудь другую религию. Как, например, христиане отнесутся к точке зрения мусульман, считающих, что у женщин нет души? Но оставим этот вопрос и обратимся к основному возражению. Мне кажется, что из приведенного выше аргумента со ссылкой на душу у человека следует серьезное ограничение всеисильности всемогущего. Пусть даже существуют определенные вещи, которые бог не может выполнить,— например, сделать так, чтобы единица оказалась равной двум; но кто же из верующих не согласился бы с тем, что бог волен вселить душу в слона, если найдет, что слон этого заслуживает? Мы можем искать выход в предположении, что бог пользуется своей силой лишь в сочетании с мутациями совершенствующимися мозг настолько, что последний оказывается в состоянии удовлетворить требованиям души, которую бог желает вселить в слона. Но точно так же можно рассуждать и в случае машин. Это рассуждение может показаться отличным лишь потому, что в случае машин его труднее «переварить». По сути дела это означает, что мы считаем весьма маловероятным, чтобы бог счел обстоятельства подходящими для того, чтобы дать душу машине, т. е. речь идет в действительности о других аргументах, которые обсуждаются в остальной части статьи. Пытаясь построить мыслящие машины, мы поступаем по отношению к богу не более непочтительно, узурпируя его способность создавать души, чем мы делаем это, производя потомство; в обоих случаях мы являемся лишь орудиями его воли и производим лишь убежища для душ, которые творит опять-таки бог. Все это, однако, пустые рассуждения. В пользу чего бы ни приводили такого рода теологические доводы, они не производят на меня особого впечатления. Однако в старину такие аргументы находили весьма убедительными. Во времена Галилея полагали, что такие церковные тексты, как: «Стояло солнце среди неба и не спешило к западу почти целый день», (Книга Иисуса Навина, глава X, стих XIII) и «Ты поставил землю на твердых основах: не поколеблется она в веки и веки» (псалмы Давида, псалом 103, стих 5) — в достаточной мере опровергали теорию Коперника. В наше время такого рода доказательство представляется беспочвенным. Но, когда современный уровень знаний еще не был достигнут, подобные доводы производили совсем другое впечатление.

## **2) Возражение со «страусовой» точки зрения**

«Последствия машинного мышления были бы слишком ужасны. Будем надеяться и верить, что машины не могут мыслить».

Это возражение редко выражают в столь открытой форме, как это сделано выше. Но оно звучит убедительно для большинства из тех, кому оно вообще приходит в голову. Мы склонны верить, что человек в интеллектуальном отношении стоит выше всей остальной природы. Лучше всего, если бы удалось доказать, что человек *необходимо* является самым совершенным существом, ибо в таком случае он может не бояться потерять свое доминирующее положение. Ясно, что популярность теологического возражения связана именно с этим чувством. Это чувство, вероятно, особенно сильно у людей интеллигентных, так как они ценят силу мышления более высоко, чем остальные люди, и более склонны основывать свою веру в превосходство человека на этой способности.

Я не считаю, что это возражение является достаточно существенным для того, чтобы требовалось какое-либо опровержение. Утешение здесь было бы более подходящим; не предложить ли искать его в учении о переселении душ?

### 3) Математическое возражение

Имеется ряд результатов математической логики, которые можно использовать для того, чтобы показать наличие определенных ограничений возможностей машин с дискретными состояниями. Наиболее известный из этих результатов — теорема Гёделя<sup>11</sup> — показывает, что в любой достаточно мощной логической системе можно сформулировать такие утверждения, которые внутри этой системы нельзя ни доказать, ни опровергнуть, если только сама система непротиворечива. Имеются и другие, в некотором отношении аналогичные, результаты, принадлежащие Черчу, Клини, Россеру и Тьюрингу. Результат последнего особенно удобен для нас, так как относится непосредственно к машинам, в то время как другие результаты можно использовать лишь как сравнительно косвенный аргумент (например, если бы мы стали опираться на теорему Гёделя, нам понадобились бы еще и некоторые средства описания логических систем в терминах машин и машин в терминах логических систем). Результат Тьюринга относится к такой машине, которая, в сущности, является цифровой вычислительной машиной с неограниченной емкостью памяти. Он устанавливает, что существуют определенные вещи, которые эта машина не может выполнить. Если она устроена так, чтобы давать ответы на вопросы, как в «игре в имитацию», то будут вопросы, на которые она или даст неверный ответ, или не сможет дать ответа вообще, сколько бы ни было ей предоставлено для этого времени. Таких вопросов, конечно, может быть много, и на вопросы, на которые нельзя получить ответ от одной машины, можно получить удовлетворительный ответ от другой. Мы здесь, разумеется, предполагаем, что вопросы принадлежат скорее к таким, которые допускают ответ «Да» или «Нет», чем к таким, как «Что вы думаете о Пикассо?». Следующего типа вопросы относятся к числу таких, на которые, как нам известно, машина не может дать ответ: «Рассмотрим машину, характеризующуюся следующим: ... Будет ли эта машина всегда отвечать «Да» на любой вопрос?» Если на место точек подставить описание (в какой-либо стандартной форме, например подобной той, которая была использована нами в разделе V) такой машины, которая находится в некотором сравнительно простом отношении к машине, к которой мы обращаемся с нашим вопросом, то можно показать, что ответ на этот вопрос окажется либо неверным, либо его вовсе не будет. В этом и состоит математический результат; утверждают, будто он доказывает ограниченность возможностей машин, которая не присуща разуму человека. Ответ на это возражение вкратце состоит в следующем. Установлено, что возможности любой конкретной машины ограничены, однако в разбираемом возражении содержится голословное, без какого бы то ни было доказательства, утверждение, что подобные ограничения не применимы к разуму человека. Я не думаю, чтобы можно было так легко игнорировать эту сторону дела. Когда какой-либо из такого рода машин задают соответствующий критический вопрос и она дает определенный ответ, мы заранее знаем, что ответ будет неверным, и это дает нам чувство известного превосходства. Не является ли это чувство иллюзорным? Несомненно, оно бывает довольно искренним, но я не думаю, чтобы ему следовало придавать слишком большое значение. Мы сами слишком часто даем неверные ответы на вопросы, чтобы то чувство удовлетворения, которое возникает у нас при виде погрешимости машин, имело оправдание. Кроме того, чувство превосходства может относиться лишь к той машине, над которой мы одержали свою — в

---

<sup>11</sup> K. G ö d e l , Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, «Monatshefte für Mathematik und Physik», B. 38, 1931, S. 173–198.

сущности весьма скромную — победу. Не может быть и речи об одновременном торжестве над *всеми* машинами. Значит, короче говоря, для любой отдельной машины могут найтись люди, которые умнее ее, однако в этом случае снова могут найтись другие, еще более умные машины, и т. д.

Я думаю, что те, кто разделяет точку зрения, выраженную в математическом возражении, как правило, охотно примут «игру в имитацию» в качестве основы дальнейшего рассмотрения. Те же, кто убежден в справедливости двух предыдущих возражений, будут, вероятно, вообще не заинтересованы ни в каком критерии.

#### 4) Возражение с точки зрения сознания

Это возражение особенно ярко выражено в выступлении профессора *Джефферсона*<sup>12</sup> на Листеровских чтениях за 1949 год<sup>13</sup>, откуда я и привожу цитату. «До тех пор, пока машина не сможет написать сонет или сочинить музыкальное произведение, побуждаемая к тому собственными мыслями и эмоциями, а не за счет случайного совпадения символов, мы не можем согласиться с тем, что она равносильна мозгу, т. е. что она может не только написать эти вещи, но и понять то, что ею написано. Ни один механизм не может чувствовать (а не просто искусственно сигнализировать, для чего требуется достаточно несложное устройство) радость от своих успехов, горе от постигших неудач, удовольствие от лести, огорчение из-за совершенной ошибки, не может быть очарованным противоположным полом, не может сердиться или быть удрученным, если ему не удастся добиться желаемого».

Это рассуждение, по-видимому, означает отрицание нашего критерия. Согласно самой крайней форме этого взгляда, единственный способ, с помощью которого можно удостовериться в том, что машина может мыслить, состоит в том, чтобы *стать* машиной и осознавать процесс собственного мышления. Свои переживания можно было бы потом описать другим, но, конечно, подобное сообщение никого бы не удовлетворило. Точно так же, если следовать этому взгляду, то окажется, что единственный способ убедиться в том, что *данный человек* действительно мыслит, состоит в том, чтобы стать именно этим человеком. Фактически эта точка зрения является солипсистской<sup>14</sup>. Быть может, подобные воззрения весьма логичны, но если исходить из них, то обмен идеями становится весьма затруднительным. Согласно этой точке зрения, **А** обязан думать, что «**А** мыслит, а **В** нет», в то время как **В** убежден в том, что «**В** мыслит, а **А** нет». Вместо того, чтобы постоянно спорить по этому вопросу, обычно принимают вежливое соглашение о том, что мыслят все.

Я уверен, что профессор Джефферсон отнюдь не желает стоять на этой крайней солипсистской точке зрения. Вероятно, он весьма охотно принял бы в качестве критерия «игру в имитацию». Эта игра (если игрок *В* не участвует) нередко применяется на практике под названием *viva voce*<sup>15</sup> для того, чтобы установить, понял ли действительно данный человек некоторую вещь или он заучил нечто «как попугай».

