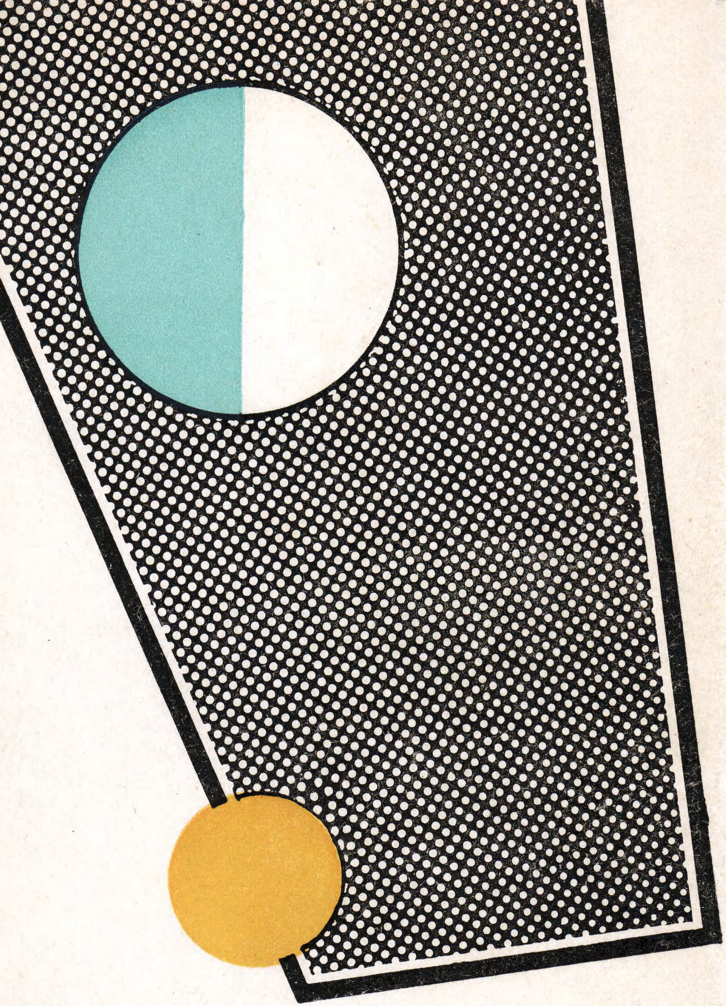
М. АРБИБ



Мозг, машина и математика

М. АРБИБ

Мозг, машина и математика

Перевод с английского
А. Д. КОРШУНОВА

Под редакцией
М. И. КРАТКО

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1968

6П2.154
А 79
УДК 519.95

Brains, Machines and Mathematics

Michael A. Arbib

McGraw — Hill Book Company
New York San Francisco Toronto
London

Майкл А. Арбиб

Мозг, машина и математика

М., 1968 г., 224 стр. с илл.

Редактор *Д. С. Фурманов*
Техн. редактор *И. Ш. Аксельрод*
Корректор *О. А. Сигал*

Сдано в набор 25/I 1968 г. Подписано к печати 8/VII 1968 г.
Бумага 84×108$^{1}/_{32}$. Физ. печ. л. 7. Условн. печ. л. 11,76.
Уч.-изд. л. 10,53. Тираж 50000 экз. Цена книги 55 коп.
Заказ № 1093.

Издательство «Наука»
Главная редакция
физико-математической литературы.
Москва, В-71, Ленинский проспект, 15.

Ленинградская типография № 2 имени Евгении Соколовой
Главполиграфпрома Комитета по печати
при Совете Министров СССР. Измайловский пр., 29.

$\frac{3-3-14}{235-67}$

Оглавление

От редактора русского перевода

Предлагаемая вниманию советского читателя книга М. Арбиба «Мозг, машина и математика» относится к разряду промежуточному между научно-популярными и научными книгами. Своеобразие ее в том, что она допускает несколько уровней изучения.

Читатель, не интересующийся глубоко проблематикой кибернетики, постарается разобраться только в определениях и теоремах, содержащихся в этой книге (не вникая в их доказательства), и понять суть рассматриваемых моделей. Для такой категории читателей книга может служить прекрасным научно-популярным руководством по кибернетике.

Значительно больше информации получит читатель, пытающийся самостоятельно воспроизвести полные доказательства в тех местах, где даются только их наметки. При таком подходе рассматривать эту книгу как научно-популярную могут только достаточно подготовленные в математическом отношении читатели — научные сотрудники, инженеры, учителя, аспиранты и студенты вузов.

И, наконец, книга может служить хорошим пособием для читателей, желающих глубоко изучить содержащийся в ней материал. Получив при первом беглом чтении книги общую картину развития соответствующих разделов кибернетики, такие читатели найдут многочисленные ссылки на работы других авторов, и захотят изучить их. Приводимая в книге краткая характеристика содержания этих работ, авторские комментарии и замечания помогут им в этом.

Следует отметить, что книга написана очень неровно. Некоторые разделы изложены достаточно подробно, ряд других — значительно менее подробно, чем того хотелось бы. Возникает впечатление, что автор сознательно включил в книгу побольше материала, даже в ущерб подробности изложения. Некоторые темы излагаются не совсем четко, во многих случаях даются только наметки на решение (например, почти все, что касается самовоспроизводящихся автоматов, многие места из главы «Кибернетика» и др.). В этих случаях читатель отсылается к специальной литературе, которая приводится как в сносках, так и в списке рекомендованной литературы.

Специально для русского издания автор прислал два дополнительных параграфа к первой главе (§§ 1.8 и 1.9). Поскольку в одном из них используются некоторые алгебраические понятия, оказалось целесообразным разъяснить их (приложение 2 написано редактором русского перевода).

Оригинал был рассчитан на американского читателя, и поэтому в нем почти нет ссылок на работы советских авторов. Во всех тех случаях, когда имеются переводы цитированных работ на русский язык или равноценные книги советских авторов, сделаны специальные примечания. Кроме того, в конце книги приводится список литературы, составленный из работ советских авторов и работ, переведенных на русский язык. Этот список был частично составлен автором для русского издания его книги и значительно дополнен переводчиком и редактором перевода. Список ни в коем случае не претендует на полноту и составлен (с учетом указанных выше трех категорий читателей), в основном из научных работ, хотя в нем и встречаются работы научно-популярного характера.

В виде приложения к книге приводится статья Э. Мура «Математика в биологии», в основном посвященная проблемам самовоспроизведения. Эта статья читается легче, чем книга, и может рассматриваться как своеобразное введение в книгу.

При редактировании мне большую помощь оказали многие товарищи, в первую очередь А. Колотов, В. Лозовский и П. Твердохлеб. Приношу им свою глубокую благодарность.

Новосибирск
Ноябрь 1967 г.

М. Кратко

Предисловие автора к русскому изданию

Проведя богатый событиями октябрь 1964 г. в Советском Союзе, я с большим удовольствием пользуюсь случаем предоставить настоящий перевод моей книги для усиления контактов с советскими учеными.

Эта книга написана в то время, когда я был в Массачусетском технологическом институте и находился под сильным влиянием Уоррена Маккаллока и Норберта Винера, у которых я тогда учился. Освещение тематики является достаточно широким, чтобы дать возможность войти в область кибернетики.

Во время работы над книгой я не считал себя компетентным во многих вопросах и поэтому мог лишь бегло упомянуть об искусственном интеллекте и эвристическом программировании (§ 4.5). Так как на русский язык переводится сборник «Вычислительные машины и мышление» под редакцией Е. А. Фейгенбаума и Дж. Фельдмана, то я рекомендую читателю этот сборник превосходных работ по данной тематике, а не беглое изложение этих вопросов в русском издании моей книги.

К первой главе русского издания я добавил два новых параграфа. Один из них посвящен последним достижениям в теории конечных автоматов, а другой — самовоспроизводящимся автоматам.

Сжатые сроки работы над книгой не дали мне возможности включить в эту книгу исследования русских ученых и поэтому я ограничиваюсь списком дополнительной литературы, которая введет русского читателя в курс отечественной кибернетической литературы.

Некоторые преподаватели предпочитают излагать кибернетику как математическую науку и поэтому ограничиваются изучением некоторых полностью формализованных разделов математики, которые можно рассматривать как основы кибернетики. Выбор тем в этом случае обусловлен такими, без сомнения, интересными исследованиями, как теория автоматов, самоорганизующиеся системы и теория алгоритмов. Эти педагоги понимают, что искусственный интеллект и изучение мозга как машины и т. п. представляют огромный интерес, но считают эти разделы еще недостаточно формализованными для математически мыслящих людей, изучающих кибернетику.

Я считаю нужным особенно подчеркнуть, что при всей важности таких направлений, как теория автоматов и теория алгоритмов, они не могут быть полным введением в общую кибернетику. Много тревог и сомнений в кибернетике вызывает проблема установления связи между вычислительными машинами или математическими моделями и сложными реальными биологическими и физиологическими системами. Существенно, что кибернетические модели имеют биологическую «основу» и эта «основа» не может быть развита посредством изучения систем, которые уже полностью формализованы.

Я представляю новое издание книги «Мозг, машина и математика» с надеждой, что мои читатели по настоящему «прочувствуют» реальные системы.

Станфордский университет,
Станфорд, Калифорния, США
Март 1966 г.

Майкл Арбиб

Посвящается
Фреду Поллоку,
мастеру образного,
остроумного и мудрого
преподавания
математики

Предисловие

Эта книга является введением в общую тему «мозг, машины и математика», где математика используется для установления аналогий между работой мозга и такими аспектами машин, как *управление, вычисление* и *связь*. Она предназначена для читателя, который в некоторой степени знаком с такими модными направлениями, как кибернетика, теория информации, теорема Гёделя, и желает знать о них больше по сравнению с тем, что он может найти в популярной литературе. Здесь читатель найдет не только то, *какими* являются некоторые результаты, но и *как* они получаются.

Эта книга имеет небольшой объем, так что для первого чтения потребуется несколько вечеров. Несмотря на это, в ней освещены многие вопросы, и читатель, который желает идти дальше, в некоторой степени будет подготовлен для серьезного чтения специальной литературы. Для подробного чтения всей книги требуется средняя математическая подготовка — один курс университетского обучения по вычислительной математике (или эквивалентная «математическая зрелость»). Однако большая часть книги должна быть понятна читателю, который опускает математические доказательства и предварительно незнаком с основами биологии или вычислительных машин.

Прежде чем говорить о содержании книги, я должен сказать несколько слов об истинном положении дел в нейрофизиологии и природе разрабатываемых нами моделей. Использование микроэлектродов, электронных микроскопов и радиоактивных индикаторов в течение последних десятилетий позволило значительно расширить наши познания в нейрофизиологии. Даже такой многотомный труд, как «Справочник по нейрофизиологии» («Handbook of Neurophysiology») не может полностью охватить все факты. Многие нейрофизиологические теории, какими бы общими они не были, с применением все более сложных приборов обнаруживают более тонкие структуры и более сложные химико-электрические клеточные механизмы. Это означает, что наши математические модели, описываемые в этой книге, основываются на весьма упрощенном взгляде на мозг и центральную нервную систему. Читатель может удивиться, что изучение таких систем может представлять интерес или иметь хоть какое-то значение. Между тем имеется много свойств таких, как память, вычисление, обучение, неопределенность, надежность при наличии ошибок в отдельных компонентах, которые, видимо, трудно отнести к «чистым механизмам». Однако в этом и состоит одна из основных причин нашего пристального внимания к ним. Исходя из математических моделей, мы доказываем существование электромеханических механизмов, которые обладают указанными свойствами. Другими словами, мы стараемся «отделить душу от машин». Мы еще не можем создавать такие механизмы, которые имеются в мозгу, но у нас есть по крайней мере модели возможных на этот счет механизмов, которые сами по себе представляют большой шаг вперед.

Имеется и другая причина такого изучения, которая с течением времени становится все более значительной. Прогресс в физической науке в основном связан с широким применением математико-дедуктивного и экспериментального методов. За последние 300 лет математика все более широко используется в физике: прикладная математика непосредственно, а чистая математика косвенно. Здесь мы вынуждены указать на то, что до сих пор математика в основном служит физике и медленно проникает в исследования мозга и его электромеханических аналогов. Возможно, что медленное развитие биологической математики будет наблюдаться до тех пор, пока не разовьются новые захватывающие разделы чистой математики.

Здесь, однако, мы применяем математику для получения общих выводов, вытекающих из имеющихся предпосылок. Мы можем проверить адекватность модели мозга, описывая ее в точных математических терминах и используя наш математический аппарат для доказательства математических теорем. В свете любых противоречий, обнаруженных между этими теоремами и экспериментами, мы можем вернуться к нашим предпосылкам, изменить их и таким образом добиться более глубокого понимания работы мозга. В дальнейшем такие теории помогут нам построить более «способные» и сложные машины.

Польза математико-дедуктивного метода состоит в том, что он позволяет нам устанавливать общие свойства наших моделей и, таким образом, является хорошим помощником при изучении моделей как в прямом, так и в переносном смысле.

Биологические системы обычно намного сложнее физических систем, так что мы не можем рассчитывать на то, что даже через несколько лет появится вполне удовлетворительная биологическая математика. Однако поиск такой математики является вполне реальным и заслуживает внимания. Я надеюсь, что результаты, которые читатель найдет в этой книге, будут для него интересными. Конечно, они составляют только очень небольшую часть того, что уже получено, но тем не менее и это является хорошим началом. Я не верю, что с применением математики удастся разрешить все наши физиологические и психологические проблемы. Однако считаю, что математико-дедуктивный метод должен занимать важное место наряду с экспериментами и клиническим изучением нейрофизиологов и психологов; он должен помочь нам понять мозг так же, как он уже помог инженерам-электрикам в построении электронных вычислительных машин, которые хотя и намного проще биологических организмов, но тем не менее имеют очень много аналогий с мозгом.

Теперь мы можем дать краткий обзор этой книги.