Вот отрывок из такой игры.

<sup>12</sup> G. J e f f e r s o n , The Mind of Mechanical Man. Lister Oration for 1949, «British Medical Journal», v. I, 1949, p. 1105–1121.

<sup>13</sup> *Листеровские чтения*. Джозеф Листер <Lister> (1827–1912) — выдающийся английский хирург. (Стр. 37.)

<sup>14</sup> *Солипсистская точка зрения*. Солипсизм (от латинского *solus* — единственный и *ipse* — сам) — взгляд, согласно которому единственной достоверной реальностью являются внутренние переживания субъекта, его ощущения и мышление. Солипсизм есть крайняя форма философии субъективного идеализма. (Стр. 38.)

<sup>15</sup> *Viva voce* (лат.) — устно. (Прим. ред.)

З а д а ю щ и й в о п р о с ы . Не находите ли Вы, что в первой строке Вашего сонета: «Сравню ль тебя я с летним днем» выражение «с весенним днем» звучало бы лучше?

О т в е ч а ю щ и й . Оно нарушало бы размер стиха.

З а д а ю щ и й в о п р о с ы . А если сказать «с зимним днем»? С размером здесь все обстоит благополучно.

О т в е ч а ю щ и й . Это так, но никто не захочет, чтобы его сравнивали с зимним днем.

З а д а ю щ и й в о п р о с ы . А разве мистер Пиквик не напоминает Вам рождество?

О т в е ч а ю щ и й . Некоторым образом, да.

З а д а ю щ и й в о п р о с ы . Но рождество — зимний день, и я не думаю, чтобы мистер Пиквик имел что-нибудь против этого сравнения.

О т в е ч а ю щ и й . Я не думаю, что Вы говорите все это всерьез. Когда говорят о зимнем дне, имеют в виду обычный зимний день, а не какой-то особенный, вроде рождества.

И так далее. Что бы сказал профессор Джефферсон, если бы машина, пишущая сонеты, могла отвечать примерно так, как это было в приведенном выше отрывке из *viva voce*? Не знаю, стал ли бы он рассматривать ответы машины лишь как «просто искусственную сигнализацию». Если бы ее ответы были столь же связными и удовлетворительными по содержанию, как в приведенном выше отрывке, я не думаю, чтобы профессор Джефферсон охарактеризовал это как дело, выполнить которое может «достаточно несложное устройство». Эту фразу из его выступления следует, по-видимому, относить к таким случаям, когда в машине имеется, скажем, граммофонная пластинка с записью сонета в чьем-либо исполнении, а также механизм, с помощью которого эту запись можно время от времени включать.

Короче говоря, я считаю, что большинство из тех, кто поддерживает возражение с точки зрения сознания *<consciousness>*, скорее откажутся от своих взглядов, чем признают солипсистскую точку зрения. В таком случае они, по-видимому, охотно примут наш критерий.

Мне не хотелось бы создавать впечатление, будто я считаю, что в сознании нет ничего загадочного. Например, неудача наших попыток локализовать сознание похожа на парадокс. Но я вовсе не думаю, что загадки, связанные с сознанием, непременно должны быть разъяснены прежде, чем мы окажемся в состоянии ответить на вопрос, рассматриваемый в настоящей статье.

## **5) Возражения, исходящие из того, что машина не все может выполнить**

Обычно эти возражения выражают в такой форме: «Я согласен с тем, что вы можете заставить машины делать все, о чем Вы упоминали, но Вам никогда не удастся заставить их делать X». При этом перечисляют довольно длинный список значений этого X. Я предлагаю читателю выбрать:

«Быть добрым, находчивым, красивым, дружелюбным (стр. 40), быть инициативным, обладать чувством юмора, отличать правильное от неправильного, совершать ошибки (стр. 41), влюбляться, получать удовольствие от клубники со сливками (стр. 40–41), заставить кого-нибудь полюбить себя, извлекать уроки из своего опыта (стр. 51 и далее), правильно употреблять слова, думать о себе (стр. 42), обладать таким же разнообразием в поведении, каким обладает человек, создавать нечто подлинно новое (стр. 44)». (Некоторые пункты из этого списка ограничений машинных возможностей будут рассмотрены особо на указанных страницах.).



Обычно в подтверждение подобные высказываний не приводят никаких доводов. Я убежден, что эти высказывания основываются главным образом на принципе неполной индукции<sup>16</sup>. Человек в течение своей жизни видел тысячи машин. Из того, что он видел, он делает ряд общих заключений. Машины безобразны, каждая из них создана для того, чтобы выполнять весьма ограниченные задачи, и если необходимо сделать нечто иное, они бесполезны, вариации их поведения крайне незначительны и т. д. и т. п. Естественно, человек делает вывод, что все это является необходимыми особенностями всех машин в целом. Многие из этих ограничений связаны с очень маленькой емкостью памяти большинства машин. (При этом я предполагаю, что понятие емкости памяти машины несколько обобщено таким образом, что охватывает и машины, отличные от машин с дискретными состояниями. Точное определение не играет здесь никакой роли, так как в настоящем рассмотрении мы не претендуем на математическую строгость.) Несколько лет назад, когда очень немногие знали о цифровых вычислительных машинах, часто приходилось встречаться с недоверчивым отношением к тому, что о них рассказывали, если об их замечательных свойствах говорили, не объясняя, как такие машины устроены. Это, вероятно, происходило из-за того, что слушавшие шаблонно применяли принцип неполной индукции. Разумеется, применение этого принципа происходило в основном бессознательно. Если ребенок, обжегшись один раз, боится огня и выражает страх перед огнем тем, что избегает его, то я бы сказал, что он применяет неполную индукцию (само собой разумеется, поведение ребенка можно описать и по-другому). Я не думаю, чтобы трудовая деятельность и обычаи человечества были особенно удачным материалом для применения неполной индукции. Большую часть пространственно-временного континуума *<space-time>* необходимо пылливо исследовать, если мы хотим получить надежные результаты. В противном случае мы можем прийти, скажем, к выводу (к которому приходит большинство английских детей), что все говорят по-английски и что глупо изучать французский язык. Однако относительно многого из того, что было названо в числе вещей, недоступных машине, следует сделать особые оговорки. Неспособность машины получать удовольствие от клубники со сливками может показаться читателю пустяком. Весьма возможно даже, что мы могли бы сделать так, чтобы машина получала удовольствие от этого изысканного блюда, но любая попытка в этом направлении была бы идиотизмом. Эта неспособность машины приобретает значение лишь в сочетании с другими труднодоступными для нее вещами, например в сочетании с трудностью установления между нею и человеком такого же отношения дружелюбия, какое бывает между двумя людьми.

Утверждение «машины не могут совершать ошибок» кажется мне курьезным. Его пытаются парировать: «А разве от этого они хуже?» Отнесемся к этому утверждению не столь враждебно и попытаемся понять, что имеют в виду в действительности. Я думаю, что возражение, содержащееся в утверждении «машины не могут совершать ошибок», можно пояснить с помощью «игры в имитацию». Требуется, чтобы задающий вопросы отличил машину от человека, просто задавая им ряд арифметических задач; машина должна разоблачить себя вследствие своей высокой точности. Ответ на эту аргументацию

---

<sup>16</sup> *Принцип неполной индукции* — принцип логики, согласно которому разрешается делать обобщающее заключение о принадлежности некоторого свойства *a* всем предметам данного класса *A* на основании того, что установлена принадлежность свойства *a* лишь некоторым (не всем) предметам класса *A*, именно тем, которые рассмотрены в ходе индукции. Вывод, основанный на принципе неполной индукции — даже при условии достоверности исходных данных, — не достоверен, а только более или менее вероятен. Выражение «неполная индукция» русского перевода соответствует выражению «scientific induction» (буквально: «научная индукция») английского оригинала. Такой перевод выбран потому, что выражение «научная индукция» употребляется у нас обычно не в том смысле, который имеет в статье Тьюринга выражение «scientific induction» (под «научной индукцией» в нашей литературе обычно понимают сложное рассуждение, основанное на совместном применении неполной индукции и дедукции, которое при определенных условиях — последние, впрочем, не уточняются — может давать достоверное заключение.) (Стр. 40.)