Сначала мы приведем очень беглый обзор нейрофизиологии и, исходя из него, опишем нашу первую грубую модель мозга в виде сети, компонентами которой являются нейроны Маккаллока — Питтса. Мы убедимся в том, что все, что может делать цифровая вычислительная машина, может делать и такая сеть. Мы выясним взаимосвязь этих сетей с конечными автоматами и машинами Тьюринга. В качестве примера сложной организации мозга приведем краткий обзор работ, посвященных изучению зрительной системы лягушки, и разберем персептрон (машину, которая «учится»). Затем приведем краткие неврологические сведения об ошибках в нейронах. Из этих сведений будет видна потребность в выяснении того, каким образом надо строить сети, которые надежно функционируют, несмотря на неправильную работу отдельных компонентов. После беглого рассмотрения одной работы фон Неймана мы изучим теорию связи Шеннона. Затем мы рассмотрим решение проблемы построения надежных устройств, предложенное Виноградом и Кованом. После этого мы вернемся к изучению кибернетики Норберта Винера и изучим управление и связь в животных и машинах. Мы рассмотрим основное понятие обратной связи, что позволит нам лучше разобраться в функционировании нервных сетей. Затем мы рассмотрим предложенную Грином схему резонансных частот в нервной сети для предостережения от слишком поспешного отождествления реального мозга с нервной сетью Маккаллока — Питтса. После обсуждения гомеостазиса и протезирования мы вернемся к образам и распознаванию общих понятий: каким способом мы постигаем звуковые и зрительные формы. Последнюю главу мы посвящаем теореме Гёделя о неполноте. Мы дадим исторический обзор тех направлений математической мысли, которые привели к работе Гёделя, докажем теорему, обсудим ее драматические философские последствия для оснований математики и, наконец, вернемся к спору о соотношении между машиной и мозгом.

Эта книга является обработкой цикла лекций, прочитанных в июне — августе 1962 г. в Университете Нового Южного Уэльса, Сидней, Австралия. Я выражаю благодарность Джону Блэтту, пригласившему меня прочесть лекции, Дерека Бруда за приглашение прочесть эти лекции по радио и Джойс Кин за превосходные машинописные копии прочитанных лекций.

Последние два года я провел в исследовательской лаборатории электроники и математическом отделении при Массачусетском технологическом институте

в качестве ассистента (работа финансировалась Министерством обороны США и Национальным институтом здравоохранения). Я должен благодарить так много людей, что не могу здесь этого сделать. Однако я особенно благодарен Уоррену Маккаллоку за его постоянную помощь, Джорджу В Цопфу, который первый настоял на опубликовании лекций. Билл Кильмер сделал полезные и критические замечания после прочтения оригинальных лекций. Годами я вынашивал надежду в этом месте сказать: «Только один он несет ответственность за оставшиеся ошибки». Однако это было бы неудачным выражением моей самой искренней благодарности ему, и я, следуя установившейся традиции, заявляю, что за все ошибки несу ответственность только я.

Наконец, я должен благодарить всех тех авторов, на работы которых я ссылался, и их издателей, любезно позволивших использовать эти материалы.

Майкл А. Арбиб

Глава 1

Нервные сети, конечные автоматы и машины Тьюринга

§ 1.1. Краткие сведения из нейрофизиологии

Я хочу начать с очень беглого обзора нейрофизиологии, который мог бы послужить основанием для нашей первой математической модели. Нервную систему человека можно представить себе как трехступенчатую систему, показанную на рис. 1.1 *). Основной гипотезой при создании модели будет предположение,

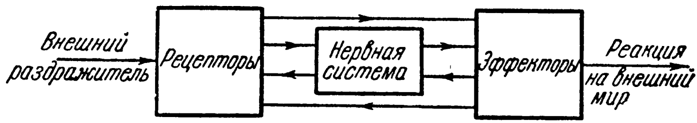


Рис. 1.1. Нервная система человека, рассматриваемая как трехступенчатая система.

что *относящееся к нашему рассмотрению функционирование нервной системы полностью определяется прохождением электрических импульсов по клеткам, которые мы называем нейронами.*

В действительности мозг человека содержит больше *глиальных* **) клеток, чем *нейронов.* До недавнего времени в нейрофизиологии считалось общепринятым, что назначение глиальных клеток состоит лишь в том, чтобы обслуживать, питать нейроны. Тогда это не относится к теме нашего рассмотрения. Однако за последние 15 лет высказывается все больше доводов в

*) Назначение стрелок, направленных влево, станет ясным из рассмотрения обратной связи в § 4.1.

**) *Глия* — ткань, сопровождающая нейроны и их отростки. *(Прим. перев.)*

пользу того, что глиальные клетки в действительности выполняют и другие функции, такие, например, как память, которые уже представляют для нас интерес. Тем не менее на протяжении этой книги мы не будем рассматривать подобные функции глиальных клеток. Мы также не будем рассматривать такие проявления взаимодействия нейронов, как постоянно меняющиеся потенциалы и обмен гормонами. Для нашего рассмотрения *возможных* механизмов нервной деятельности будет вполне достаточно импульсной активности нейронов. В дальнейшем, бесспорно, потребуется уделять гораздо больше внимания, чем сейчас, остальным нейрофизиологическим механизмам, а возможно, и глии.

В свете принятой нами основной гипотезы мы будем рассматривать нервную систему как обширную сеть нейронов, имеющую сложную структуру и исключительно сложные внутренние соединения. Эта сеть воспринимает информацию от большого числа рецепторов: палочек и колбочек глаз, тактильных рецепторов кожи, температурных рецепторов и т. д., которые преобразуют возбуждения, поступающие из внешней среды или возникающие внутри тела человека, во множество электрических импульсов, посредством которых информация передается в сеть. Эти импульсы взаимодействуют с чрезвычайно сложными «картинами» импульсов, которые циркулировали до этого в нейронах (установлено, что в нервной сети человеческого мозга содержится порядка 10^{10} нейронов!), в результате чего на выходе сети появляются импульсы, управляющие эффекторами, например мускулами или железами. В этом и выражается реакция нашего организма на раздражение. Таким образом, нервную систему человека можно условно разделить на три основные подсистемы: рецепторы, нервная сеть и эффекторы.

Здесь мы не собираемся создавать модели рецепторов или эффекторов; нас интересует модель нервной сети. Для этого необходимо предварительно создать модель нейрона. Нейроны нашей нервной системы весьма разнообразны по форме, но мы ограничимся рассмотрением нейронов, имеющих вид, показанный на рис. 1.2.

Нейрон, как и все клетки, имеет ядро, содержащееся в *соме* или *теле* клетки. *Дендриты* *) можно представлять себе в виде чрезвычайно разветвленного куста, каждое волокно которого тоньше аксона, а сам *аксон* представляет собой длинный и тонкий цилиндр,

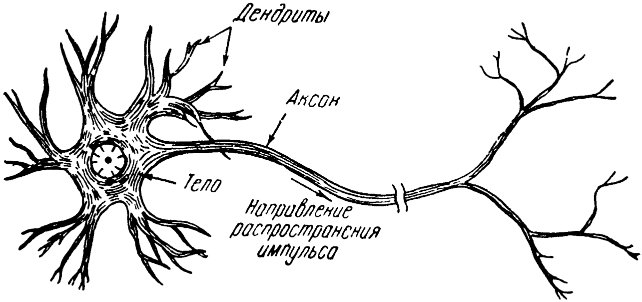


Рис. 1.2. Схематическое изображение нейрона.

передающий импульсы от сомы к другим клеткам. Аксон переходит в густое древовидное разветвление, ветви которого заканчиваются маленькими *пластинками* (бляшками), которые почти касаются дендритов нейрона. Такой участок близкого контакта называется *синапсом* **). Достигающие синапса импульсы вызывают переменный электрический сигнал в дендритах ***), примыкающих к данному синапсу. Передача сигналов осуществляется иногда электрическими импульсами, иногда путем химической диффузии. Нейрон выдает импульс (поступающий на аксон) лишь

*) Дендрит проводит нервный импульс с периферии к телу клетки. (*Прим. перев.*)

**) Синапсы — образования, в которых происходит контакт нервных клеток друг с другом. По синапсу возбуждение передается только в одном направлении (как правило, лишь с окончаний аксона одного нейрона на дендриты другого). (*Прим. перев.*)

***) Возможен также синапсический переход на другой аксон. Это «афферентное взаимодействие» рассматривается в § 31.

в том случае, когда на концевые бляшки, находящиеся у его дендритов, поступает достаточное число импульсов в очень малый промежуток времени, называемый *периодом латентного накопления.* В действительности эти импульсы могут либо способствовать, либо препятствовать возбуждению нейрона и соответственно называются *возбуждающими* или *тормозящими.* Условие выдачи импульса нейроном заключается в том, чтобы возбуждение превосходило торможение на некоторую критическую величину, называемую *порогом нейрона.* Если каждому возбуждающему синапсу приписать подходящий положительный вес, а каждому тормозящему синапсу — подходящий отрицательный вес, то можно сказать, что

нейрон возбуждается тогда и только тогда, когда сумма весов синапсов, на которые поступают импульсы в период латентного накопления, превышает порог нейрона. (1.1.1)

Предположение о таком простом линейном суммировании опять-таки является очень сильным упрощением. Далее, порог представляет собой параметр, меняющийся во времени; однако это изменение редко рассматривается при моделировании формального нейрона и не учитывается в моделях, которые будут рассматриваться ниже. Читателю, готовому впасть в отчаяние при виде все возрастающего разрыва с действительностью, можно рекомендовать для поднятия духа перечитать предисловие.

Между периодом латентного накопления и прохождением соответствующего импульса по аксону и концевым бляшкам существует небольшое время задержки, так что поступление импульсов на дендриты нейрона повлечет появление импульса на его аксоне лишь через некоторое время.

После того как по аксону прошел импульс, наступает *период рефрактерности*, в течение которого аксон не способен передавать импульсы. Следовательно, за период времени, равный одному периоду рефрактерности, по аксону может быть передан не более чем один импульс. Выбрав в качестве единицы измерения времени период рефрактерности нейрона, можно описать поведение нейрона, указав для каждого интервала времени, когда нейрон возбужден, а когда не возбужден.

Таким образом, мы ввели упрощающее предположение (еще более удаляющее нас от действительности), что наш нейрон может возбуждаться только в моменты времени $t=1, 2, 3, 4, \ldots$ при наличии подходящих условий. Наконец, введем еще одно *упрощающее* предположение: будем использовать единую *дискретную шкалу времени* для всех нейронов сети, т. е. будем считать, что любое возбуждение нашей сети полностью определяется картиной возбуждения отдельных нейронов в дискретные моменты времени $t=1, 2, 3, \ldots$ В связи с этим мы полагаем, что выдача импульса на аксон в данный момент времени полностью определяется возбуждением синаптических входов этого нейрона в предыдущий момент времени.

§ 1.2. Модель Маккаллока — Питтса

Исходя из весьма упрощенных нейрофизиологических рассмотрений предыдущего параграфа, введем следующую модель нейрона (или модуля), которая впервые была предложена Маккаллоком и Питтсом.

Определение 1.2.1. *Модулем* (*формальным нейроном*) называется элемент с m входами $x_1, \ldots, x_m$ $(m \geqslant 1)$ и одним выходом d. Он характеризуется $m+1$ числом: порогом θ и весами $\omega_1, \ldots, \omega_m$. Каждому входу x_i $(1 \leqslant i \leqslant m)$ сопоставляется вес ω_i. Модуль работает в дискретном времени $t=1, 2, 3, 4, \ldots$ Его выход в момент $n+1$ зависит только от входов в момент n. Эта зависимость задается следующим правилом (ср. с утверждением (1.1.1.)): модуль в момент $n+1$ передает импульс по своему аксону в том и только в том случае, когда сумма всех весов возбужденных входов в момент n превышает порог модуля.

Введем обозначения:

$x_i(t)=0$ для «i-й вход не возбужден в момент t»,

$x_i(t)=1$ для «i-й вход возбужден в момент t»,

$d(t) = 0$ для «выход не возбужден в момент t»,
$d(t) = 1$ для «выход возбужден в момент t».

Тогда приведенное выше правило формально можно записать следующим образом:

$$d(n+1) = 1$$

тогда и только тогда, когда

$$\sum \omega_i x_i(n) \geqslant \theta.$$

Заметим, что положительный вес $\omega_i > 0$ указывает на то, что i-й вход (синапс) является возбуждающим, а отрицательный вес $\omega_i < 0$ означает, что i-й вход является тормозящим входом.

Пользуясь этой крайне упрощенной моделью нейрона, определим теперь модель нервной сети.

Определение 1.2.2. *Модульной сетью* (абстрактной нервной сетью) называется множество модулей, соединенных между собой так, что выход каждого модуля ветвится на подходящее число линий, каждая из которых присоединяется к некоторому входу какого-либо модуля. Выход каждого модуля может быть соединен с любым числом входов, но каждый вход не может быть соединен более чем с одним выходом. Все модули сети работают в одной и той же шкале времени.

Входными линиями сети называются те входы $l_0, l_1, \ldots, l_{m-1}$ модулей сети, которые не подсоединены к выходам модулей*). *Выходными линиями* сети называются те линии $p_0, p_1, \ldots, p_{r-1}$ выходов модулей, которые не соединены со входами модулей.

В сети, показанной на рис. 1.3, имеются три входные и четыре выходные линии. Отметим, что входные линии могут ветвиться, а выходные линии не обязаны идти от различных модулей.

---

*) Это не совсем точно. Свободные входы нейронов $l_0, l_1, \ldots, l_{m-1}$ в действительности не являются входными линиями, а присоединяются к входным линиям сети. Каждый из этих входов соединен в точности с одной входной линией. Таким образом, в сети имеется не более m входных линий (см. рис. 1.3, где свободных входов 5, а входных линий). (*Прим. ред.*)

Читателю, который представляет себе модель как реальную схему с проводами и транзисторами, может показаться несколько странным употребление нами слова «модель». Поэтому позвольте обратить внимание на то, что в этой книге слово «модель» используется только в математическом смысле. *Инженер* говорит, что он имеет модель системы, если он фактически построил аппарат, который будет работать так

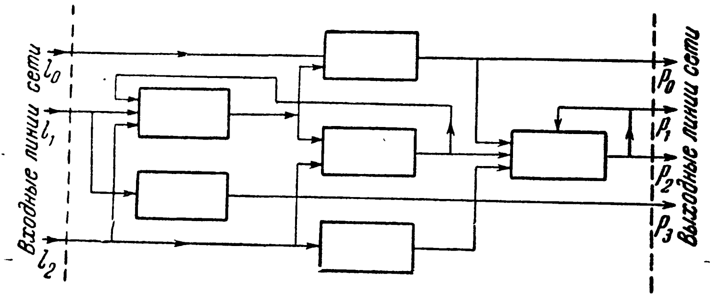


Рис. 1.3. Простая модульная сеть.

же, как и первоначальная система. *Математик* же говорит, что он построил модель системы, если он «выразил» некоторые свойства этой системы в точных терминах математических определений и аксиом так, что он может дедуктивно вывести такие новые свойства этой «формальной» (т. е. математической) *модели*, которые, как он ожидает, *объяснят* известные свойства оригинальной системы и *предскажут* новые, ранее неизвестные свойства. Понятие «модульная сеть» определено в точных математических терминах (в следующих параграфах будут доказаны теоремы о модульных сетях), поэтому в этом математическом смысле мы и рассматриваем ее как модель мозга.

Прежде чем приступить к изучению этой модели, укажем на то, *что она получена при следующих весьма существенных ограничениях*: а) предполагалось, что все нейроны работают синхронно; б) предполагалось, что порог и вес каждого нейрона не меняются с течением времени; в) мы не учитывали химических воздействий (например, алкоголя) и влияния гормонов на изменение поведения мозга; г) мы пренебрегли всеми взаимодействиями между нейронами (например, с помощью электрического поля, возникающего в результате импульсов), кроме синапсических; д) нами игнорировались глиальные клетки.

Этот список ограничений можно продолжить. Отсюда следует, *что предложенная модель является только отправным пунктом для изучения и, конечно, далека от окончательной.* Несмотря на это, принятые ограничения не обесценили полностью нашу модель, ибо модульные сети могут хранить информацию и производить вычисления. Эти возможности модульных сетей будут продемонстрированы в § 1.4 при рассмотрении цифровых вычислительных машин как модульных сетей. Но сначала в § 1.3 мы введем понятие «конечный автомат» и выясним его взаимосвязь с модульной сетью. Кроме того, в § 1.7 мы рассмотрим одну простейшую модель восприятия, которая задается в точных математических терминах входных последовательностей конечного автомата.

§ 1.3. Конечные автоматы и модульные сети

Понятие конечного автомата мы введем так, чтобы было очевидно, что любая модульная сеть является конечным автоматом. Тогда основная цель будет заключаться в установлении того факта, что поведение «вход — выход» любого конечного автомата всегда можно воспроизвести на подходящим образом построенной модульной сети. Так как построение конечного автомата для определенных целей, вообще говоря, намного легче, чем построение соответствующей модульной сети, то этот результат позволяет судить о том, что могут делать модульные сети (см. § 1.4).

Пусть модульная сеть N состоит из g модулей, имеет m входных и r выходных линий.

Будем говорить, что задан *вход* сети N, если известно, какие входные линии возбуждены, а какие — не возбуждены. Так как значения «возбужден» и «не возбужден» m входным линиям можно приписать 2^m

различными способами, то в сети N имеется 2^m входов. Аналогично в ней имеется 2^r *выходов*.

Будем говорить, что задано *состояние* сети N в момент t, если известно, какие модули в этот момент возбуждены, а какие не возбуждены. Таким образом, N имеет 2^g состояний.

Обозначим через Q множество состояний, через X — множество входов и через Y — множество выходов сети N. Возбуждение любого модуля сети N в момент $t+1$ определяется тем, какие входы этого модуля возбуждены в момент t и, следовательно, определяется состоянием и входом всей сети N в момент t. Но это означает, что вход и состояние сети в момент t однозначно определяют выход и состояние сети в момент $t+1$. Следовательно, конечный автомат теперь можно определить как простое обобщение модульной сети. В дальнейшем конечный автомат будем рассматривать как «черный ящик» с конечным числом входов, конечным числом «внутренних состояний» (это могут быть либо состояния модульной сети, либо позиции зубцов часового механизма и т. п.) и конечным числом выходов. Как и выше, мы предполагаем, что вход и состояние в момент t определяют выход и состояние в момент $t+1$. Более точно эти требования можно сформулировать следующим образом.

Определение 1.3.1. *Конечным автоматом* называется пятерка

$$A=(X,\ Y,\ Q,\ \lambda,\ \delta),$$

где X — конечное множество (множество входов), Y — конечное множество (множество выходов), Q — конечное множество (множество внутренних состояний), $\lambda: Q\times X\to Q$ — функция перехода *), $\delta: Q\times X\to Y$ — функция выхода. Предполагается, что A работает в дискретной шкале времени так, что если в момент t он находился в состоянии q и воспринимал вход a, то в момент $t+1$ он перейдет в состояние $\lambda(q, a)$, а $\delta(q, a)$ будет его выходом.

*) Объяснение этих и ряда следующих понятий читатель найдет в Приложении 1.

Легко видеть, что любая модульная сеть является конечным автоматом. Более неожиданным является тот факт, что любой конечный автомат можно заменить подходящей модульной сетью. Так как доказательство этого факта не требует привлечения никаких глубоких идей и не имеет большого значения для последующего изложения, то читатель, который не интересуется им, сразу может переходить к § 1.4.

Действительно, пусть конечный автомат A имеет m возможных входов $x_0, x_1, \ldots, x_{m-1}$ и r возможных выходов $y_0, y_1, \ldots, y_{r-1}$. Сопоставим этому автомату модульную сеть N с *m входными линиями* $h_0, h_1, \ldots$ $\ldots, h_{m-1}$ (заметим, что N, таким образом, имеет 2^m *входов*) и r выходными линиями $p_0, p_1, \ldots, p_{r-1}$. Вход x_j автомата A сопоставим такому входу $\bar{x}_j$ сети N, в котором возбуждена только h_j-я входная линия. Аналогично определим выход $\bar{y}_j$ сети N. Искомая сеть N состоит из $mn+r$ модулей*): mn модулей сопоставляются возможным парам «вход — состояние» автомата A, а остальные r модулей — выходам A. Модуль, соответствующий k-му состоянию и j-му входу автомата A, обозначим (k, j), а модуль, который соответствует k-му выходу, обозначим (k). Выходная линия p_k нашей сети N подсоединена к модулю (k).

Модули соединим таким образом, чтобы модуль (k, j) возбуждался в момент $t+1$ тогда и только тогда, когда автомат A в момент t имеет вход x_j и находится в k-м состоянии; модуль (k) возбуждается в момент $t+1$ тогда и только тогда, когда автомат A в момент t дает выход y_k.

Пусть $\{k_1, k_2, \ldots, k_{n(k)}\}=\{(i, j) \mid \lambda(i, j)=k\}$, т. е. это множество таких пар *«состояние — вход»*, которые переводят автомат A в *состояние* k. Тогда модуль (k, l) должен быть возбужден в момент $t+1$ тогда и только тогда, когда **)

$$h_l(t) \wedge [k_1(t) \vee \ldots \vee k_{n(k)}(t)],$$

---

*) Предполагается, что автомат A имеет n внутренних состояний. (*Прим. ред.*)

**) $\wedge$ обозначает логическое «и», а $\vee$ — логическое «или». Заметим, что запись $k_1(t)$ надо читать «k_1 возбуждено в момент t» и т. д.

т. е. тогда и только тогда, когда автомат A в момент t находится в k-м состоянии и имеет l-й вход.

Пусть

$$\{\overline{k}_1, \overline{k}_2, \ldots, k_{m(k)}\} = \{(i, j) \mid \delta(i, j) = k\},$$

т. е. это множество таких пар «состояние — вход», при которых автомат A имеет *выход* k.

Тогда модуль (k) возбужден в момент $t+1$ тогда и только тогда, когда $\overline{k}_1(t) \vee \ldots \vee \overline{k}_{m(k)}(t)$.

Легко убедиться, что каждый нейрон сети N действительно может быть реализован при условии подходящего выбора весов и порога. Таким образом, наша сеть N удовлетворяет следующей теореме.

Теорема 1.3.1. *Пусть* $A = (X, Y, Q, \lambda, \delta)$ — *произвольный конечный автомат, где*

$$X = \{x_0, \ldots, x_{m-1}\}, \quad Y = \{y_0, \ldots, y_{r-1}\},$$
$$Q = \{q_0, \ldots, q_{n-1}\}.$$

Тогда существуют модульная сеть N *и подмножества ее входов* $\{\overline{x}_0, \ldots, \overline{x}_{m-1}\}$, *выходов* $\{\overline{y}_0, \ldots, \overline{y}_{r-1}\}$ *и состояний* $\{\overline{q}_0, \ldots, \overline{q}_{n-1}\}$ *такие, что если автомат* A, *первоначально находящийся в состоянии* q_j, *под действием входа* $x_{j_1}, \ldots, x_{j_s}$ *выдает* $y_{k_1}, \ldots, y_{k_s}$, *то сеть* N, *первоначально находящаяся в состоянии* $\overline{q}_j$, *под действием входов* $\overline{x}_{j_1}, \ldots, \overline{x}_{j_s}$ *выдает* $\overline{y}_{k_1}, \ldots, \overline{y}_{k_s}$ (*возможно с единичной задержкой*). Это означает, что любой автомат с заданным поведением «вход — выход» можно заменить подходящей модульной сетью с тем же поведением.

§ 1.4. Конечные автоматы и цифровые вычислительные машины

Исходя из того (показано в § 1.3), что любой конечный автомат можно заменить модульной сетью, теперь нетрудно убедиться в том, что модульные сети способны хранить информацию и производить вычисления. Для этого достаточно показать, что любая цифровая вычислительная машина *) может быть представлена в виде надлежащего соединения конечных автоматов, поскольку нам известно, что (возможно после некоторой возни с временными задержками) мы сможем заменить эти автоматы модульными сетями. Цифровая вычислительная машина, снабженная самой лучшей программой (списком команд) 1964 г., является, конечно, менее «интеллектуальным» объектом **), нежели мозг, получивший лучшее образование 1964 г. Здесь же мы только хотим показать, что наша модель мозга, как бы груба она ни была, относится к категории вычислительных машин.

Любая цифровая вычислительная машина обычно состоит из четырех устройств: устройства ввода и вывода, устройства памяти, устройства управления и арифметического устройства. Устройство ввода и вывода служит для чтения команд и исходных данных с входной ленты и их пересылки в устройство памяти и, наоборот, для вывода из устройства памяти результатов вычислений на выходную ленту. Устройство памяти состоит из конечного числа «ячеек», каждая из которых имеет свой адрес и вмещает одно «слово», где «слово» может быть как числом (исходные данные, промежуточный или окончательный результат), так и командой. Устройство управления в каждый такт работы машины выбирает из устройства памяти по одной команде и выполняет ее. Если эта команда является операцией ввода или вывода или операцией обращения к устройству памяти, то устройство управления приводит в действие либо устройство ввода и вывода, либо устройство памяти; если команда является арифметической операцией, то устройство управления приводит в действие арифметическое устройство; если условным переходом, то оно осущест-

*) Имеется два основных типа вычислительных машин — аналоговые и цифровые. Здесь обсуждаются только цифровые вычислительные машины.

**) Однако уже имеются такие программы, которые позволяют производить вычисления с целью демонстрации различных сторон «интеллектуального» поведения; краткое описание двух из них приводится в § 4.5.

вляет необходимую проверку и принимает решение о выполнении следующей команды.

Для определенности рассмотрим вычислительную машину, команды которой состоят из кода выполняемой операции и трех адресов памяти:

ОП адрес 1-го числа адрес 2-го числа адрес следующей команды.

По первому и второму адресам устройство управления отыскивает данные, которые нужны для выполнения *ОП* (операции), а последний адрес указывает, где найти следующую команду. Например, команда

Сл. 3275 3119 4006

должна быть истолкована следующим образом: сложить числа, находящиеся в ячейках (устройства памяти) с адресами 3275 и 3119, и перейти к выполнению команды, находящейся в ячейке с адресом 4006.

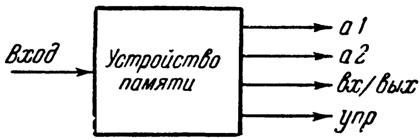


Рис. 1.4. Автомат «устройство памяти».

Теперь опишем конечный автомат «устройство памяти». Состояние устройства памяти (рис. 1.4), когда оно содержит слова $x_1, x_2, \ldots, x_n$ в своих n «ячейках», обозначим $(x_1, x_2, \ldots, x_n)$. Его входы имеют вид $(m, b, 0)$ и $(m, 0, d)$. Если входом является $(m, b, 0)$, то слово b поступает в m-ю ячейку устройства памяти, т. е. состояние $(x_1, \ldots, x_{m-1}, x_m, x_{m+1}, \ldots, x_n)$ заменяется состоянием $(x_1, \ldots, x_{m-1}, b, x_{m+1}, \ldots, x_n)$, а выход отсутствует. Устройство памяти имеет четыре выходные линии, две из которых $a1$ и $a2$ связаны с арифметическим устройством, одна *вх/вых* — с вход-выходным устройством и одну *упр* — с устройством управления. Если входом является $(m, 0, d)$, то слово, находящееся в ячейке с адресом m, отсылается по выходной линии d (где d есть $a1$, $a2$, *вх/вых* или *упр*) в соответствующее устройство, т. е. состояние $(x_1, \ldots \ldots, x_m, \ldots, x_n)$ остается неизменным, а x_m поступает на выход d. Ясно, что устройство памяти является конечным автоматом.

Арифметическое устройство (рис. 1.5) имеет три входные линии: одну (*ОП*) из устройства управления

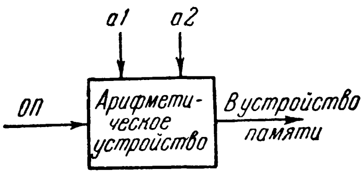


Рис. 1.5. Автомат «арифметическое устройство».

и две ($a1$ и $a2$) из устройства памяти. Оно состоит из двух регистров, а его состоянием является пара (B_1, B_2), где B_1 — содержимое первого, а B_2 — содержимое второго регистра. Входами арифметического устройства являются тройки вида (*ОП*, $a1$, $a2$). Если входом является тройка $(0, a1, 0)$, то $a1$ записывается в первый регистр: состояние (B_1, B_2) переходит в состояние $(a1, B_2)$, а на выход ничего не поступает. Аналогично вход $(0, 0, a2)$ состояние (B_1, B_2) переводит в состояние $(B_1, a2)$. В случае входа (*ОП*, 0, 0) над (B_1, B_2) производится *ОП*, вычисляется результат C, некоторый промежуточный результат D, и арифметическое устройство переходит в состояние (C, D). В случае входа (*Пам*, $a1$, 0) арифметическое устройство отсылает результат вычислений в ячейку памяти $a1$, т. е. состояние (C, D) остается неизменным, а выход $(a1, C, 0)$ поступает в память. Ясно, что арифметическое устройство является конечным автоматом — оно по-разному реагирует на каждую пару «вход — состояние», число которых конечно.

Входами устройства управления (рис. 1.6) являются команды программы, находящейся в устройстве памяти. Вход *ОПabc* переводит его в состояние (*ОП*, *a*, *b*, *c*). Если *ОП* является кодом некоторой функции, то на выход устройства управления один за другим выдаются четыре значения: (*a*, 0, *a*1) и (*b*, 0, *a*2) поступают в устройство памяти для вызова исходных данных; (*ОП*, 0, 0) поступает в арифметическое устройство для выполнения этой операции и, наконец,

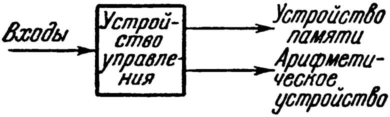


Рис. 1.6. Автомат «устройство управления».