очень прост. Можно сделать так, чтобы машина (запрограммированная для участия в игре) не стремилась давать *правильные* ответы на арифметические задачи. Она может в известной мере специально вводить ошибки в вычисления для того, чтобы сбить с толку задающего вопросы. Что касается ошибок, связанных с механическими неисправностями, то такие ошибки обнаружат себя, по-видимому, тем, что ошибочный результат в этом случае окажется трудно подвести под некоторый общий род типичных арифметических ошибок. Однако даже такая интерпретация данного возражения не является приемлемой. Размеры настоящей статьи не позволяют нам остановиться на этом более подробно. Мне кажется, что это возражение возникает потому, что смешивают ошибки двух видов. Их можно называть «ошибками функционирования» и «ошибками вывода». Ошибки функционирования происходят вследствие некоторых механических или электрических неисправностей, в результате которых машина ведет себя не так, как это было намечено. В философских дискуссиях обычно отвлекаются от возможности ошибок такого рода; поэтому подвергают рассмотрению «абстрактные машины». Эти абстрактные машины — математические фикции, а не реально существующие объекты. По определению, они не могут иметь ошибок функционирования. В этом смысле мы действительно можем сказать, что «машины никогда не могут ошибаться». Ошибки вывода могут возникать лишь тогда, когда сигналу на выходе машины придан определенный смысл. Например, машина может выдавать в печатном виде математические уравнения или какие-нибудь высказывания на русском языке. Если при этом печатается ложное предложение, мы говорим, что машина совершила ошибку вывода. У нас, очевидно, вовсе нет оснований для утверждения, что машина не может совершать ошибок этого рода. Например, она может галка и делать, что печатать « $0 = 1$ ». В качестве более естественного примера рассмотрим машину, располагающую каким-то методом для того, чтобы делать заключения на основе неполной индукции. Мы должны ожидать, что такой метод в отдельных случаях будет давать ошибочные результаты.

На утверждение о том, что машина не может иметь предметом своей мысли самое себя, можно, конечно, дать ответ лишь в том случае, если бы было возможно показать, что машина вообще имеет *какие-либо* мысли, выражающие *какое-либо* предметное содержание. Все же выражение «предметное содержание машинных операций» имеет некоторый смысл, по крайней мере для тех, кто имеет дело с машинными вычислениями. Если, например, машина решает уравнение  $x^2 - 40x - 11 = 0$ , то уравнение можно считать частью предметного содержания операций машины в данный момент. В этом смысле содержанием операций машины, безусловно, может быть она сама. Ее можно использовать при составлении своей собственной программы или для предсказания последствий, вызываемых изменениями в ее устройстве. Наблюдая результаты своего поведения, машина сможет изменять свои собственные программы с тем, чтобы быть более эффективной в достижении некоторой цели. Все это станет возможно в ближайшем будущем; это не утопические мечты.

Возражение, состоящее в том, что машина не отличается разнообразием поведения, является всего лишь способом выражения того обстоятельства, что она не обладает большой емкостью памяти. До самого последнего времени емкость памяти даже в тысячу цифр была очень редкой.

Все возражения, которые мы сейчас разбираем, часто являются просто замаскированной формой возражения с точки зрения сознания. Обычно, если утверждают, что машина *может* выполнить что-нибудь из того, что было перечислено в начале раздела 5), и при этом описывают сущность метода, которым пользуется машина, это не производит большого впечатления. Считают, что, в чем бы ни состоял этот метод, он должен быть весьма элементарным, так как носит механический характер. Сравните сказанное с тем, что говорит Джефферсон (см. приведенный на стр. 37 отрывок из его выступления; слова, взятые в скобки).

## б) Возражение леди Лавлейс

Наиболее подробные сведения, которыми мы располагаем об Аналитической машине Бэббеджа, берутся из воспоминаний леди Лавлейс<sup>17</sup>. В них она высказывает такую мысль: «Аналитическая машина не претендует на то, чтобы создавать что-то действительно новое. Машина может выполнить все то, что мы умеем ей предписать» (курсив леди Лавлейс). Это высказывание цитируется Хартри<sup>18</sup>, который добавляет: «Отсюда не следует, что невозможно сконструировать электронное устройство, которое «мыслит», или в котором, пользуясь биологическими терминами, можно вырабатывать условные рефлексы, на основе которых становится возможным «обучение». Увлекательный и будирующий вопрос, подсказанный некоторыми из последних достижений, состоит в том, осуществимо это принципиально или нет. Однако не видно, чтобы машины, построенные или запроектированные до настоящего времени, обладали этим свойством».

Я полностью согласен с Хартри по этому вопросу. Следует отметить, что он вовсе не утверждает в категорической форме, что машины, о которых идет речь, не обладают этим свойством. Он лишь замечает, что данные, которыми располагала госпожа Лавлейс, не позволяли ей допустить этого. Весьма возможно, что машины, о которых шла речь, в некотором смысле обладали этим свойством. Действительно, пусть некоторая машина с дискретными состояниями обладает рассматриваемым свойством. Аналитическая машина Бэббеджа была универсальной цифровой вычислительной машиной; это значит, что если бы она обладала нужной емкостью памяти и необходимой скоростью работы, то, будь в нее введена соответствующая программа, она могла бы подражать этой машине. По-видимому, этот довод не приходил в голову ни Бэббеджу, ни графине Лавлейс. Во всяком случае от них нельзя требовать, чтобы они исчерпали все, что можно сказать по этому вопросу.

Весь этот вопрос будет рассмотрен еще раз в разделе, посвященном обучающимся машинам.

Один из вариантов аргумента госпожи Лавлейс — это утверждение, гласящее, что машина «никогда не может создать ничего подлинно нового». На секунду возразим поговоркой, что вообще «ничто не ново под луной». Кто может быть уверенным в том, что выполненная им «оригинальная работа» не была ростком из зерна, посеянного образованием, или просто результатом применения хорошо известных общих принципов? Более удачный вариант этого возражения состоит в утверждении, что машина никогда «не может ничем поразить человека». Это утверждение представляет собой прямой вызов, который, однако, мы можем принять, не уклоняясь. Лично меня машины удивляют очень часто. В основном это происходит потому, что я не могу точно рассчитать, чего можно, а

---

<sup>17</sup> Леди Лавлейс, Ада Августа <Ada Augusta, the Countess of Lovelace> принадлежала к тем немногим современникам Бэббеджа, которые вполне оценили значение его идей. Она была дочерью английского поэта Байрона (родилась в 1815 г., умерла в 1852 г.). Лавлейс получила хорошее математическое образование, сначала под руководством своей матери, а потом под руководством проф. Августа де Моргана (Augustus de Morgan), одного из создателей математической логики. С Бэббеджем и его машинами она познакомилась еще в юности. В 1840 г. Бэббедж посетил Турин (Италия) и прочел там серию лекций. Идеи Бэббеджа заинтересовали одного из итальянских офицеров — Менабреа, который опубликовал их изложение в *Bibliothèque Universelle de Geneve* (№ 82, Oktober 1842). Лавлейс перевела на английский язык эту работу и опубликовала ее в *Scientific Memoirs* (ed. by R. Taylor, v. 3, 1842, p. 691–731), присоединив к ней обширные **П р и м е ч а н и я п е р е в о д ч и к а**, более чем в два раза превосходившие по объему текст Менабреа. Эти Примечания относились к принципам работы Аналитической машины и ее применению и были высоко оценены Бэббеджем. См. F a s t e r t h a n T h o u g h t . A Symposium on Digital Computing Machines. Ed. by V. V. Bowden. London, 1953, chapter I. В приложении к книге воспроизведены работа Менабреа в переводе Лавлейс (Sketch on the Analitical Engine invented by Charles Babbage, Esq. by L. F. Menabrea, of Turin, Officer of the Military Engineers) и работа самой Лавлейс (Notes by the Translator). (Стр. 43.)

<sup>18</sup> D. R. H a r t r e e , Calculating Instruments and Machines, New York, 1949

чего нельзя ожидать от них, или (это бывает чаще) потому, что, хотя я и провожу необходимые расчеты, однако делаю это в спешке, неряшливо, рискуя ошибиться. Вот я говорю себе: «По-видимому, электрическое напряжение здесь должно быть таким же, как там: во всяком случае, будем исходить из этого предположения». Само собой разумеется, что в таких случаях я часто ошибаюсь, и получающийся результат оказывается для меня неожиданностью, так как к тому времени, когда эксперимент заканчивается, сделанное допущение уже забыто мною. Эти предположения и натяжки я оставляю открытыми до лекции на тему о моих порочных методах работы. Однако я нисколько не сомневаюсь в том, что действительно испытываю удивление перед машинами.

Я не жду, что этот ответ заставит замолчать моего противника. Вероятно, он скажет, что это удивление происходит вследствие некоторого творческого умственного акта с моей стороны и отражает мое недоверие к машине. Но такая аргументация уводит от вопроса о том, может ли машина чем-либо удивить человека, и возвращает снова к возражению с точки зрения сознания. Этот способ аргументации должен, таким образом, считаться исчерпанным, хотя, быть может, стоит все же отметить то обстоятельство, что если нечто поражает нас своей неожиданностью, то удивление, которое мы испытываем,— независимо от того, что является его источником: человек, книга, машина или еще что-нибудь — требует «творческого умственного акта».

Мнение о том, что машины не могут чем-либо удивить человека, основывается, как я полагаю, на одном заблуждении, которому в особенности подвержены математики и философы. Я имею в виду предположение о том, что коль скоро какой-то факт стал достоянием разума, тотчас же достоянием разума становятся все следствия из этого факта. Во многих случаях это предположение может быть весьма полезно, но слишком часто забывают, что оно ложно. Естественным следствием из него является взгляд, что якобы нет ничего особенного в умении выводить следствия из имеющихся данных, руководствуясь общими принципами.

### **7) Возражение, основанное на непрерывности действия нервной системы**

Нет сомнения в том, что нервная система не является машиной с дискретными состояниями. Небольшая ошибка в информации относительно силы нервного импульса, действующего на нейрон, может привести к значительному изменению импульса на выходе. Исходя из этого, можно было бы как будто предполагать, что нельзя имитировать поведение нервной системы с помощью машины с дискретными состояниями.