(*c*, 0, *упр*) поступает в устройство памяти для вызова очередной команды, которую должно выполнять устройство управления.

Если *ОП* является операцией обращения к памяти *Пам*, то возникает два выхода: сначала выход (*Пам*, *a*, 0) поступает в арифметическое устройство, а затем выход (*c*, 0, *упр*) — в устройство памяти.

Если *ОП* является операцией условного перехода *УП* (так что *УП abc* означает следующее: надо проверить некоторое свойство слова, находящегося в ячейке с адресом *a*; если данное слово обладает этим свойством, то устройство управления должно переходить к выполнению команды, находящейся в ячейке с адресом *b*, в противном случае оно должно переходить к выполнению команды, находящейся в ячейке с адресом *c*), то устройство управления за конечное число шагов должно выполнить соответствующую проверку и в устройство памяти выдать либо (*b*, 0, *упр*), либо (*c*, 0, *упр*). Так или иначе устройство управления является конечным автоматом.

Этого достаточно для построения нашей вычислительной машины. Правда, здесь не рассмотрено устройство ввода — вывода как конечный автомат (а также вид соответствующих входов для устройства управления). Предлагаем это читателю в качестве самостоятельного упражнения.

Читатель, знакомый с устройством вычислительных машин, знает, что они представляют собой схемы, состоящие из десятков тысяч элементов: магнитных сердечников, транзисторов, триггеров. Это помогаст читателю убедиться, что вычислительная машина является модульной сетью, хотя эти модули и не являются в строгом смысле нейронами Маккаллока — Питтса. Для новичка, однако, приведенное выше рассмотрение должно дать некоторую полезную информацию сверх того, что вычислительная машина представляет собой набор радиодеталей. Отметим, что различные вычислительные машины имеют неодинаковую внутреннюю структуру, так что мы лишь обсуждаем общие идеи, а не конкретные схемы существующих машин.

Закончим этот параграф двумя ссылками. Небольшим по объему является доступный труд John von Neumann «The Computer and the brain» *).

Подробное описание вычислительных машин и принципов их работы имеется в «Computers and Data Processing» ed. Grabbe, Ramo and Wooldrige «Handbook of automation, computation and control», v. 2, J. Wiley & Sons, Inc., N. Y., 1958—1961.

§ 1.5. Машины Тьюринга

Из приведенных выше рассуждений следует, что любая цифровая вычислительная машина является конечным автоматом или модульной сетью, осуществляющей определенные операции над своими входами и выходами, причем так, что вход в данный момент может быть не использован, возможно использование более ранних входов и т. д. Покажем теперь, что любое вычисление, которое можно выполнить на цифровой вычислительной машине (используя конечную

*) Yale University Press, New Haven, Conn., 1958. [Русский перевод: Джон фон Нейман, Вычислительная машина и мозг, «Кибернетический сборник», вып. 1, ИЛ, 1960.]

программу), может быть также выполнено на машине Тьюринга, где

Определение 1.5.1. *Машиной Тьюринга Z* называется конечный автомат A, снабженный потенциально бесконечной лентой, разбитой по длине на одинаковые квадраты (ячейки), и устройством D для считывания содержимого ячейки ленты в каждый момент времени, печатания нового символа в этой (обозреваемой) ячейке и передвижения ленты на одну

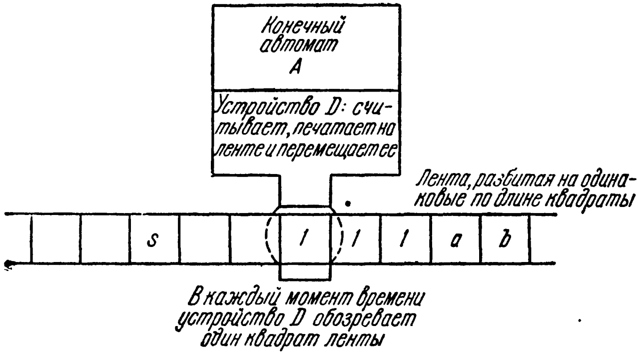


Рис. 1.7. Машина Тьюринга.

ячейку влево или вправо. Символы ленты принадлежат конечному алфавиту X (для удобства считаем, что в X содержится пустой символ $\wedge$, т. е. пустая ячейка). Каждая ячейка ленты либо пуста, либо в ней содержится только один символ. В каждый момент времени лента содержит только конечное число непустых символов.

Входом для A (рис. 1.7) в каждый момент времени является символ из X, считываемый устройством D. Выходом A, который однозначно определяется по считываемому символу из X и внутреннему состоянию A, является элемент из $(X \times M) \cup$ (стоп) *), где $M = \{Л, П, Н\}$. Если выходом является «стоп», то

*) Напомним (см. Приложение 1), что $\cup$ означает теоретико-множественное объединение.

машина Z прекращает вычисления. Если выходом является пара (x, m), то автомат A с помощью устройства D печатает *) в обозреваемую ячейку символ x и перемещается на одну ячейку влево (если $m=Л$), вправо (если $m=П$) или остается на месте (если $m=H$).

Определение 1.5.2. Каждой машине Тьюринга Z сопоставим функцию $F_Z(n_1, n_2, \ldots, n_r)$ такую, что если Z в начальном состоянии q_0 обозревает самый левый символ слова

$$\underbrace{1 \ldots 1}_{n_1+1} \wedge \underbrace{1 \ldots 1}_{n_2+1} \wedge \ldots \wedge \underbrace{1 \ldots 1}_{n_r+1},$$

по обе стороны которого находятся только пустые символы, то $F_Z(n_1, n_2, \ldots, n_r)$ равно числу единиц, произвольным образом размещенных на ленте машины Z в момент ее остановки; если Z не останавливается, то считается, что функция F_Z не определена **).

Читатель должен помнить, что в приведенном выше определении предполагается, что машина Z обращается с пустой ячейкой ленты так, как будто бы в ней напечатан специальный символ «$\wedge$».

Введем специальное название для функции F_Z одного аргумента.

Определение 1.5.3. Функция f натурального аргумента ***) называется *частично рекурсивной*, если

$$f(n) = F_Z(n)$$

для некоторой машины Тьюринга Z и всех натуральных чисел (считается, что если одна часть равенства неопределена, то такой же является и вторая часть этого равенства).

Частично рекурсивная функция называется *рекурсивной*, если она определена при всех n.

---

*) Ранее записанный символ стирается. (*Прим. перев.*)

**) На данном наборе $(n_1, \ldots, n_r)$. (*Прим. ред.*)

***) Под множеством натуральных чисел понимается множество $N=\{0, 1, 2, 3, \ldots\}$.

В определении 1.5.2 содержится ряд ограничений *). Однако общность этого определения как раз и заключается в том, что множество всех функций F_Z (когда Z принимает всевозможные значения) остается неизменным независимо от того, имеют или не имеют место эти ограничения **). В частности, нам достаточно того, что класс всех частично рекурсивных функций формально (т. е. математически) выражает интуитивные представления о функциях, которые могут быть вычислены на цифровых вычислительных машинах.

Пример ***). Построим машину Тьюринга (+) для сложения двух натуральных чисел, т. е.

$$F_{(+)}(n_1, n_2) = n_1 + n_2.$$

Машина (+), начав работу со слова

$$\underbrace{1 \ldots 1}_{n_1+1} \wedge \underbrace{1 \ldots 1}_{n_2+1},$$

должна при остановке иметь на ленте n_1+n_2 единиц. Это, конечно, очень просто. Легко видеть, что машина (+) задается автоматом $(\{1, \wedge\}, ((\wedge) \times M) \cup (\text{стоп}), \{q_0, q_1, q_2\}, \lambda, \delta)$, где

$$\begin{aligned}
\lambda(q_0, 1) &= q_1, & \delta(q_0, 1) &= (\wedge, \Pi),\\
\lambda(q_1, \wedge) &= q_1, & \delta(q_1, \wedge) &= (\wedge, \Pi),\\
\lambda(q_1, 1) &= q_2, & \delta(q_1, 1) &= (\wedge, \Pi),\\
& & \delta(q_2, \wedge) &= \delta(q_2, 1) = \text{стоп},
\end{aligned}$$

а в остальных случаях функции λ и δ определяются произвольно.

---

*) К таким ограничениям, например, относятся: способ записи исходных данных на ленте машины Тьюринга; в начальный момент в «поле зрения» машины всегда находится самая левая непустая ячейка; в конце работы на ленте должно находиться только значение функции и т. п. (*Прим. перев.*)

**) Этого мы здесь не доказываем. Интересующегося читателя отсылаем к книге: M. Davis, Computability and unsolvability [9].

***) Более содержательные и поучительные примеры содержатся, например, в брошюре Б. А. Трахтенброт, Алгоритмы и машинное решение задач, Физматгиз, М., 1960. (*Прим. перев.*)

Читатель может проследить работу машины Тьюринга (+) по приводимой ниже таблице (состояние записывается левее символа, который машина обозревает в данный момент времени).

$n_1 = 0$	$n_1 > 0$
$q_0 1 \wedge 1 \ldots$	$q_0 11 \ldots$
$\wedge q_1 \wedge 1 \ldots$	$\wedge q_1 1 \ldots$
$\wedge \wedge q_1 1 \ldots$	$\wedge \wedge q_2 \ldots$
$\wedge \wedge \wedge q_2 \ldots$	стоп
стоп	

Таким образом машина (+) удаляет первые две единицы из исходного слова*).

Из определения 1.5.1 машины Тьюринга следует, что вся интересующая нас информация о ее структуре уже содержится в структуре входящего в ее состав конечного автомата

$$A = (X,\ (X \times M) \cup (\text{стоп}),\ Q,\ \lambda,\ \delta).$$

Поскольку X и Q — конечные множества, мы можем легко перенумеровать их элементы

$$X = \{x_1,\ x_2,\ \ldots,\ x_m\}, \qquad Q = \{q_1,\ q_2,\ \ldots,\ q_n\}.$$

Теперь автомат A можно представить следующим списком пятерок

$$q_k x_l x_n m q_p, \tag{1.5.1}$$

где для каждой пары (q_k, x_l) такой, что $\delta(q_k, x_l) \neq$ $\neq$стоп, имеется в точности одна пятерка такая, что

$$(x_n,\ m) = \delta(q_k,\ x_l), \quad q_p = \lambda(q_k,\ x_l).$$

Ясно, что имеется взаимно однозначное соответствие между такими списками и конечными автоматами

---

*) Работа машины (+) протекает так. Начальные условия: в поле зрения машины устанавливается самая левая единица, и машина находится в состоянии q_0. В первом такте устройство D стирает обозреваемую единицу и сдвигается вправо, а машина переходит в состояние q_1. Во втором такте, если обозревается единица, то D стирает ее и перемещается вправо, а машина переходит в стоп-состояние q_2; если же обозревается пустой символ, то D оставляет его и перемещается вправо, а машина остается в состоянии q_1. Начинается третий такт работы, после которого машина останавливается. (*Прим. перев.*)

рассматриваемого типа. Так как множество значений каждой из величин k, l, n и p нумеруемо*), а величина m принимает три значения, то совокупность всех пятерок вида (1.5.1) может быть также перенумерована. Таким образом, перечислимо множество всех конечных списков таких пятерок и, следовательно, множество всех интересующих нас конечных автоматов.

Далее, эти автоматы можно перечислить эффективным способом**), т. е. так, что по любому заданному числу r можно восстановить r-й автомат. Определим *вес* пятерки вида (1.5.1) как натуральное число $k+l+n+m+p$ (полагая $Л=1$, $П=2$, $H=3$). Ясно, что имеется только конечное число автоматов, задаваемых списками из не более чем r пятерок таких, что вес каждой пятерки не превосходит r. Все такие списки можно расположить в лексикографическом порядке. Автоматы, задаваемые такими списками, назовем *автоматами степени r*. Таким образом, все интересующие нас автоматы можно эффективно перечислить (возможно с повторениями, что не имеет значения) следующим образом: перечисление начинается с автоматов степени 1, затем автоматов степени 2 и т. д., а автоматы одной степени перечисляются в лексикографическом порядке. Таким образом, имеет место следующая теорема.

Теорема 1.5.1. *Множество всех машин Тьюринга эффективно перечислимо*:

$$Z_1,\ Z_2,\ Z_3,\ \ldots$$

Отсюда непосредственно имеем

Следствие 1.5.2. *Множество всех рекурсивно перечислимых множеств эффективно перечислимо:*

$$S_1,\ S_2,\ S_3,\ \ldots,$$

при этом

*) Напомним, что множество называется нумеруемым (перечислимым), если его элементы можно расположить по порядку: первый, второй, третий и т. д.

**) Смысл, который мы вкладываем в термин «эффективный», будет разъяснен в § 1.6.

Определение 1.5.4. Множество S называется *рекурсивно перечислимым*, если $S=\{f(n) \mid n \in N\}$ для некоторой частично рекурсивной функции f.

Наконец, введем еще одно определение, которое потребуется в следующем параграфе.

Определение 1.5.5. Множество S называется *рекурсивным*, если его характеристическая функция *) является рекурсивной функцией.

§ 1.6. Рекурсивные множества и тезис Тьюринга

В этом параграфе будут установлены некоторые результаты из теории рекурсивных функций. В §§ 5.3—5.5 они будут использованы для доказательства теоремы Геделя, которое приводится по следующим двум причинам. Во-первых, потому, что это одна из очень важных теорем, которую, к сожалению, понимают очень мало людей, хотя ее доказательство доступно даже неспециалисту. Во-вторых, четкое понимание этого доказательства позволит обсудить вопрос о потенциальных возможностях мозга и вычислительных машин.

Уточним теперь идею *эффективной процедуры*. Существуют определенного рода вычисления, которые выполняются по механическим правилам или *алгоритмам*, например алгоритм Евклида для нахождения наибольшего общего делителя двух целых чисел. Бесспорно, что любое вычисление, которое можно выполнить на электронной вычислительной машине, осуществляется по чисто механическим правилам. В подобных случаях мы говорим, что имеется эффективная процедура для выполнения таких вычислений. Есть много случаев, когда, с одной стороны, мы уверены в существовании эффективной процедуры для выполнения необходимых вычислений; с другой стороны, мы затрудняемся описать эту процедуру, т. е. составить программу, по которой можно осуществить вычисле-

*) Характеристическая функция $C_S(n)$ множества S равна 0, если n не принадлежит S, и равна 1, если n принадлежит S.