То, что машина с дискретными состояниями должна отличаться от машины непрерывного действия, это, конечно, справедливо. Однако если мы будем придерживаться условий «игры в имитацию», то задающий вопросы не сможет использовать это различие. Данную ситуацию можно сделать яснее, рассмотрев другую, более простую, машину непрерывного действия. Для этого особенно хорошо подходит дифференциальный анализатор. (Дифференциальный анализатор — это машина определенного рода, не относящаяся к типу машин с дискретными состояниями, применяемая для вычислений некоторых видов<sup>19</sup>. Некоторые из дифференциальных

---

<sup>19</sup> *Дифференциальный анализатор* — вычислительная машина, разработанная В. Бушем (Vannevar Bush) и его сотрудниками в Массачусеттском технологическом институте в Кембридже (США) в конце 20-х годов и предназначенная для решения широкого класса обыкновенных дифференциальных уравнений. Дифференциальный анализатор — машина непрерывного действия; при решении задач мгновенные значения переменных выражаются положениями вращающихся валов машины (с учетом числа сделанных валом полных оборотов и направления вращения). Первая модель машины была чисто механическим устройством. В дальнейшем дифференциальный анализатор был усовершенствован его автором и превратился в электромеханическую машину. См. В у ш и С. К о л д в е л л , Новый дифференциальный анализатор, Успехи математических наук, т. 1, вып. 5–6 (15–16) (новая серия), М.–Л., 1946, стр. 113–171;

анализаторов выдают ответы в напечатанном виде и поэтому пригодны для игры в имитацию. Цифровая вычислительная машина не может предсказать, какие в точности ответы даст дифференциальный анализатор, решая некоторую задачу, но зато она может сама находить ответы правильного характера на ту же задачу. Например, если требуется найти значение числа  $\pi$  (в действительности приблизительно равное 3,1416), то цифровая вычислительная машина могла бы осуществлять случайный выбор его значения из множества чисел 3,12; 3,13; 3,14; 3,15; 3,16, имеющих соответственно такие (например) вероятности выбора: 0,05; 0,15; 0,55; 0,18; 0,06. При этих условиях задающему вопросы будет очень трудно отличить дифференциальный анализатор от цифровой вычислительной машины.

### 8) Возражение с точки зрения неформальности поведения человека

Невозможно выработать правила, предписывающие, что именно должен делать человек во всех случаях, при всех возможных обстоятельствах. Например, пусть имеется правило, согласно которому человеку следует остановиться, если включен красный свет светофора, и продолжать движение, если свет зеленый; но как быть, если по ошибке оба световых сигнала появятся одновременно? По-видимому, безопаснее всего остановиться. Однако это решение в дальнейшем может быть источником каких-либо новых затруднений. Рассуждая так, мы приходим к заключению, что любая попытка сформулировать правила действия, предусматривающие любой возможный случай, обречена на провал, даже если ограничиться областью транспортной сигнализации. Со всем этим я согласен.

Основываясь на сказанном, доказывают, что мы не можем быть машинами. Я попытаюсь воспроизвести это доказательство, хотя боюсь, что вряд ли сумею сделать это хорошо. Выглядит оно приблизительно так: «Если бы каждый человек обладал определенной совокупностью правил действия, следуя которым он живет, он был бы не чем иным, как машиной. Однако таких правил не существует. Следовательно, человек не может быть машиной». В этом рассуждении бросается в глаза ошибка, связанная с нераспределенностью термина. Я не думаю, чтобы когда-нибудь это возражение излагали именно в такой форме, однако я убежден, что рассуждение этого рода все же находит применение. Однако оно основано на смешении терминов «правила действия» (*rules of conduct*) и «законы поведения» *<laws of behaviour>*, что затемняет вопрос. Под «правилами действия» я понимаю такие предписания, как «остановитесь, если увидите красный свет»; такие предписания могут определять наши действия и осознаваться нами. Под «законами поведения» я понимаю управляющие человеком естественные законы, например, «если человека ущипнуть, он вскрикнет». Если в приведенном выше рассуждении вместо «правил действия, которыми человек руководствуется в своей жизни» подставить «законы поведения, управляющие жизнью человека», то ошибка, связанная с нераспределенностью термина, оказывается вполне устранимой<sup>20</sup>. Ибо мы

---

D. R. H a r t r e e , *Calculating Instruments and Machines*, Cambridge, 1950, chapters 2 and 3; ср. также стр. 65–66 настоящей книги. В настоящее время разнообразные, главным образом электронные, машины непрерывного действия (они называются иногда также *аналоговыми*) получили широкое распространение. (Подробнее о машинах этого класса см. Н. Е. К о б р и н с к и й , *Математические машины непрерывного действия*, Гостехиздат, М., 1954.) (Стр. 45.)

<sup>20</sup> В этом абзаце автор разбирает логическую ошибку в рассуждениях своих оппонентов, привлекая понятие о *распределенности терминов категорического силлогизма*. Категорический силлогизм можно описать как рассуждение, в котором из данного в посылках отношения по объему двух каких-либо терминов (понятий) к третьему следует их отношение друг к другу. Примером силлогизма может быть следующий вывод: (а) все млекопитающие являются позвоночными; (б) все копытные животные являются млекопитающими; (в) значит, все копытные животные являются позвоночными. Здесь из отношения между понятиями *млекопитающие* и *позвоночные* (посылка (а)) и между понятиями

убеждены не только в том, что быть управляемым законами поведения — значит быть некоторым родом машины (не обязательно машиной с дискретными состояниями), но что и, наоборот, быть такой машиной означает быть управляемым законами поведения<sup>21</sup>. Однако в отсутствии законов поведения, которые в своей совокупности полностью определяли бы нашу жизнь, нельзя убедиться столь же легко, как в отсутствии законченного списка правил действия. Единственно известный нам способ отыскания таких законов есть научное наблюдение, и, конечно, мы никогда и ни при каких обстоятельствах не можем сказать: «Мы уже достаточно исследовали. Законов, которые полностью бы определяли нашу жизнь и поведение, не существует».

---

копытные животные и млекопитающие (посылка (б)) выводится отношение между понятиями копытные животные и позвоночные, составляющее содержание заключения (в). Если в этом выводе заменить указанные понятия переменными  $A$ ,  $B$ ,  $C$ , мы выявим форму этого умозаключения: «Все  $A$  суть  $B$ ; все  $C$  суть  $A$ ; следовательно, все  $C$  суть  $B$ », или в виде единого выражения в условной форме: «Если все  $A$  суть  $B$  и все  $C$  суть  $A$ , то все  $C$  суть  $B$ ». Обратная замена переменных любыми понятиями всегда порождает логически правильное рассуждение, т. е. такое, в котором — при условии истинности посылок — всегда получается верное заключение, например: «Все люди смертны, Сократ — человек, следовательно, Сократ смертен» (пример силлогизма, издавна приводящийся в руководствах по формальной логике).

Формы — так называемые модусы — категорического силлогизма могут быть различными. Вышеприведенные силлогизмы являются силлогизмами модуса *Barbara* (латинские названия модусов были придуманы в средние века). Модусы различаются различным расположением терминов и характером входящих в них предложений: в них могут фигурировать не только утвердительные и всеобщие суждения (как в *Barbara*), но и отрицательные и частные суждения («некоторые животные приспособились к холодному климату» и пр.).

Следование заключения из посылок категорического силлогизма происходит по правилам и аксиомам, которые могут быть сформулированы различным образом. Обычно в числе правил фигурируют некоторые, относящиеся к так называемой распределенности терминов. Термин называется распределенным в данном суждении (в которое он входит либо как логическое подлежащее, либо как логическое сказуемое), если суждение служит для выражения информации, относящейся к каждому предмету класса, который имеется в виду в данном термине. Например, в предложении «все млекопитающие позвоночные» термин подлежащего (млекопитающие) распределен (так как суждение выражает информацию о том, что каждое млекопитающее есть позвоночное животное), а термин сказуемого (позвоночные) не распределен (так как в суждении не выражена информация о том, что каждое позвоночное есть млекопитающее). К числу правил, относящихся к распределенности терминов, принадлежит следующее: термин, не распределенный в посылке, не может быть распределен и в заключении. Смысл этого правила состоит в том, что, если в посылке нет информации о каждом члене класса, который имеется в виду в данном термине, она не может появиться и в заключении. Нарушение этого правила будет, конечно, логической ошибкой.

Ошибочное рассуждение, которое рассматривается в тексте, таково: «Если бы все действия человека определялись некоторой совокупностью правил, то он был бы машиной. Но у человека нет такой совокупности правил. Значит, человек не есть машина».

Это рассуждение логически неправильно, так как из отрицания основания условного суждения не следует отрицание его заключения. Представленное в виде категорического силлогизма, оно выглядит так: (1) Все то, чьи действия полностью определены некоторой совокупностью правил, есть машина. (2) Люди не действуют согласно некоторой совокупности правил, полностью определяющей их поведение. (3) Следовательно, люди не машины. Это рассуждение содержит логическую ошибку, так как в посылке (1) термин «машина» не распределен, в то время как в заключении он является распределенным (поскольку в нем выражена информация, относящаяся к каждой машине, именно, что она не есть человек). Если в приведенном выше рассуждении вместо «правил действия» подставить «законы поведения» (в смысле, разъясненном в тексте), то логическая ошибка легко устраняется за счет замены посылки (1) обратным ей суждением: «Все машины отличаются тем, что их поведение полностью определено некоторыми законами» (в истинности которого, говорит Тьюринг, мы убеждены), в котором термин «машины» распределен (так как речь идет о всех машинах). Но тут оказывается, что — в отличие от случая, когда речь шла о «правилах действия», — истинность второй посылки вызывает сомнения; по мнению Тьюринга, мы не имеем возможности убедиться в ее достоверности. (Стр. 47.)

<sup>21</sup> Напомним, что под «машиной» Тьюринг имеет в виду значительно более общее понятие, чем понятие машины, которую можно действительно построить. Критические замечания по поводу излагаемых здесь взглядов Тьюринга см. в Предисловии к русскому переводу. (Стр. 47)

Мы можем с большей убедительностью показать, что любое утверждение такого рода является неоправданным. Действительно, допустим, что мы были бы в состоянии отыскать такие законы (если они существуют). Тогда, если нам будет дана некоторая машина с дискретными состояниями, становится возможным получить посредством наблюдения над ней достаточно данных, чтобы предсказать ее поведение в будущем, причем сделать это можно будет в приемлемый срок, скажем в 1000 лет. Но, по-видимому, дело обстоит не так. Я вводил в Манчестерскую вычислительную машину небольшую программу, занимающую 1000 ячеек памяти, используя которую машина в ответ на введенное в нее 16-значное число выдает в течение двух секунд другое 16-значное число. Попробуйте-ка извлечь из этих ответов такую информацию о программе машины, которая была бы достаточно для предсказания ее ответа на любое, еще не испытанное число. Держу пари, что вам это не удастся.

### 9) Возражение с точки зрения сверхчувственного восприятия

Я предполагаю, что читателю знакомо понятие о сверхчувственном восприятии и его четырех разновидностях, именно: о телепатии, ясновидении, способности к прорицанию и психокинезисе<sup>22</sup>. Эти поразительные явления, по-видимому, опровергают все наши обычные научные представления. Как бы нам хотелось доказать их несостоятельность! К несчастью, статистические данные, по крайней мере в случае телепатии, на их стороне. Очень трудно перестроить наши представления так, чтобы охватить и эти новые факты, ибо тот, кто верит в сверхчувственное восприятие, по-видимому, не так уже далек от веры в чертей и духов. Ведь взгляд, что жизнь и деятельность человека подчиняются только естественным законам *<laws of physics>* — как тем, которые уже известны нам, так и тем, которые еще не открыты, но которые предполагаются в некотором смысле аналогичными уже открытым, — напрашивается прежде всего<sup>23</sup>.

---

<sup>22</sup> *Телепатия, ясновидение, способность к прорицанию и психокинезис.* В словаре английского языка Н.Вебстера смысл этих терминов разъясняется следующим образом. *Т е л е п а т и я* (от греч. *tele* — вдаль, далеко и *pathos* — чувство, страдание) — «связь одного ума с другим, осуществляющаяся без участия органов чувств; передача мыслей на расстоянии» (Webster, New International Dictionary of English Language, Second Edition, Unabridged, 1958, p. 2594). *Я с н о в и д е н и е* (англ. *clair-voyance*) — «явление или способность различения объектов, которые не воспринимаются чувствами, но считаются имеющими объективную реальность» (там же, стр. 494). *П р о р о ч е с т в о* (англ. *precognition*) — «способность предвидеть события; предугадывание будущего» (там же, стр. 1944). *П с и х о к и н е з и с* (от греч. *psyche* — душа и *kinesis* — движение) — «вызывание физического движения с помощью психических средств» (там же, стр. 2001). (Стр. 48.)

<sup>23</sup> Представления о сверхчувственном восприятии издавна служили средством пропаганды суеверий и мистики. При этом для подкрепления этих представлений использовались как вымышленные, мнимые факты (иллюзии, самообман или сознательное введение в заблуждение), так и некоторые явления, имевшие под собой основание, но не получившие научного объяснения на данной ступени развития человеческих знаний. Долгое время к явлениям последнего рода относились, например, факты, связанные с гипнозом или повышенной кожной и мышечно-суставной чувствительностью, свойственной некоторым людям (к чему, по-видимому, сводятся многие «психологические опыты» по «передаче мыслей на расстоянии»). Развитие науки принесло естественнонаучное, материалистическое объяснение этих явлений. Так, например, можно считать установленным, что словесные раздражители — в том числе и внутренняя речь — вызывают изменения в потоке нервных импульсов, идущих к мышцам, что ведет к сокращению отдельных групп мышечных волокон даже в тех случаях, когда соответствующее движение не реализуется. Люди, обладающие повышенной кожной и мышечно-суставной чувствительностью, могут воспринимать даже незначительные мышечные сокращения у других людей и на этом основании получать некоторую информацию о содержании внутренней речи — т. е. о мыслях — другого человека. Что касается тех фактов, на которые и сейчас ссылаются сторонники сверхчувственного восприятия, особенно телепатии, то непроверенность их — на фоне известных случаев внушения и самовнушения и вообще некорректности опытов — в значительной мере признают даже сами сторонники этих явлений.

Возражение с точки зрения сверхчувственного восприятия, по моему мнению, является достаточно серьезным. На него можно было бы возразить, сказав, что многие научные теории, несмотря на весь шум вокруг сверхчувственного восприятия, остаются применимыми на практике, так что в действительности можно прекрасно обойтись и без него, попросту забыв о его существовании. Это, пожалуй, слабое утешение; есть опасение, что мышление относится как раз к одному из тех явлений, к которым сверхчувственное восприятие имеет, быть может, непосредственное отношение.

Не в столь общей форме возражение, основанное на сверхчувственном восприятии, может быть выражено так: «Будем играть в имитацию, используя в качестве отвечающих на вопросы: человека, способного воспринимать телепатические воздействия, и цифровую вычислительную машину. Задающий вопросы может сформулировать, например, такой вопрос: «Какой масти карта в моей правой руке?» Человек с помощью телепатии или ясновидения дает правильные ответы в 130 случаях из 400. Ответы же машины могут только случайно оказаться правильными, и она сможет угадать масть, скажем, лишь в 104 случаях. Это позволит задающему вопросы отличить человека от машины».

Здесь открывается интересная возможность. Допустим, что в нашей цифровой вычислительной машине имеется устройство случайной выдачи чисел. Тогда весьма естественно использовать его для получения ответов на вопросы. Но тогда это устройство случайной выдачи чисел будет находиться под влиянием психокинетических способностей задающего вопросы. Возможно, что психокинезис приведет к тому, что машина будет давать правильные ответы гораздо чаще, чем этого следует ожидать, исходя из вероятностных расчетов, так что задающий вопросы может оказаться не в состоянии сделать правильное заключение относительно того, кто из участников человек, а кто машина. С другой стороны, он может, вообще не задавая никаких вопросов, узнать это с помощью ясновидения: если в дело вмешивается сверхчувственное восприятие, возможно еще и не такое.

Если считать, что телепатия возможна, необходимо ввести ограничения в наш критерий. Можно, например, требовать, чтобы ситуация была аналогична той, которая возникает, когда задающий вопросы обращается к самому себе, а один из участников игры подслушивает его через стенку. Чтобы удовлетворить всем требованиям нашей игры, отвечающих на вопросы следовало бы поместить в комнату, «защищенную от телепатии».

## VII. ОБУЧАЮЩИЕСЯ МАШИНЫ

Читатель, вероятно, уже почувствовал, что у меня нет особенно убедительных аргументов позитивного характера в пользу своей собственной точки зрения. Если бы у меня были такие аргументы, я не стал бы так мучиться, разбирая ошибки, содержащиеся в мнениях, противоположных моему собственному. Сейчас я изложу те доводы, которыми я располагаю.

Вернемся на секунду к возражению графини Лавлейс, согласно которому машина может выполнять лишь то, что мы ей приказываем. Можно сказать, что человек «вставляет» в машину ту или иную идею и машина, прореагировав на нее некоторым

---

Обсуждение вопросов, относящихся к так называемому мысленному внушению (телепатии) — вместе с рассмотрением некоторых научных гипотез, выдвинутых для их объяснения, — читатель может найти в книге: Л. Л. В а с и л ь е в , Таинственные явления человеческой психики, Госполитиздат, М., 1959, а также в брошюре; П. И. Г у л я е в , Мозг и электронная машина, Л., 1960. Нет сомнения, что какие бы факты, относящиеся к способам передачи информации — быть может, связанные с опытами по телепатии, получившими в настоящее время известное распространение, — ни были в дальнейшем установлены наукой, они неизбежно получат естественнонаучное объяснение. Вся история науки свидетельствует о материальности мира и закономерном характере его развития, несовместимом ни с какими формами суеверий и мистики. (Стр. 48.)