ния. В других же случаях мы не только затрудняемся написать подобную программу, но даже склонны считать, что такой эффективной процедуры в принципе не может существовать (например, для предсказания результата при бросании монеты).

В § 1.5 было показано, что любой процесс, который может быть выполнен на цифровой вычислительной машине, может быть также выполнен и на подходящей машине Тьюринга. Большая общность тех процессов, которые могут быть выполнены на условной машине, описанной в § 1.4, соблазняет ввести следующую гипотезу, которая была впервые высказана Тьюрингом в 1936 г.

Тезис Тьюринга. *Неформальное интуитивное понятие эффективной процедуры над последовательностями символов совпадает с точным понятием процедуры над последовательностями символов, которая может быть выполнена машиной Тьюринга.*

Эта гипотеза никогда не может быть формально доказана просто потому, что для формального доказательства нужны определения всех имеющихся понятий и, таким образом, надо формально определить, что значит «интуитивно» *). Уверенность в справедливости этой гипотезы основана на том, что до последнего времени в теории рекурсивных функций имеет место следующий факт: если удается установить существование некоторой эффективной процедуры, то удается также построить машину Тьюринга, которая бы в точности воспроизводила этот процесс.

Дальнейшие рассуждения этого параграфа будут вполне строги, за исключением следующего: если интуитивно ясно, что данная процедура эффективна, то я буду считать, что существует соответствующая машина Тьюринга. Чтобы успокоить Вас, замечу, что во всех таких случаях имеется строгое доказательство [9]. Давайте перефразируем теперь приведенные выше

*) Но могла бы быть опровергнута, если бы удалось указать такую процедуру, которую все бы признали эффективной (в интуитивном смысле), и доказать, что эта процедура не может быть выполнена никакой машиной Тьюринга. (*Прим. ред.*)

определения на интуитивном языке эффективных процедур.

Определение 1.6.1. *Функция* называется *рекурсивной*, если существует эффективная процедура для ее вычислений.

Определение 1.6.2. *Множество* называется *рекурсивным*, если существует эффективная процедура для выяснения того, принадлежит или не принадлежит произвольный элемент к этому множеству.

Определение 1.6.3. *Множество* называется *рекурсивно перечислимым*, если существует эффективная процедура для последовательного порождения (перечисления) его элементов.

Пример. Множество квадратов целых чисел рекурсивно перечислимо. Для получения элементов этого множества достаточно последовательно брать числа 1, 2, 3, ... и возводить их в квадрат. Более того, это множество является также рекурсивным. Проверка принадлежности любого числа к данному множеству сводится к разложению этого числа на простые множители, что и позволяет выяснить, является ли оно квадратом. Позднее мы увидим, что любое рекурсивное множество рекурсивно перечислимо и что обратное утверждение является неверным. Более глубоким из этих двух результатов является последний.

Итак, мы имеем два вида определений: формальные и интуитивные. Если функция рекурсивна в формальном смысле, то совершенно ясно, что она рекурсивна и в интуитивном смысле. Действительно, из формального определения следует существование машины Тьюринга, вычисляющей эту функцию. А это как раз и означает существование эффективной процедуры. Аналогично, множество, которое рекурсивно (или рекурсивно перечислимо) в формальном смысле, должно быть рекурсивно (или рекурсивно перечислимо) в интуитивном смысле. Обратное же утверждение, т. е. утверждение о том, что если множество рекурсивно в интуитивном смысле (т. е. в смысле определения 1.6.2), то оно рекурсивно и в формальном смысле (т. е. в смысле определения 1.5.5), как раз и дает нам тезис Тьюринга. Если мне потребуется убедить Вас в том, что существует программа для выполнения определенных вычислений (сколь угодно длинная и работающая достаточно долго), то я просто представлю Вам блок-схему вычислений, а не программу на языке машины. Приводимые нами доказательства связаны со строгими доказательствами так же, как блок-схемы программ связаны с программами на языке машин. Ниже мы останавливаемся только на существенных вопросах, опуская детали, подробное рассмотрение которых могло бы затемнить основное содержание. Теперь докажем теоремы о рекурсивных и рекурсивно перечислимых множествах.

Теорема 1.6.1. *Если R и S — рекурсивно перечислимые множества, то рекурсивно перечислимы множества**) $R \cup S$ и $R \cap S$.

Доказательство. Если P_R и P_S — эффективные процедуры (алгоритмы) для получения множеств R и S, то алгоритм для получения множества $R \cup S$ получается из исходных путем их одновременного применения.

Пусть алгоритм P_R последовательно порождает элементы $m_1, m_2, m_3, \ldots$ множества R, а P_S последовательно порождает элементы $n_1, n_2, n_3, \ldots$ множества S. Тогда алгоритм для перечисления элементов множества $R \cap S$ заключается в следующем.

Поочередно используем алгоритмы P_R и P_S для перечисления элементов m_1, n_1, m_2, n_2, m_3, n_3 и т. д. Каждый вновь порожденный элемент m_i сравнивается со всеми ранее порожденными элементами n_j. Если m_i совпадает с одним из них, то он включается во множество $R \cap S$. В противном случае надо переходить к порождению элемента n_i и применить к нему аналогичную проверку. Описанная процедура позволяет эффективно перечислить все элементы множества $R \cap S$, что и требовалось доказать.

Теорема 1.6.2. *Множество положительных целых чисел S рекурсивно тогда и только тогда, когда S и $\bar{S}$ рекурсивно перечислимы**).

---

*) См. Приложение 1.

Доказательство. Пусть S — рекурсивное множество. Тогда существует алгоритм, который для каждого натурального числа 0, 1, 2, ... проверяет, принадлежит оно множеству S или не принадлежит. Этот алгоритм позволяет перечислить все элементы множества S. Аналогичным образом перечисляется множество $\bar{S}$.

Наоборот, пусть S и $\bar{S}$ — рекурсивно перечислимые множества; $n_1, n_2, n_3, \ldots$ — рекурсивное перечисление множества S и $m_1, m_2, m_3, \ldots$ — рекурсивное перечисление множества $\bar{S}$. Тогда для выяснения принадлежности произвольного числа n к множеству S достаточно породить элементы $n_1, m_1, n_2, m_2, n_3, m_3, \ldots$ и каждый вновь получаемый элемент сравнить с n. Так как n должно принадлежать либо S, либо $\bar{S}$, то через конечное число шагов получим либо n_i, либо m_i, совпадающее с n. Следовательно, этот метод и является алгоритмом для определения принадлежности произвольного натурального числа к множеству S. Поэтому S рекурсивно, что и требовалось доказать.

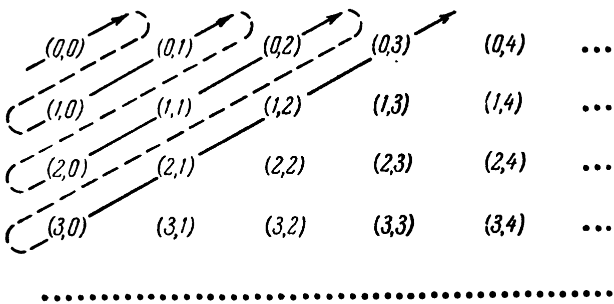


Рис. 1.8. Диагональное перечисление упорядоченных пар.

Прежде чем переходить к следующей теореме, напомним читателю, что существует эффективное перечисление всех упорядоченных пар натуральных чисел, которое называется *диагональным методом*. Оно показано на рис. 1.8 и заключается в следующем. Перечисление осуществляется последовательным прохождением по диагонали, начиная с верхнего левого угла. Первыми парами этого перечисления являются

$$(0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2), (3, 0), \ldots$$

Читатель может проверить, что в этой последовательности пара (x, y) находится на $1/2(x^2+2xy+ +y^2+3y+x+2)$-м месте.

Теорема 1.6.3. *Существует рекурсивно перечислимое, но не рекурсивное множество положительных целых чисел.*

Доказательство. Из теоремы 1.6.2 следует, что этот факт эквивалентен существованию рекурсивно перечислимого множества натуральных чисел, дополнение к которому *не* является рекурсивно перечислимым.

Пусть $S_1, S_2, \ldots$ — эффективное перечисление всех рекурсивно перечислимых множеств (см. следствие 1.5.2), а $\tau(x, y)$ — номер пары (x, y) какого-нибудь эффективного перечисления всех пар натуральных чисел, например диагонального метода. Тогда на $\tau(x, y)$-м этапе вычисляется x-й элемент множества S_y. Если этот элемент совпадает с y, то включим его во множество U. Таким образом, y принадлежит множеству U тогда и только тогда, когда y принадлежит S_y. Ясно, что U является рекурсивно перечислимым множеством. Так как множество *) $\bar{U}$ состоит из всех таких y, что y не принадлежит S_y, то $\bar{U}$ отличается от любого рекурсивно перечислимого множества хотя бы одним целым числом **). Поэтому $\bar{U}$ не является рекурсивно перечислимым, а U — рекурсивным множеством, что и требовалось доказать.

---

*) $\bar{U}$ — множество тех натуральных чисел, которые не входят в U. См. приложение 1. (*Прим. ред.*)

**) В самом деле, если бы $\bar{U}$ было рекурсивно перечислимым, оно встречалось бы в нашем пересчете с каким-то номером, скажем $\bar{U}=S_j$. Число j либо принадлежит S_j, либо нет. Если $j \in S_j$, то $S_j \neq \bar{U}$, так как тогда $j \in U$. Если $j \notin S_j$, то S_j также не равно $\bar{U}$, так как тогда $j \in \bar{U}$ и $j \notin S_j$. Значит, $\bar{U}$ не может быть рекурсивно перечислимым. (*Прим. ред.*)

Эта теорема фактически включает в себя в неявном виде теорему Гёделя. Однако в §§ 5.3—5.5 предстоит еще провести кропотливую работу для такого понимания идеи Гёделя, чтобы можно было извлечь его теорему из очень общего утверждения теоремы 1.6.3.

Рассмотрение машин Тьюринга нельзя закончить без упоминания об универсальных машинах Тьюринга, т. е. таких машин Тьюринга, которые способны моделировать работу *любой* машины Тьюринга. Для уточнения этого понятия обратимся к теореме 1.5.1 и каждой машине Тьюринга Z сопоставим число n, где Z_n — первое вхождение Z в эффективном перечислении Z_1, Z_2, Z_3, ... машин Тьюринга. Теперь можно точно установить существование универсальной машины Тьюринга (ср. с определением 1.5.2).

Теорема 1.6.4. *Существует универсальная машина Тьюринга U такая, что*

$$F_{Z_n}(x) = F_U(n, x)$$

для всех машин Тьюринга Z_n и всех целых чисел x.

Доказательство. Пусть n и x — произвольные натуральные числа. По числу n можно эффективно построить машину Тьюринга Z_n и на ней эффективно вычислить функцию $F_{Z_n}(x)$. Таким образом, функция F_U эффективно вычислима, а тогда, согласно тезису Тьюринга, существует искомая машина Тьюринга U, что и требовалось доказать.

Читателя, который посчитает приведенное выше использование тезиса Тьюринга мошенничеством, мы отсылаем к работе Тьюринга [8], в которой подробно описывается конструкция универсальной машины Тьюринга (отметим только, что там имеется несколько мелких ошибок).

§ 1.7. Регулярные и представимые события

«Жизнь» конечного автомата протекает в дискретные моменты $t=1, 2, 3, \ldots$, и его «общение» с внешним миром осуществляется только посредством входов, которые он воспринимает в эти моменты. Следовательно, если в некоторый момент спросить «его», наступило ли некоторое событие (предварительно договорившись интерпретировать некоторые выходные символы как «да», а другие — как «нет»), то «его ответ» может зависеть только от входного слова *). Таким образом, в «жизнь» автомата может проникнуть только такое событие, которое состоит из некоторого множества входных слов, где

Определение 1.7.1. *Событием* называется любое подмножество множества всех входных слов в некотором конечном алфавите.

Если a_1, a_2, a_3, a_4 — входы автомата, то будем считать, что $a_1a_2a_3a_4$ является входным словом, в котором a_1 предшествует a_2 и т. д.

Рассмотрим теперь следующую модель распознавания события. Будем говорить, что некоторое событие E «распознаваемо» (или «представимо») в конечном автомате A, если A отвечает «да» на всякое входное слово из события E. Более формально

Определение 1.7.2. Событие E называется *представимым*, если существуют конечный автомат

$$A=(X, Y, Q, \lambda, \delta)$$

с начальным состоянием q_0 и разбиение Y на два различных подмножества «да» и «нет» такие, что входное слово $a_1a_2 \ldots a_n$ принадлежит E тогда и только тогда, когда A, находясь в начальном состоянии q_0, последовательно воспринимает входы $a_1, a_2, \ldots, a_n$ и на последнем такте на выход выдает «да».

Пусть входом A является 1. Тогда событие $E=\{1 \ldots 1 (n \text{ раз}) | n \text{ является квадратом}\}$ непредставимо. Это как раз и означает, что существуют непредставимые события. Доказательство и обсуждение этого факта будет проведено в конце этого параграфа. Сейчас же будет дана точная характеристика представимых событий. Без большого ущерба для понимания этой характеристики читатель может не вникать в доказательства лемм 1.7.1 и 1.7.2.

*) И начального состояния автомата. (*Прим. ред.*)

Бесспорно, пустое событие и события, каждое из которых состоит из единственного однобуквенного входного слова, представимы.

Введем три операции над событиями: дизъюнкцию, умножение и итерацию.

Определение 1.7.3. Если E и F — два события (множества слов), то определим:

$E \cup F$ (*дизъюнкция* E и F): $x \in E \cup F$ тогда и только тогда, когда $x \in E$ или $x \in F$.

$E \cdot F$ (*произведение* E и F): $x \in E \cdot F$ тогда и только тогда, когда слово x может быть записано в виде ef, т. е. за словом e из E следует слово f из F.

E^* (*итерация* E): $x \in E^*$ тогда и только тогда, когда слово x может быть записано в виде $e_1 \cdot \ldots \cdot e_n$ при некотором n, а слова $e_1, \ldots, e_n$ принадлежат E, или x есть пустое слово λ.

Если положить $E^0 = \lambda$, $E^1 = E$, $E^{n+1} = E^n \cdot E$, то легко видеть, что $E = E^0 \cup E^1 \cup E^2 \cup E^3 \cup E^4 \cup \ldots$

Лемма 1.7.1. *Если E и F — представимые события, то такими же являются события $E \cup F$, $E \cdot F$ и E^*.*

Доказательство. Пусть автомат $A(E) = (X, Y, Q, \lambda, \delta)$ с начальным состоянием q_0 представляет событие E, а автомат $A(F) = (X', Y', Q', \lambda', \delta')$ с начальным состоянием q_0' представляет событие F. Без потери общности можно считать, что $X = X'$, $Y = Y' = \{\textit{да, нет}\}$.

Утверждение леммы получается из следующих построений.

а) $$A(E \cup F) = (X, Y, Q \times Q', \lambda \times \lambda', \delta'')$$

с начальным состоянием (q_0, q_0');

$$(\lambda \times \lambda')((q, q'), x) = (\lambda(q, x), \lambda'(q', x))$$

и

$$\delta''((q, q'), x) = \textit{да}$$

тогда и только тогда, когда

$$\delta(q, x) = \textit{да},$$

либо

$$\delta'(q', x) = \textit{да}.$$

б) Построение автомата, представляющего событие $E \cdot F$, является более сложным. Через $2^{Q'}$ обозначим множество всех подмножеств множества Q'. Тогда

$$A(E \cdot F) = (X,\ Y,\ Q \times 2^{Q'},\ \bar{\lambda},\ \bar{\delta})$$

с начальным состоянием *) $(q_0, \varnothing)$;

$$\bar{\lambda}((q,\ T),\ x) = (\lambda(q,\ x),\ \{\lambda'(t,\ x) \mid t \in T\} \cup R),$$

где

$$R = \begin{cases} \varnothing, & \text{если} \quad \delta(q,\ x) = \textit{нет}, \\ \{q_0'\}, & \text{если} \quad \delta(q,\ x) = \textit{да}; \end{cases}$$

$\bar{\delta}((q,\ T),\ x) = \textit{да}$ тогда и только тогда, когда $\delta'(t,\ x) = \textit{да}$ при некотором $t \in T$.

с) $$A(F^*) = (X,\ Y,\ 2^{Q'},\ \bar{\lambda},\ \bar{\delta})$$

с начальным состоянием $\{q_0'\}$;

$$\bar{\lambda}(T,\ x) = \{\lambda'(t,\ x) \mid t \in T\} \cup R,$$

где $R = \{q_0'\}$ тогда и только тогда, когда $\delta'(t,\ x) = \textit{да}$ при некотором $t \in T$; в противном случае $R = \varnothing$. $\delta(T,\ x) = \textit{да}$ тогда и только тогда, когда $\delta'(t,\ x) = \textit{да}$ при некотором $t \in T$. Что и требовалось доказать.

Таким образом, показано, что все регулярные события представимы, где

Определение 1.7.4. Множество слов E называется *регулярным*, если: а) E пусто или содержит в точности один элемент, или б) E может быть получено из пустого множества и одноэлементных множеств с помощью конечного числа дизъюнкций, умножений и итераций.

Покажем теперь, что все представимые события регулярны, т. е. регулярные события являются точной характеристикой представимых событий. Для доказательства этого факта требуется следующая лемма.

---

*) $\varnothing$ есть пустое множество. Поэтому $\varnothing$ является подмножеством множества Q', так что $\varnothing \in 2^{Q'}$.

Лемма 1.7.2. *Пусть T — конечное множество и R — бинарное отношение на T (т. е. для каждой пары a и b либо «aRb» (a и b находятся в отношении R), либо «не aRb»). Слово $a_1a_2 \ldots a_m$ из элементов T называется R-переходным словом тогда и только тогда, когда a_iRa_{i+1} для $i=1, \ldots, m-1$. Тогда для любых двух элементов a и b из T множество всех R-переходных слов, начинающихся с a и оканчивающихся b, регулярно.*

Доказательство проведем индукцией по n — числу символов в T.

а) $$n=1, \quad T=\{a\}.$$

Пусть aRa. Тогда искомым множеством слов является регулярное множество $\{a, aa, aaa, \ldots\}=\{a\}^*$. Если же на aRa, то искомое множество состоит из одного элемента $\{a\}$ и, следовательно, регулярно.

б) $n>1$. Предположим, что теорема доказана для всех $m<n$. Тогда любое R-переходное слово, начинающееся с a и оканчивающееся b, может быть записано в виде

$$aA_1aA_2a \ldots aA_maBb, \tag{1.7.1}$$

причем в словах A_i и B нет вхождения буквы a. Пусть

$$C=\{x \mid aRx\}, \quad D=\{x \mid xRa\}, \quad E=\{x \mid xRb\}.$$

Тогда каждое A_i является переходным словом, начинающимся с элемента из C и оканчивающимся элементом из D, а B является переходным словом, начинающимся с элемента из C и оканчивающимся элементом из E. Пусть R_{xy} — множество R-переходных слов, начинающихся x и оканчивающихся y. По предположению индукции оно регулярно (ибо алфавит $T\setminus\{a\}$ состоит из $n-1$ элементов), когда $x\in C$, $y\in D$, или когда $x\in C$, $y\in E$.

Так как операция дизъюнкции сохраняет свойство регулярности множества, то

$$\bigcup\{R_{xy} \mid x\in C,\ y\in D\}=F$$

и

$$\bigcup\{R_{xy} \mid x\in C,\ y\in E\}=G$$

регулярны. Используя условие (1.7.1), получаем

$$R_{ab} = (\{a\} \cdot F)^* \cdot (\{a\} \cdot G \cdot \{b\}),$$

и поэтому R_{ab} регулярно, что и требовалось доказать.

Теперь в качестве множества T возьмем множество $Q \times X$ пар «состояние — вход» автомата A из определения 1.7.2, представляющего событие E. Бинарное отношение определим так: $(c, d) R (e, f)$ тогда и только тогда, когда $e = \lambda(c, d)$, т. е. (c, d) переводит автомат A в состояние e. Входное слово $a_0 \ldots a_m$ принадлежит E тогда и только тогда, когда (q_0, a_0), (q_1, a_1), ..., (q_m, a_m) является R-переходным словом таким, что $\delta(q_m, a_m) = да$. Следовательно, множество

$$E' = \bigcup \{R_{(q_0, a_0)(q_m, a_m)} \mid a_0 \in X, \ \delta(q_m, a_m) \in да\}$$

регулярно. Поскольку E получается из E' путем извлечения только второго элемента пары, а эта операция не нарушает регулярности (см. определения 1.7.3 и 1.7.4), то нами доказана следующая теорема.

Теорема 1.7.1. *Событие представимо тогда и только тогда, когда оно регулярно.*

Значение этой теоремы состоит в том, что она позволяет математически выразить возможности конечных детерминированных автоматов. Язык регулярных событий становится основным орудием инженеров, проектирующих такие автоматы как сети, построенные из электромеханических реле или других подобных элементов.

Покажем теперь, что событие

$$S = \{1 \ldots 1 \,(n \text{ раз}) \mid n \text{ является квадратом}\}, \qquad (1.7.2)$$

рассмотренное в начале этого параграфа в качестве примера, непредставимо *ни в одном* конечном автомате A. Для этого достаточно убедиться лишь в том, что это событие не является регулярным событием. Пусть P_k означает, что существует k последовательных чисел $m = n, \ldots, n+k-1$ таких, что $1 \ldots 1$ (m раз) не принадлежит E, а $1 \ldots 1$ ($n+k$ раз) принадлежит E. Тогда событие E из (1.7.2) обладает свойством P_k при всех k. Используя определение 1.7.4, читатель может убедиться (например, индукцией) в том, что не существует регулярного события, обладающего свойством P_k при всех k. Следовательно, «множество квадратов» *нерегулярно*.

Однако известно (см. пример в § 1.6), что множество всех квадратов *рекурсивно*. Это означает, что машины Тьюринга «сильнее» конечных автоматов. С другой стороны, известно, что если конечному автомату добавить рецепторы и эффекторы, с помощью которых он мог бы в некоторой степени *воздействовать* на окружающий мир (например, ленту с лентопротяжным, считывающим и записывающим устройством, за счет чего автомат может работать как машина Тьюринга), можно радикально изменить его возможности поведения. Более детально такие варианты поведения будут рассмотрены в § 4.1 при обсуждении понятия «обратной связи».

§ 1.8. Еще некоторые сведения из теории конечных автоматов

Этот параграф посвящен одному из новых математических направлений, изучающему автоматы, абстрактно задаваемые в виде пятерок

$$A = (X,\ Y,\ Q,\ \lambda,\ \delta).$$

Обозначим через X^* множество всех конечных последовательностей (слов) входных букв и включим в него Λ — «пустое слово», не содержащее ни одной буквы. X^* — это полугруппа *) относительно операции приписывания (сцепления) слов; сцепление ассоциативно $(x_1x_2)x_3 = x_1(x_2x_3)$, а Λ является единицей полугруппы. Область определения λ и δ расширим так, что λ отображает $Q \times X^*$ в Q, а δ отображает его в Y. Это осуществляется повторным применением следующих равенств (рис. 1.9):

$$\lambda(q, x'x'') = \lambda(\lambda(q, x'), x''),\quad \delta(q, x'x'') = \delta(\lambda(q, x'), x'').$$

С каждым состоянием q автомата A можно связать вполне определенный способ получения выхода

*) Определение полугруппы и других используемых понятий алгебры см. в Приложении 2. (*Прим. ред.*)

для каждого входного слова. Он задается функцией $A_q: X^* \to Y$, где $A_q(x) = \delta(q, x)$. Очевидно, что поведение A будет одинаковым, если он начинает работу в двух состояниях q и q' с одной и той же функцией «вход — выход», т. е. если $A_q(x) = A_{q'}(x)$ для всех входных слов x. Таким образом, если нас интересует только внешнее поведение A, то мы можем заменить

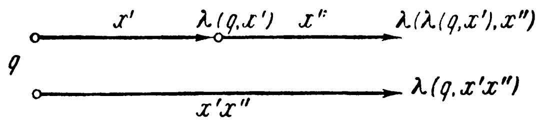


Рис. 1.9.

его на автомат в *приведенной форме*, который имеет одно состояние для каждой функции A_q. Пусть автомат A в состоянии q имеет функцию «вход — выход» f, и пусть q' — состояние, *достижимое* из q, т. е. можно найти $x \in X^*$ такое, что $\lambda(q, x) = q'$. Тогда q' имеет функцию $g = A_{q'}$, где

$$A_{q'}(x') = A_{\lambda(q,x)}(x') = \delta(\lambda(q, x), x') = \\ = \delta(q, xx') = f(xx') = fL_x(x'),$$

а $L_x: X^* \to X^*$ есть функция «умножение слева на x»: $L_x(x') = xx'$. Теперь в приведенной форме автомата A каждое состояние, достижимое из q, заменяем функцией «вход — выход» fL_x и считаем, что приведенная форма (ограниченная состояниями, достижимыми из q) имеет конечное число состояний; во всяком случае имеется конечное число различных функций fL_x (так что бесконечное множество слов должно давать одну и ту же fL_x). Таким образом, наше внимание будет обращено на автоматы вида

$$A(f) = (X, Y, Q_f, \lambda_f, \delta_f),$$

где f — заданная функция, отображающая X^* в Y, а

$$Q_f = \{g: X^* \to Y \mid g = fL_x \text{ для некоторого } x \in X^*\},$$
$$\lambda_f(g, x) = gL_x, \quad \delta_f(g, x) = g(x),$$

и особенно когда f такова, что Q_f — конечное множество.

С автоматом $A(f)$ свяжем полугруппу S_f, а именно совокупность отображений множества состояний Q_f, индуцированную входными словами в $A(f)$, т. е.

$$S_f = \{s\colon Q_f \to Q_f \mid \exists x \in X^* \text{ такое, что } s(q) = \\ = \lambda(q, x) \text{ для всех } q \in Q\}.$$

Эта полугруппа конечна тогда и только тогда, когда Q_f конечно. Если A не находится в приведенной форме, мы связываем с ним (и некоторым заданным начальным состоянием q) полугруппу приведенного автомата $A(f)$, где f является как раз функцией A_q.

Мы можем ввести функцию $i_f: S_f \to Y$, определив

$$i_f(s) = f(x), \quad \text{если} \quad s(q) \equiv \lambda(q, x).$$

Это не зависит от выбора X. Затем мы можем определить новый «полугрупповой автомат» с выходной функцией i_f

$$A(s_f, i_f) = (s_f, s_f, Y, \lambda, \delta),$$

где $\lambda(s, s') = s \cdot s'$ (символ означает умножение в полугруппе) и $\delta(s, s') = i_f(s, s')$.

1. Регулярные события

Мы видели (теорема 1.7.1), что множество представимо тогда и только тогда, когда оно регулярно. Например, регулярное множество

$$10^* \cup (0 \cup 10^*1) \cdot (1 \cup 0^*1)^* 0^*0$$

может быть получено как множество слов, для которых $\delta(q_1, x) = 1$ в автомате, показанном на рис. 1.10 (стрелка, помеченная парой a/b и идущая из кружка q к кружку q', означает, что $\lambda(q, a) = q'$, а $\delta(q, a) = b$).

Введем теперь понятие регулярного выражения более абстрактно. Отметим, что регулярное выражение является выражением *метаязыка* и служит для *обозначения* регулярного события.

Пусть $X = \{x_1, \ldots, x_n\}$. Ввведем новое множество символов $\overline{x}_1, \ldots, \overline{x}_n$, $\overline{\Lambda}$ и $\overline{\varnothing}$. Скажем, что слово, состоящее из букв алфавита $\{\overline{x}_1, \ldots, \overline{x}_n, \overline{\Lambda}, \overline{\varnothing}, +, 0, *,), (\}$, является *регулярным выражением*, если оно может быть получено из выражений $\overline{x}_1, \ldots, \overline{x}_n, \overline{\Lambda}, \overline{\varnothing}$ посредством конечного числа операций $\alpha, \beta \to (\alpha+\beta)$;

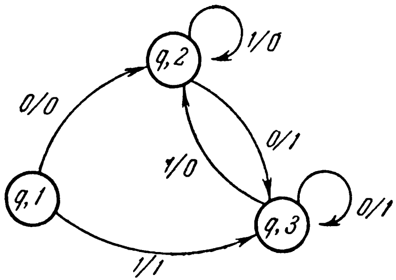


Рис. 1.10.

$\alpha, \beta \to (\alpha \circ \beta)$; $\alpha \to \alpha^*$. Пусть L — множество регулярных выражений, и пусть K — множество всех подмножеств X^*. Определим функцию $|\ |: L \to K$ по индукции:

$$|\overline{\varnothing}| = \varnothing; \quad |\overline{\Lambda}| = \Lambda; \quad |\overline{x}_i| = x_i, \qquad i = 1, \ldots, n;$$

$$|(\alpha+\beta)| = |\alpha| \cup |\beta|; \qquad |(\alpha \circ \beta)| = |\alpha| \cdot |\beta|; \qquad |\alpha^*| = |\alpha|^*.$$

Выражение $|\alpha| = E$ читается так: α обозначает множество E. Заметим, что множество регулярно, если оно может быть задано регулярным выражением. Мы часто пишем $\alpha\beta$ вместо $\alpha \circ \beta$ и опускаем скобки там, где это не приводит к недоразумению.