образом, возвращается затем к состоянию покоя, подобно фортепианной струне, по которой ударил молоточек. Другое сравнение: атомный реактор, размеры которого не превышают критических. Идея, вводимая человеком в машину, соответствует здесь нейтрону, влетающему в реактор извне. Каждый такой нейтрон вызывает некоторое возмущение, которое в конце концов замирает. Но если величина реактора превосходит критические размеры, то весьма вероятно, что возмущение, вызванное влетевшим нейтроном, будет нарастать и приведет в конце концов к разрушению реактора. Имеют ли место аналогичные явления в случае человеческого разума и существует ли нечто подобное в случае машин? В первом случае, кажется, следует дать утвердительный ответ. Большинство умов, по-видимому, являются «подкритическими», т. е. соответствуют, если пользоваться проведенным выше сравнением, подкритическим размерам атомного реактора. Идея, ставшая достоянием такого ума, в среднем порождает менее одной идеи в ответ. Несравненно меньшую часть умов составляют умы надкритические. Идея, ставшая достоянием такого ума, может породить целую «теорию», состоящую из вторичных, третичных и еще более отдаленных идей. Ум *<mind>* животных, по-видимому, явным образом подкритичен. Развивая нашу аналогию, мы ставим вопрос: «Можно ли сделать машину надкритической?»

Для уяснения поставленного вопроса имеет смысл прибегнуть еще к одной аналогии, именно уподобить человеческий разум луковице. Рассматривая функции ума или мозга, мы обнаруживаем определенные операции, которые возможно полностью объяснить в терминах чисто механического процесса. Можно сказать, что они не соответствуют подлинному разуму: это своего рода «кожица», которую следует удалить для того, чтобы обнаружить настоящий разум. Однако рассматривая оставшуюся часть, мы снова найдем «кожицу», которую следует удалить, и т. д. Возникает вопрос: если мы будем продолжать этот процесс, удастся ли нам прийти когда-нибудь к «настоящему» разуму или же в конце концов мы снимем кожицу, под которой ничего не останется? В последнем случае мы считаем, что разум имеет механический характер. (Правда, он не может быть машиной с дискретными состояниями. Этот вопрос мы уже рассматривали.)

Два последних абзаца вовсе не претендуют на роль убедительных доказательств. Их скорее следовало бы считать аргументами риторического характера.

Единственно убедительное доказательство, которое могло бы подтвердить правильность нашей точки зрения, приведено в начале раздела VI (стр. 32) и состоит в том, чтобы подождать до конца нашего столетия и провести описанный эксперимент. А что же можно сказать в настоящее время? И что можно было бы предпринять уже сейчас, если исходить из предположения, что эксперимент окажется успешным?

Как я уже объяснял, проблема заключается главным образом в программировании. Прогресс в инженерном деле также необходим, однако маловероятно, чтобы затруднение возникло с этой стороны. Оценки емкости памяти человеческого мозга колеблются от  $10^{10}$  до  $10^{15}$  двоичных единиц. Я склоняюсь к нижней границе и убежден, что лишь очень небольшая доля емкости памяти человека используется в высших типах мышления, причем из того, что используется, большая часть служит сохранению зрительных восприятий. Для меня было бы неожиданностью, если бы оказалось, что для игры в имитацию на удовлетворительном уровне требуется емкость памяти, превышающая  $10^9$ , во всяком случае, если бы игра велась против слепого человека. (Заметьте: емкость *Британской энциклопедии*, 11-е изд., составляет  $2 \cdot 10^9$ .) Емкость памяти, равная  $10^7$ , практически представляется вполне осуществимой, даже при современном состоянии техники. Вероятно, нет необходимости вообще далее увеличивать скорость машинных операций<sup>24</sup>. Те части современных машин, которые можно рассматривать как аналоги

---

<sup>24</sup> Говоря о том, что нет необходимости в дальнейшем увеличении скорости машинных операций, автор, по-видимому, имеет в виду главным образом требования, возникающие при игре в имитацию. Во всяком случае, при математических применениях быстродействующих вычислительных машин, обусловленных развитием современной науки и техники, требования к скорости их работы непрерывно возрастают. Если

нервных клеток, работают примерно в тысячу раз быстрее последних. Это создает «запас надежности», могущий компенсировать потери в быстроте, возникающие во многих случаях. Перед нами стоит задача составить машинную программу для игры в имитацию. В настоящее время скорость моей работы как программиста составляет примерно тысячу знаков в день; если исходить из такой скорости программирования, то получится, что шестьдесят работников могли бы полностью закончить работу, о которой идет речь, если бы они работали непрерывно в течение пятидесяти лет, при условии, конечно, *что* ничего не пойдет в корзину для бумаг. Желателен, по-видимому, какой-нибудь более производительный метод<sup>25</sup>.

Пытаясь имитировать ум *<mind>* взрослого человека, мы вынуждены много размышлять о том процессе, в результате которого человеческий интеллект достиг своего нынешнего состояния. Мы, можем выделить три компоненты:

- 1) первоначальное состояние ума, скажем, в момент рождения;
- 2) воспитание, объектом которого он был;
- 3) другого рода опыт, воздействовавший на ум,— опыт, который нельзя назвать воспитанием.

Почему бы нам вместо того, чтобы пытаться создать программу, имитирующую ум взрослого, не попытаться создать программу, которая бы имитировала ум ребенка? Ведь если ум ребенка получает соответствующее воспитание, он становится умом взрослого человека. Как можно предположить, мозг ребенка в некотором отношении подобен блокноту, который мы покупаем в киоске: совсем небольшой механизм и очень много чистой бумаги. Наш расчет состоит в том, что механизм в мозгу ребенка настолько несложен, что устройство, ему подобное, может быть легко спрограммировано. В качестве первого приближения можно предположить, что количество труда, необходимое для

---

лучшие из ныне эксплуатируемых автоматических цифровых вычислительных машин работают со скоростью порядка десятков тысяч операций (сложений и умножений) в секунду, то ряд машин, которые осваиваются в настоящее время, имеют скорость работы уже порядка сотен тысяч операций в секунду. Что касается проектируемых типов машин, то для них выдвигается требование доведения скорости вычислений до миллионов операций в секунду. (Стр. 51.)

<sup>25</sup> Вместе с развитием автоматических быстродействующих цифровых вычислительных машин начали разрабатываться и методы, облегчающие программирование задач для решения на этих машинах. Одним из наиболее распространенных методов, ускоряющих ручное программирование, явился *метод стандартных подпрограмм*. Суть метода состоит в том, что составляется коллекция («библиотека») программ для часто встречающихся участков вычислительных процедур (типовые, или стандартные, подпрограммы); при подготовке задач к решению на быстродействующих цифровых вычислительных машинах типовые подпрограммы в различных сочетаниях включаются в составляемую программу. А. Тьюринг в период своей работы в Манчестерском университете принимал активное участие в разработке метода библиотечных подпрограмм в применении К машине Ф. Вильямса и Т. Килберна (ср. примеч. 6).

Впоследствии наряду с методами, облегчающими ручное программирование, начали развиваться методы *автоматизации* программирования. Автоматизация программирования заключается в передаче некоторых этапов подготовки программ для решения задач на быстродействующих цифровых вычислительных машинах самим этим машинам. Главным в автоматическом, или машинном, программировании является использование так называемой *п р о г р а м м и р у ю щ е й п р о г р а м м ы*. Машина, в которую введена программирующая программа, получив сравнительно небольшую по объему исходную информацию о данной задаче и методе ее решения, автоматически составляет программу решения задачи. Начиная с 1954–1955 гг. в Советском Союзе построен и используется ряд программирующих программ. Автоматизация программирования значительно облегчает труд по подготовке машинного решения задач и сокращает время программирования.

Несомненно, что программирование машин для игры в имитацию — разумеется, при условии, что оно сопровождается необходимыми уточнениями в постановке задачи, — является чрезвычайно сложной проблемой, возможность практического решения которой, по-видимому, существенно зависит от дальнейшего развития автоматического программирования, в области которого в настоящее время ведутся исследования по различным направлениям. Об автоматизации программирования см., например: 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. X. (Стр. 51.)

воспитания такой машины, почти совпадает с тем, которое необходимо для воспитания ребенка.

Таким образом, мы расчленили нашу проблему на две части: на задачу построить «программу-ребенка» и задачу осуществить процесс воспитания. Обе эти части тесно связаны друг с другом. Вряд ли нам удастся получить хорошую «машину-ребенка» с первой же попытки. Надо провести эксперимент по обучению какой-либо из машин такого рода и выяснить, как она поддается научению. Затем провести тот же эксперимент с другой машиной и установить, какая из двух машин лучше. Существует очевидная связь между этим процессом и эволюцией в живой природе, которая обнаруживается, когда мы отождествляем:

структуру «машины-ребенка» — с наследственным материалом, изменения, происходящие в «машине-ребенке», — с мутациями, решение экспериментатора — с естественным отбором.

Тем не менее можно надеяться, что этот процесс будет протекать быстрее, чем эволюция. Выживание наиболее приспособленных является слишком медленным способом оценки преимуществ. Экспериментатор, применяя силу интеллекта *<intelligence>*, может ускорить процесс оценки. В равной степени важно и то, что он не ограничен использованием только случайных мутаций. Если экспериментатор может проследить причину некоторого недостатка, он, вероятно, в состоянии придумать и такого рода мутацию, которая приведет к необходимому улучшению.