Для регулярных выражений α и β мы пишем $\alpha \equiv \beta$ (читается α и β тождественны), если α и β содержат одинаковые символы в одинаковом порядке; мы пишем $\alpha = \beta$ (читается α и β равны), если α и β обозначают одно и то же множество, т. е. $|\alpha| = |\beta|$. Например, $\overline{0}^*\, \overline{0} + \overline{0} + \overline{1} = \overline{1} + \overline{0}^*$. Легко проверить, что для любых регулярных выражений α, β и γ выполняются следующие равенства:

A_1: $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$; A_2: $\alpha(\beta\gamma) = (\alpha\beta)\gamma$;

A_3: $\alpha + \beta = \beta + \alpha$; A_4: $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$;

A_5: $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$; A_6: $\alpha + \alpha = \alpha$;

A_7: $\overline{\Lambda}\alpha = \alpha$; A_8: $\overline{\varnothing}\alpha = \overline{\varnothing}$; A_9: $\alpha + \overline{\varnothing} = \alpha$;

A_{10}: $\alpha^* = \overline{\Lambda} + \alpha^*\alpha$; A_{11}: $\alpha^* = (\overline{\Lambda} + \alpha)^*$.

Саломаа показал *), что *все* равенства для регулярных выражений могут быть получены из равенств A_1—A_{11} повторным применением двух правил вывода **):

$R1$ (замена). Допустим, что γ' есть результат замены вхождения α на β в γ. Тогда из равенства $\alpha = \beta$ и $\gamma = \delta$ можно вывести равенства $\gamma' = \delta$ и $\gamma' = \gamma$.

$R2$ (решение равенства). Пусть $\Lambda \notin |\beta|$. Тогда из равенства $\alpha = \alpha\beta + \gamma$ можно вывести равенство $\alpha = \gamma\beta^*$.

Однако Редько показал ***), что если использовать только правило замены, то ни одна конечная система аксиом не является достаточной для выведения всех регулярных выражений. Если равенства A_1—A_{11} пополнить счетным множеством аксиом

$$\alpha^* = \Lambda + \alpha + \alpha^2 + \ldots + \alpha^{k-1} + \alpha^k\alpha^*, \quad k = 1, 2, 3, \ldots$$

(где $\alpha^2 = \alpha \cdot \alpha$, $\alpha^3 = \alpha \cdot \alpha \cdot \alpha$ и т. д.), то из этой системы уже можно вывести все равенства регулярных выражений, применяя только $R1$.

Отметим, что если α — регулярное выражение, содержащее n букв (считаем, что $\bar{0}\bar{1}^* + \bar{0}^*\bar{1}\bar{0}^*$ имеет пять букв, а не две ****)), то существует автомат не более чем с 2^n состояниями, который представляет $|\alpha|$.

*) A. Salomaa, Two complete axiom systems for the algebra of regular events, J. ACM, **13** (1966), 158—159.

**) Ср. с § 5.3.

***) Редько В. Н., Об определяющей совокупности соотношений алгебры регулярных событий. Укр. матем. журнал **16** (1964), 120—126 (*Прим. ред.*)

****) То есть считается каждое вхождение буквы. (*Прим. ред.*)

2. *Моделирование и делимость*

Будем говорить, что автомат A *моделирует* автомат A', если при некотором кодировании и декодировании входа и выхода A перерабатывает слова точно так же, как это делает A'. Мы требуем, чтобы кодирующее и декодирующее устройства не имели памяти (т. е. заменяли символ на символ) для того, чтобы A производил всю вычислительную работу, включая и память (см. рис. 1.11).

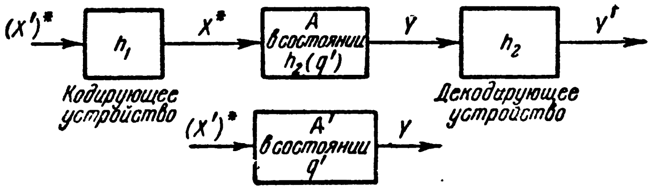


Рис. 1.11. A моделирует A', если можно выбрать h_1, h_2 и h_3 так, что обе системы имеют одинаковое поведение «вход — выход».

Если A моделирует A', то мы пишем $A'|A$ и говорим, что A' *делит* A. Если $A|A'$ и $A'|A$, то мы говорим, что A и A' *слабо эквивалентны*. Полезность полугрупп в теории конечных автоматов выражается, например, в том, что с их помощью легко получить результат: $A(s_f, i_f)$ слабо эквивалентен $A(f)$. Это получается потому, что полугруппы, как и автоматы, имеют естественное понятие делимости. Мы говорим, что полугруппа S *делит* полугруппу S', если в S' существует подполугруппа S'', гомоморфным образом которой является S (при некотором гомоморфизме Z):

$$S' \supseteq S'' \xrightarrow{Z} S.$$

Если S и S' имеют отображения $i : S \to Y$, $i' : S' \to Y'$, то мы говорим, что (S, i) *делит* (S', i'), если верно все вышесказанное и, кроме того, существует отображение $H : Y' \to Y$ такое, что

$$i(Z(s)) = H(i'(s)) \quad \text{для всех} \quad s \in S''.$$

Связь между понятием делимости для автоматов и полугрупп выражается в том, что автомат $A(g)$ делит автомат $A(f)$ тогда и только тогда, когда пара (S_g, i_g) делит пару (S_f, i_f).

3. Теория разбиений без циклов

Теорема 1.3.1 гласит, что каждый конечный автомат можно промоделировать сетью из модулей (т. е. конечных автоматов с *одним состоянием*), если в сети допускаются циклы. Подобные конструкции являются центральными в любом учебнике по теории релейных схем. Действительно, при построении любой такой сети мы можем использовать копии только одного модуля (модуль *) «штрих Шеффера»). Другими словами, достаточно очень простого набора компонент для построения любого конечного автомата, *если только разрешаются циклы.* В этом пункте мы приведем несколько теорем об ограничениях, вытекающих из запрета циклов между автоматами (конечно, циклы могут быть внутри автоматов, которые мы соединяем).

Пусть даны автоматы A, A' и отображение $Z: X' \times Y \to X$. Полупрямое произведение $A' \times_Z A$ автоматов A' и A относительно отображения Z определим

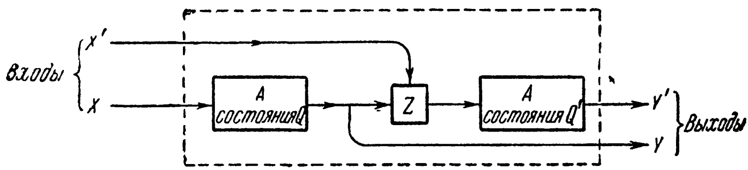


Рис. 1.12.

посредством диаграммы, показанной на рис. 1.12. Чтобы получить *последовательное* соединение (хотя бы сохраняющее выход A), мы делаем Z независимым от X'; чтобы получить *параллельное* соединение,

*) $(a|b)=1$ только в том случае, когда $a=0$ и $b=0$. См. § 3.2.

мы берем $Z(x', y)$ таким, чтобы оно совпадало с x'. Наше точное понятие соединения без циклов состоит в том, что, повторяя образования полупрямых произведений и кодирования без памяти, мы получаем автоматы, моделируемые такими соединениями. Пусть $SD(\mathfrak{A})$ означает множество автоматов, полученных таким способом из множества автоматов $\mathfrak{A}$.

Будем говорить, что автомат A *несократим*, если для *любой* системы автоматов $\mathfrak{A}$ такой, что $A \in SD(\mathfrak{A})$, существует автомат A', который моделирует A, т. е. A не может быть заменен комбинацией более простых автоматов.

Автомат A назовем *тождественно возвратным*, если каждый вход действует либо как сигнал «сброса», переводя автомат в состояние, полностью определяемое этим входом ($\lambda(q, x) = q_x$ для всех $q \in Q$), либо оставляет данное состояние неизменным ($\lambda(q, x) = q$ для всех $q \in Q$).

Особое значение имеет тождественно возвратный автомат с двумя состояниями, который называется триггером T. Триггер имеет состояния (q_0, q_1) и входы (e, x_0, x_1) с $\lambda(q, e) = q$ (e — единица) и $\lambda(q, x_i) = q_i$ (x_i возвращает триггер в состояние q_i); выход T совпадает с состоянием. Отметим, что любой тождественно возвратный автомат можно построить путем подходящего соединения без циклов нескольких триггеров. Триггер имеет полугруппу $V_3 = \{1, r_0, r_1\}$, элементы которой являются отображениями состояний, соответствующих входам e, x_0, x_1. Пусть U_n означает множество полугрупп, которые делят V_3. На самом деле,

$$U_n = \{v_0, v_1, v_2, v_3\},$$

где $v_0 = \{1\}$, $v_1 = \{1, r_0\}$ и $v_2 = \{r_0, r_1\}$.

Следующие изящные результаты были установлены Кроном и Родесом *).

1. Автомат $A(f)$ неприводим тогда и только тогда, когда s_f является либо элементом U_n, либо простой

*) K. B. Krohn and J. L. Rhodes, Algebraic theory of machines, Proc. Symp. Math. Theory Automata, Brooklyn, N. J. Polytechnic Press, 1963. (*Прим. ред.*)

группой (т. е. не имеет нетривиальных нормальных делителей).

2. Данный автомат $A(f)$ можно построить в виде соединения без циклов из триггеров и автоматов (не обязательно всех), простые группы которых делят полугруппу S_f.

Полное изложение и доказательство этих результатов читатель может найти в [13]. Другой подход к соединениям без циклов с использованием теории структур можно найти в [14].

4. *Справки*

Существует много и других тем в теории конечных автоматов, даже если мы ограничимся приведенным выше абстрактным подходом. Вместо того чтобы кратко осветить эти темы, мы приводим список работ, которые позволят познакомиться с этими исследованиями. Отличная библиография работ, опубликованных до 1963 г., содержится в книге: М. А. Айзерман и др., Логика. Автоматы. Алгоритмы, Физматгиз, М., 1963.

§ 1.9. Самовоспроизводящиеся автоматы *)

Следует отметить, что, обычно, в теории автоматов мы имеем дело с устройствами, осуществляющими переработку входной информации в выходную, не интересуясь изменениями, происходящими внутри автомата. При изучении самовоспроизводящихся автоматов внимание переносится на то, каким образом исходная информация обусловливает рост и изменения

*) Этот параграф обобщает итоги работы автора по самовоспроизводящимся автоматам, опубликованные в двух статьях:

A. «A simple self-reproducting universal automation», Information and Control, 9, № 2 (1966), 177—189.

B. «Some comments on self-reproducting automata», presented at the Conference on Computer Science and systems, London, Ontario, September 1965. (*Прим. автора.*)

Этот параграф был напечатан в журнале «Автоматика», № 2 (1966): М. А. Арбіб, Універсальні конструюючі машини і автомати, що самовідтворюються. (*Прим. ред.*)

структуры автомата. Будет небезынтересно познакомиться с тем, что может дать подобный подход в решении наших классических проблем. Изучение процессов самовоспроизведения — это лишь одна ветвь обширных исследований, посвященных растущим автоматам.

Пусть Z — некоторая машина Тьюринга, а $e(Z)$ — ее описание (код). Как известно, можно построить универсальную машину Тьюринга, т. е. такую машину, что если ее снабдить описанием $e(Z)$, то она сможет моделировать все вычисления машины Z. Это навело фон Неймана (в 1951 г.) на мысль о существовании такого автомата A, который, будучи снабжен описанием любого другого автомата M (состоящего из определенного набора элементарных частей), был бы способен создать копию M, и он показал, как можно использовать подобный «универсальный конструктор» A для решения проблемы самовоспроизведения.

Сначала укажем две теоремы о неподвижной точке. Можно было бы ожидать, что универсальный конструктор, снабженный перечнем элементов, и синтезирующий новые машины из этих элементов, сам должен быть выполнен из более сложных компонентов. Доказав, что это предположение неверно, получаем первую теорему о неподвижной точке: *существует такой набор компонент, из которого можно построить универсальный конструктор для автоматов, создаваемых из этого же набора элементов.*

Универсальный конструктор A, снабженный описанием некоторого автомата, построит его копию. Тем не менее A *не является* самовоспроизводящимся: A, *снабженный* копией своего собственного описания, построит копию A, не имеющую своего собственного описания. Переход от универсального конструктора к «самоконструктору» представляет по существу вторую теорему о неподвижной точке.

Первое описание самовоспроизводящегося автомата было дано фон Нейманом в работе «Теория автоматов — конструирование, воспроизведение, однородность», которая осталась незаконченной и которую

А. Беркс подготовил для издания в Иллинойском университете. Он заменил резервуар с плавающими в нем организмами «решеткой» из идентичных элементов, каждый из которых является конечным автоматом с 29 состояниями.

Тэтчер (Thatcher) в 1965 г. дал блестящий вариант схемы фон Неймана*), используя те же составные элементы. В [A]**) показывается, каким образом можно в значительной мере упростить программирование, используя более сложные элементы.

Мы строим сейчас двумерные модели. Использование третьего измерения несколько упрощает конструирование автоматов. Свой элемент я сделаю достаточно сложным, чтобы он мог работать почти в трех измерениях; Вы можете считать мой элемент слоем конечной толщины, состоящим из компонент фон Неймана. При этом каждый элемент, оставаясь двумерным, обеспечивает прохождение пересекающихся потоков информации, не допуская их смешивания. При переходе в третье измерение необходимо иметь в виду высоту отдельных блоков; этот фактор приходится учитывать при выполнении соединений. В одном измерении, надо полагать, возможно ограниченное число самовоспроизведений (при этом конечным совокупностям моих элементов соответствуют единичные элементы одномерной модели).

1. КТ-машины (Тэтчер, 1965) совмещают в себе функции конструирующей машины и машины Тьюринга (рис. 1.13). КТ-машина снабжается программой, использующей следующий конечный список команд:

К-команды (*конструирование*)

u, d, r, l — переместить головку (конструирующую) вверх, вниз, вправо, влево;

$C(x)$ — в просматриваемом квадрате построить x;

*) Thatcher J. W., Universalitu in the von Neumann cellular model, In «Essays in Cellular Automata», ed A. W. Burks, 1965.