Невозможно применять в точности один и тот же процесс обучения как к машине, так и к нормально развитому ребенку. Например, машину нельзя снабдить ногами; поэтому ее нельзя попросить выйти и принести ведро угля. Машина, по-видимому, не будет обладать глазами. И как бы хорошо ни удалось восполнить эти недостатки с помощью различных остроумных приспособлений, такое существо нельзя будет послать в школу без того, чтобы другие дети не потешались над ним. И вот такое существо мы должны чему-то научить. Отметим, что не стоит особенно беспокоиться относительно ног, глаз и т. д. Пример мисс Елены Келлер<sup>26</sup> показывает, что воспитание возможно, если только удастся тем или иным способом установить двустороннюю связь между учителем и учеником.

Обычно процесс обучения в нашем представлении связан с наказаниями и поощрениями. Идея применения какой-либо формы этого принципа обучения может лежать в основе конструирования и программирования некоторых простых «машин-детей». В этом случае машину следует устроить таким образом, чтобы поступление в нее сигнала-«наказания» приводило к резкому уменьшению вероятности повторения тех реакций машины, которые непосредственно предшествовали этому сигналу, в то время как сигнал-«поощрение», наоборот, увеличивал бы вероятность тех реакций, которые ему предшествовали (которые его «вызвали»). Все это не предполагает со стороны машины никаких чувств. Я проделал несколько экспериментов с одной такой «машинкой-ребенком» и достиг кое-какого успеха в обучении ее некоторым вещам, но метод обучения был слишком необычен, чтобы эксперимент можно было считать действительно успешным.

Применение поощрений и наказаний в лучшем случае может быть лишь частью процесса обучения. Проще говоря, если у учителя нет других средств общения со своими учениками<sup>27</sup>, то количество информации, которую может получить ученик, не превышает

---

<sup>26</sup> Елена Келлер (Keller) (род. в 1880 г.) — американская слепоглухонемая, получившая высшее образование. В возрасте полутора лет в результате болезни потеряла зрение и слух и стала немой. Когда девочке было 6 лет, родители пригласили воспитательницу, которая через посредство осязания научила ребенка говорить, а затем читать и писать (по системе для слепых). Е. Келлер прошла школьный курс, изучила языки, окончила университет; она является автором ряда книг.

Случай Е. Келлер — не единственный случай воспитания слепоглухонемых. В Академии педагогических наук РСФСР в качестве научного сотрудника работает О. И. Скороходова, которая с 5 лет потеряла зрение и слух. Она была воспитана в Харьковской клинике для слепоглухонемых детей. Известна книга О. И. Скороходовой «Как я воспринимаю и представляю окружающий мир», М., 1954. (Стр. 53.)

<sup>27</sup> Кроме сигналов поощрения и наказания. (Прим. ред.)

общего числа примененных к нему поощрений и наказаний. Вероятно, к тому времени, когда ребенок выучит наизусть стихотворение «Касабьянка»<sup>28</sup>, он будет до крайности измучен, если процесс обучения будет идти по методу игры в «20 вопросов»<sup>29</sup>, причем каждое «нет» учителя будет принимать для ученика форму подзатыльника. В силу этого необходимо иметь какие-то другие, «неэмоциональные» каналы связи. Если такие каналы имеются, то, применяя поощрения и наказания, машину можно было бы научить выполнять команды, отдаваемые на каком-либо — например символическом — языке. Эти команды следует передавать по «неэмоциональным каналам». Применение такого символического языка значительно снизит число требуемых поощрений и наказаний. О том, какая степень сложности является наиболее пригодной для «машины-ребенка», могут быть различные мнения. Можно стремиться к тому, чтобы «машина-ребенок» была настолько простой, насколько этого возможно добиться, не нарушая общих принципов. Можно идти противоположным путем: «встраивать»<sup>30</sup> сложную систему логического вывода<sup>31</sup>. В последнем случае значительную часть запоминающего устройства заняли бы определения и суждения (*propositions*). Суждения по своему характеру должны быть различного рода, например: утверждения о хорошо известных фактах, предположения, математически доказанные теоремы, высказывания авторитетных лиц, выражения, по своей логической форме являющиеся суждениями, но не претендующие на верность. Некоторые из этих суждений могут быть охарактеризованы как «приказания». Машину следует устроить так, чтобы, как только некоторое приказание будет оценено ею как «вполне достоверное» *<well-established>*, автоматически выполнялась соответствующая операция. Чтобы пояснить это, предположим, что учитель говорит машине: «Теперь выполняй домашнее задание», а машина реагирует на это тем, что ситуация «Учитель говорит машине: «Теперь выполняй домашнее задание»» включается в число вполне достоверных фактов. Другим фактом такого же рода в ней может быть: «Все, что говорит учитель, истинно». Комбинация этих фактов может<sup>32</sup>, в заключение, вести к тому, что приказание: «Теперь выполняй домашнее задание» тоже будет включено в разряд вполне надежных фактов, а это в свою очередь будет значить, в силу устройства нашей машины, что последняя действительно начнет выполнять домашнее задание, — что нам и было нужно. Процесс логического вывода, применяемый машиной, вовсе не обязательно должен быть таков, чтобы он удовлетворял требованиям самых строгих логиков. Например, может отсутствовать иерархия типов<sup>33</sup>. Но это отнюдь не означает, что

---

<sup>28</sup> «Касабьянка» (*Casabianca*) — стихотворение английской поэтессы Хеманс (*Felicia Hemans*, 1793–1835). Повествует о мальчике 10 лет, сыне капитана Касабьянки, который вместе с отцом погиб на горящем военном корабле, отказавшись покинуть судно, взорванное своим командиром Касабьянкой вовремя морского боя. (Стр. 53.)

<sup>29</sup> «Двадцать вопросов» — распространенная в Англии игра в вопросы и ответы. Обычно ведется так. Один из играющих задумывает некоторое понятие. Другой играющий отгадывает задуманное, задавая вопросы, предполагающие ответы (обязательно правдивые) вида «да» или «нет». Количество вопросов, которое имеет право задать отгадчик, не должно превышать некоторого заранее установленного числа. Отгадчик выигрывает, если при указанных условиях отгадает, что же было задумано первым играющим. (Стр. 54)

<sup>30</sup> В «машину-ребенка». (Прим. ред.)

<sup>31</sup> Лучше сказать: «впрограммировать», так как наша «машина-ребенок» будет программироваться на цифровой вычислительной машине. Однако указанная логическая система не будет обучаемой.

<sup>32</sup> В случае, если в упомянутой выше системе логического вывода, встроенной в машину, содержатся соответствующие аксиомы и правила выведения следствий. (Прим. ред.)

<sup>33</sup> Здесь имеется в виду *иерархия типов*, предложенная Берtrandом Расселом с целью избежать противоречий (антиномий), обнаруженных в логике и теории множеств в конце XIX — начале XX столетий (см. 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., М., 1947, гл. IV, §§ 4–6 и Приложение I, а также С. К. К л и н и , Введение в метаматематику, перев. с англ., М., 1957, гл. III, §§ 11–12). Обнаружение этих противоречий сыграло значительную роль в выяснении трудностей, связанных с задачами логического обоснования математики, но на развитии самой теории множеств и особенно ее математических приложений стало сказываться лишь в самое последнее время в связи с созданием конструктивных математики и логики. Ни к каким противоречиям в самой математике применение так называемой «наивной» теории множеств, во всяком случае, не привело. Этим, по-видимому, объясняется

вероятность связанной с этой ошибкой, которую может сделать машина, больше вероятности того, что человек может упасть в пропасть, если ее край не будет огорожен. В рассматриваемом случае подходящие приказания (выраженные *внутри* системы формального вывода, а не составляющие часть ее правил), например, такие, как: «Не используйте некоторый класс, если он не является подклассом класса, который ранее упоминался учителем», могут иметь эффект, аналогичный тому, какой имеет предупреждение: «Не подходите слишком близко к краю обрыва».

Приказания, которые может выполнять машина, не имеющая ни рук, ни ног, должны касаться преимущественно интеллектуальных сторон деятельности, как это было в приведенном выше примере (с домашним заданием). Из такого рода приказов наиболее важными будут приказания, определяющие порядок, в котором следует применять правила рассматриваемой логической системы. Ибо на каждой стадии применения логической системы перед нами открывается большое число возможных шагов, которые исключают друг друга и любой из которых мы можем осуществить, следуя правилам рассматриваемой системы. Как производится такой выбор — в этом и выражается различие между глубоким и посредственным умом, но это не имеет отношения к правильности или неправильности рассуждений. Суждения, которые порождают приказания такого рода<sup>34</sup>, могут быть, например, такими: «Если упоминается Сократ, то применяй силлогизм модуса *Вагвага*» или: «Если один метод приводит к результату быстрее, чем второй, не применяй более медленный». Одни из них могут исходить от «авторитетного лица», другие же могут вырабатываться самой машиной, например, с помощью неполной индукции.

Некоторым читателям мысль об обучающейся машине может показаться парадоксальной. Как могут меняться правила, по которым машина производит операции? Ведь правила должны полностью описывать поведение машины независимо от того, какова была ее предыстория и какие изменения она претерпела. Таким образом, правила должны быть абсолютно инвариантными относительно времени. Все это, конечно, верно. Объяснение этого парадокса состоит в том, что правила, которые меняются в процессе научения, не претендуют на это, ибо их применимость носит преходящий характер. Читатель может провести параллель с Конституцией Соединенных Штатов<sup>35</sup>.

Важная особенность обучающейся машины состоит в том, что ее учитель в значительной мере не осведомлен о многом из того, что происходит внутри нее, хотя он все же в состоянии в известных пределах предсказывать поведение своей ученицы. Сказанное особенно применимо к дальнейшему воспитанию машины, прошедшей уже хорошую подготовку и вышедшей из начальной стадии «машины-ребенка». Такое положение, очевидно, в корне отличается от обычного подхода, связанного с применением машин для вычислений, когда мы стремимся к тому, чтобы иметь ясное представление о состоянии машины в любой момент вычисления, достичь чего можно лишь с трудом. В свете сказанного взгляд, что «машина может выполнить только то, что мы умеем ей предписать», кажется странным. Большинство программ, которые мы можем ввести в машину, вызывают в ее работе кое-что такое, что мы вообще не в состоянии осмыслить или рассматриваем как чисто случайное поведение. Интеллектуальное <*intelligent*> поведение предполагает, по-видимому, некоторое отступление от абсолютно детерминированного <*disciplinedly*> поведения в процессе вычисления; это отступление,

---

выраженная Тьюрингом ниже уверенность в том, что и без логического устройства, соответствующего иерархии типов, вероятность для машины впасть в логическую ошибку — при наличии у нее достаточно осторожного «учителя» — не больше вероятности для человека упасть в пропасть, не огороженную забором. (Смр. 55.)

<sup>34</sup> То есть приказания, касающиеся порядка применения правил логической системы. (Прим. ред.)

<sup>35</sup> Автор имеет здесь в виду то, что к Конституции США (выработана и утверждена в 1787–1789 гг.) — при сохранении ее основного содержания (изменения и дополнения к американской конституции обставлены весьма сложной процедурой) — за истекшее со времени ее принятия время был сделан целый ряд поправок (ныне их число составляет 22). (Смр. 56.)

однако, должно быть очень незначительным, чтобы не вызвать полностью беспорядочного поведения или бессмысленных повторений отдельных циклов. Другой важный результат обучения, как способа подготовки нашей машины для участия в игре в имитацию, состоит в том, что «присущая человеку склонность к ошибкам» будет, по-видимому, обойдена естественным образом, т. е. без специального «натаскивания». (Читатель должен связать это с точкой зрения, выраженной на страницах 39–41.) Процесс обучения не обязательно должен быть успешным во всех случаях; если бы это было так, то не встречались бы случаи неудачи в обучении.

Вероятно, в обучающуюся машину имеет смысл ввести случайный элемент (см. стр. 26). Случайный элемент довольно полезен, когда мы ищем решение какой-нибудь задачи. Пусть, например, требуется найти число, расположенное между 50 и 200 и равное квадрату суммы своих цифр; мы можем сначала проверить число 51, затем 52 и продолжать до тех пор, пока не найдем то, которое удовлетворяет условию задачи. Но мы можем поступить иначе: выбирать числа наугад до тех пор, пока не получим то, которое нам нужно. Этот метод имеет то преимущество, что не требует хранения в памяти уже проверенных значений; однако он имеет и отрицательную сторону, состоящую в том, что одно и то же число может быть подвергнуто проверке повторно, но это не так уж существенно, если задача имеет несколько решений. Систематический метод имеет другой недостаток: именно, может случиться, что придется проверять массу значений, не содержащих ни одного решения, прежде чем будет найдено первое число, обладающее нужным свойством.

В нашем случае процесс обучения можно рассматривать как поиски такой формы поведения, которая бы удовлетворяла требованиям учителя (или какому-нибудь другому критерию). Поскольку в этом случае, по-видимому, имеется весьма большое число решений, отвечающих предъявленным требованиям, постольку метод случайного выбора представляется нам предпочтительнее систематического. Следует отметить, что метод случайного выбора применяется и в другом аналогичном процессе — в эволюции. Но там систематический метод невозможен вообще. Неясно, каким образом, было бы возможно в процессе эволюции сохранять информацию о тех разнообразных генетических комбинациях, которые были испробованы, с тем, чтобы предупредить возможность их повторного применения.

Мы можем надеяться, что машины в конце концов будут успешно соперничать с людьми во всех чисто интеллектуальных областях. Но какие из этих областей наиболее пригодны для того, чтобы начать именно с них? Решение даже этого вопроса наталкивается на затруднения. Многие считают, что начать лучше всего с какой-нибудь очень абстрактной деятельности, например с игры в шахматы. Другие предлагают снабдить машину хорошими органами чувств, а затем научить ее понимать и говорить по-английски. В этом случае машину можно будет обучать, как ребенка: указывать на предметы и называть их и т. д. В чем состоит правильный ответ на этот вопрос, я не знаю, но думаю, что следует испытать оба подхода.

Мы можем заглядывать вперед лишь на очень небольшое расстояние, но уже сейчас очевидно, что нам предстоит еще очень многое сделать в той области, которая была предметом настоящей статьи.

---

# **ДЖ. фон НЕЙМАН ОБЩАЯ И ЛОГИЧЕСКАЯ ТЕОРИЯ АВТОМАТОВ<sup>36</sup>**

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.

---

<sup>36</sup> 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 приходим 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 система<sup>37</sup>. 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

---

<sup>37</sup> 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 синапс<sup>38</sup> — 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

---

<sup>38</sup> *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 метаболизм<sup>39</sup>. 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 лет<sup>40</sup>. 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)<sup>41</sup>.

---

<sup>39</sup> *Метаболизм* — обмен веществ. (Стр. 74.)

<sup>40</sup> 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.)

<sup>41</sup> *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,— которая 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.)<sup>42</sup> 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

---

<sup>42</sup> *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 метод<sup>43</sup> 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

---

<sup>43</sup> *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*<sup>44</sup> 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*<sup>45</sup> 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

---

<sup>44</sup> Сами по себе; как таковые (лат.) (Прим. ред.)

<sup>45</sup> 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*<sup>46</sup> 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.

---

<sup>46</sup> 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» логики<sup>47</sup>, 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.

---

<sup>47</sup> *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»<sup>48</sup>, 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.

---

<sup>48</sup> *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.)

## ИМЕННОЙ УКАЗАТЕЛЬ

Аристотель 5

Байрон Дж. (J. N. G. Byron) 104  
Барроу И. (I. Barrow) 103  
Бернайс П. (P. Bernays) 17  
Больцман Л. (L. Boltzmann) 81  
Брозин Г. (H. W. Brosin) 4  
Буш В. (V. Bush) 65, 104  
Бэббедж Ч. (Ch. Babbage) 27, 43, 44, 103, 104  
Вебстер Н. (N Webster) 106  
Вильяме Ф. (F. C. Williams) 103, 107

Галилей Г. 34  
Гёдель К. (K. Godel) 35  
Гэллуп Г. (G. Gallup) 19, 102

Джеффрис Л. (L. A. Jeffress) 4, 109  
Джефферсон (G. Jefferson) 37, 38, 43  
Дирак П. (P. A. M. Dirac) 103

Есенин-Вольпин А. С. 18

Касабьянка Л. (L. de Casablanca) 18, 108  
Келлер Е. (H. Keller) 53, 108  
Килберн Т. (T. Kilburn) 103, 107  
Клини С. К. (S. C. Kleene) 35  
Коперник Н. 34

Лавлейс (Ada Augusta, the Countess of Lovelace) 18, 43, 44, 49, 104  
Лаплас П. (P. Laplace) 29  
Ленин В. И. 10, 13, 18  
Листер Дж. (J. Lister) 103  
Лорент де Но (R. Lorente de No) 4  
Лэшли К. (K. S. Lashley) 4  
Люкас Г. (H. Lucas) 103

Маккаллох У. (W. S. McCulloch) 4, 6, 13, 88, 89, 109  
Марков А. А. 16  
Менабреа Л. (L. F. Menabrea) 104  
де Морган А. (A. de Morgan) 104

Нейман Дж. (J. von Neumann) 3, 7, 11, 13–18, 109  
Ньюман Дж. (J. V. Newman) 102  
Ньютон И. 103

Пикассо П. (P. Picasso) 36  
Питтс У. (W. Pitts) 88, 89

Рассел В. (B. Russell) 33, 108  
Россер Дж. (J. V. Rosser) 35

Скороходова О. И. 108  
Сократ 55, 105

Трахтенброт Б. А. 31, 36, 103

Тьюринг А. (A. M. Turing) 3–14, 16, 18, 31, 35, 93–99, 103, 104, 106, 107

Фома Аквинский 33

Хальстед У. (W. C. Halstead) 109

Хартри Д. (D. R. Hartree) 43

Хеманс Ф. (F. D. Hemans) 108

Хиксон Ф. (F. P. Nixon; 109

Чёрч А. (A. Church) 35