**) См. сноску на стр. 56.

Т-команды (обращение к ленте)

$+$ сдвинуть ленту влево на один шаг;

$-$ сдвинуть ленту вправо на один шаг;

l — стереть просматриваемый квадрат (отпечатать 0);

m — пометить просматриваемый квадрат (отпечатать 1);

$t+(n)$, $t-(n)$ — если команда k есть $t\pm(n)$ и просматриваемый на ленте квадрат помечен, перейти

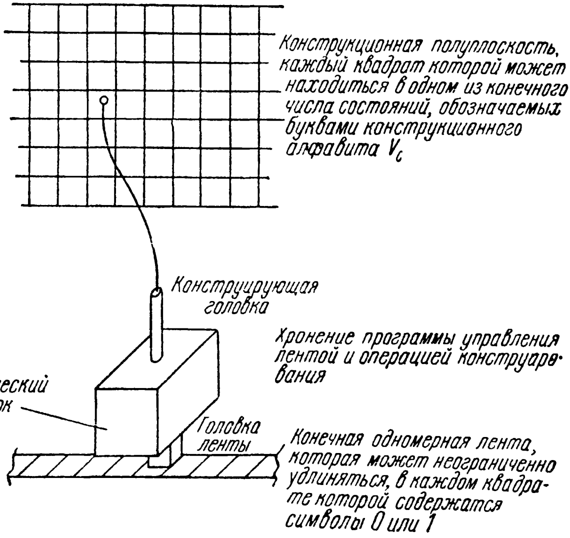


Рис. 1.13. КТ-машина.

к команде $k\pm n$; в противном случае перейти к команде $k+1$;

* останов.

Все элементы машины — идентичные конечные автоматы, отличающиеся один от другого только внутренними состояниями. Одно из состояний считается состоянием покоя. Введя в нашей решетке целочисленную систему координат (m, n), пометим направления следующим образом:

u (=вверх)	увеличение n;
d (=вниз)	уменьшение n;
l (=влево)	уменьшение m;
r (=вправо)	увеличение m.

У каждого квадрата (m, n) есть четыре соседа — элементы с координатами $(m+1, n)$, $(m-1, n)$, $(m, n+1)$ и $(m, n-1)$. Мы программируем поведение элементов и используем операцию их соединения («спайки») для формирования из элементов конфигураций, которые могут перемещаться по плоскости как одно целое: в этом случае лента машины Тьюринга моделируется при помощи одномерной последовательности соединенных между собой элементов, которые по команде могут передвигаться влево или вправо. Наша конструкция обладает следующим свойством: если данные компоненты обладают некоторым уровнем сложности, мы, используя их, можем получить самовоспроизводящиеся конфигурации любого более высокого уровня сложности (например, универсальную машину Тьюринга), где сложность понимается в смысле Рабина, Ритчи, Арбиба и Блюма (см., например, [24, 25, 26]).

Структура нашего основного элемента приведена на рис. 1.14. В каждом из четырех направлений есть входные, выходные каналы и регистры соединений. В состав элемента входят двоичный регистр BR и 20 регистров для записи внутренней программы. Штриховкой отмечены комбинационные цепи, которые служат для надлежащего соединения входов с окружающими регистрами и определения движения в период между моментами t и $t+1$. Кроме того, они служат для определения новых окружающих регистров и выхода модуля в момент $t+1$.

В состоянии покоя все 25 регистров (включая четыре регистра соединений) установлены в нулевое состояние.

Каждый из соединительных регистров может быть включен (состояние 1) или выключен (состояние 0). Говорят, что два соседних элемента соединены, если хотя бы один из соответствующих соединительных регистров находится в состоянии 1.

Пусть W — транзитивное отношение связи элементов, т. е. $W(a, b) \leftrightarrow a$ и b являются связанными

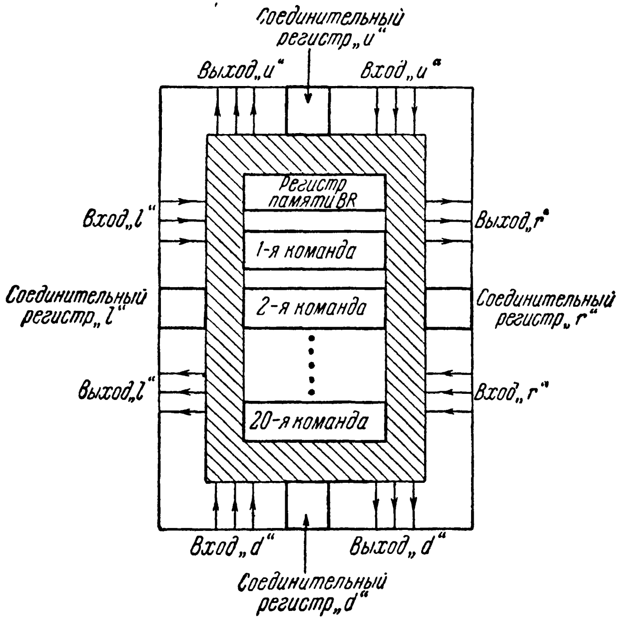


Рис. 1.14. Основной элемент.

соседями, или существует c такое, что $W(a, c)$ и $W(c, b)$. Совокупность C элементов называется *подвижной*, если

$$a \in C \rightarrow b \in C \leftrightarrow W(a, b).$$

Условимся считать, что если *какой-либо* элемент подвижной совокупности C в момент времени t получает команду двигаться в направлении x и ни один другой элемент этой совокупности не получил команды двигаться в другом направлении, то каждый элемент из C в промежуток времени между t и $t+1$ сдвинется на один квадрат в направлении x.

Общая схема машины крайне проста (рис. 1.15). Программа состоит из линейной последовательности подвижных элементов, расчлененной на подпоследовательности из одного или более соседних элементов таких, что самый левый из них имеет $\langle BR\rangle = 1$, а

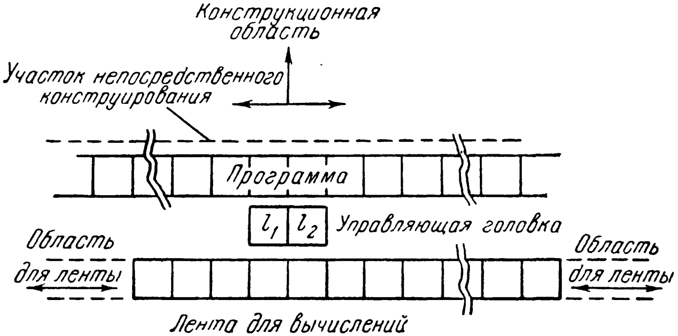


Рис. 1.15. Схема КТ-машины, выполненной на решетке.

остальные (если они есть) имеют $\langle BR\rangle = 0$. m-я последовательность слева представляет собой m-ю команду КТ-программы. Для записи каждой команды используется один элемент. Исключение составляет команда $t\pm(n)$, требующая для своей записи $n\mp 1$ элементов.

Головка для работы с лентой, состоящая из двух элементов, предназначена для считывания T-команд программы и выполнения их. Она уже управляет операциями построения новых элементов, расположенных выше программы, а также сдвигом программы. «Лента» состоит из линейной последовательности подвижных элементов, каждый из которых играет роль квадрата ленты; в регистре $\langle BR\rangle$ каждого элемента находится 1 или 0 в зависимости от того, помечен или не помечен данный квадрат.

2. *Коды команд элементов.* Ниже под A будем понимать одно из направлений: u, l, d, r или lr; $b\in\{0, 1\}$ обозначает содержимое регистра BR или регистра соединений; k, $k'\in\{BR, 1, 2, \ldots, 20\}$ обозначают регистры. Команда «$\langle A\rangle x$» указывает элементу: «выдать команду x в направлении A».

«соединить Ab»	элемент должен заменить содержимое регистра соединений на b в направлении A;
«передача Ak»	элемент должен передать содержимое своего k-го регистра в направлении A;
«переместиться в направлении A» «$A=0$: да (k), нет (k')»	команда перемещения элемента: если сигнал, поступающий с направления A, равен нулю, то в следующий такт надо выполнять команду k; в противном случае — команду k';
«перейти к k»	элемент должен перейти к выполнению команды k (это можно рассматривать как сокращенную форму записи «$A=0$: да (k), *нет* (k')»);
«$\langle A\rangle k$; k'»	в результате выполнения этой команды элементом α содержимое k-го регистра передается в регистр k' того элемента, который является ближайшим соседом в направлении A;
«A поместить b»	элемент в направлении A должен поместить в свой регистр BR величину b;
«останов»	эта команда выполняется только при исчезновении сигналов на входах элемента до тех пор, пока на его входы не поступит команда передачи управления на какую-либо другую команду внутренней программы.

В [A] показано, что элемент, изображенный на рис. 1.14, позволяет реализовать команды m, e, $+$, $-$ и $t\pm(n)$. У нас отсутствует конструирующая головка, подобная той, которая была в пункте 1 для КТ-машины (см. рис. 1.13). Вместо этого мы выполняем операции конструирования (или печати) только в двух квадратах, расположенных над C_1, а затем перемещаем конструируемые элементы программным путем.

3. КТ-машина на любой стадии может быть описана четверкой: P — собственная программа; I — выполняемая в данный момент команда программы; T — состояние ленты; C — состояние конструкционной области.

Для приведенной выше конструкции справедливы

Теорема 1. *Любая конфигурация вида* (P, I, T, C) *эффективно вкладывается в решетку.*

Теорема 2. *Существует эффективная процедура, при помощи которой для любого КТ-автомата* A *можно получить КТ-автомат* $c(A)$ *(«конструктор* A*»), обладающий следующим свойством. Если в начале работы на управляющую головку* $c(A)$ *поступает команда «перейти к* I*», то он начинает строить копию* A *в трех рядах конструкционной области непосредственно над собой, а затем передает на управляющую головку этой копии команду «перейти к* I*».*

Так как при выполнении процедуры, подробно описанной в [A], на один элемент автомата A приходится около трех элементов $c(A)$, то по-прежнему может казаться, что любая машина в состоянии конструировать лишь более простые машины.

Все конфигурации вида (P, I, T, C) эффективно нумеруемы.

Теорема 3. *Существует универсальный конструктор* M_u*, обладающий тем свойством, что если на его ленте записать число* n *в двоичном коде* I_n*, то он построит в своей конструкционной области* n*-ю конфигурацию* M_n

$$I_n\colon M_u \to M_n.$$

Это утверждение можно «доказать» либо путем обращения к модифицированной гипотезе Тьюринга, либо путем построения блок-схемы универсального конструктора M_u.

Почему же M_u не является самовоспроизводящимся? Потому, что хотя $I_n : M_u \to M_u$, но во «втором поколении» $M_u \to ?$, и воспроизведение прекращается. (Элемент, создающий свою копию и не создающий генов, не является самовоспроизводящимся.) Следуя Майхиллу, в [22] доказываются две теоремы.

Теорема. *Для каждой рекурсивной функции h существует такая машина M_c, что*

$$I_n \colon M_c \to M_{h(n)} \quad (c \text{ не зависит от } h).$$

Теорема (фон Нейман). *Существует самовоспроизводящаяся машина.*

4. Совершенно ясно, что описанные нами модели самовоспроизведения могут быть легко выполнены в виде электронных схем, подчиняющихся законам квантовой механики. В связи с этим несколько озадачивают высказывания Вигнера (Wigner, 1961) *), в которых утверждается невозможность процессов самовоспроизведения в квантово-механической системе (он отстаивает точку зрения, что для объяснения многих биологических процессов, в том числе и сознания, недостаточно только одних физических законов; нужны еще особые «биотонические» законы **). В шестом пункте работы [*B*] ***) показывается, в чем, по нашему мнению, заключается ошибка Вигнера.

5. Предложенный автомат можно рассматривать как программу, выполняющую несколько вычислений параллельно (Холланд, Вагнер). В таком случае с точки зрения программирования становится интересным получение воспроизводящейся программы, т. е.

---

*) Wigner E. P, The probability of the existence of a self-reproducting unit, «The Logic of Personal Knowledge: Essays presented to Michael Polanyi», The Free Press, Glencoe, Illinois, 1961, pp. 231—238.

**) Мне кажется, что мы как ученые, не можем утверждать, что биотонические законы не нужны; просто их необходимость не следует из доводов Вигнера.

***) См. сноску на стр. 56.

программы, выполняющей нужные сдвиги в пределах решетки, в результате чего производятся необходимые модификации числовых массивов, и вычисления продолжаются. К сожалению, все публикации, относящиеся к подобного рода системам, являются предварительными. Это говорит о том, что мы еще мало продвинулись в этой области.

6. Пункт 7(*b*) работы [*B*] посвящен некоторым свободным аналогиям с биологией. Мы пользовались значительно более сложными элементами, чем элементы фон Неймана, которые имели 29 состояний. По отношению к проблемам биологии это кажется превосходным, ибо живая клетка, имеющая сложную структуру, включающую сотни метаболических путей, может выполнять любую операцию нашего модуля, так как в управляющей клетке *ДНК*, как мы думаем, содержится больше бит информации, чем в нашем элементе. Так что мы, возможно, во многом проигрываем, чрезмерно ограничивая информационное содержимое нашего элемента.

Сравнивая нашу модель с воспроизведением живых организмов, мы отмечаем следующее:

1. Наша программа была выполнена в виде последовательности клеток, в то время как биологические программы представляют собой последовательности, хранимые в каждой клетке.

2. Мы используем полное описание, в то время как «природа использует» неполное описание.

3. Мы используем лишь незначительную долю вычислительных возможностей нашей модели (последовательное выполнение операций вместо параллельного).

4. Созданные нами конфигурации пассивны, в то время как биологические процессы определяются активным взаимодействием и взаимовлиянием между отдельными подсистемами.

В [*B*] показывается, как модифицировать нашу модель с учетом п. 1. Учесть пп. 2—4 сложнее, и мы не стали рассматривать эту задачу, хотя и отметили, как ее увязать с параллельным способом работы наших устройств.

Глава 2

Структура и случайность

§ 2.1. Зрительная система лягушки

В этом параграфе мы обсудим один частный пример *структуры* мозга. С этой целью мы рассмотрим работы Леттвина (Lettvin), Матурана (Maturana), Маккаллока и Питтса *). Вместе с ними мы изучим структуру зрительной системы лягушки.

Лягушки питаются насекомыми, которых они обнаруживают при помощи зрения. Они ловят только движущихся насекомых; неподвижные предметы никогда не привлекают их внимания. Большой движущийся объект вызывает у них двигательную реакцию в виде прыжка. Неподвижный же объект, по-видимому, не изменяет их поведения. Скорее всего они опознают свою жертву и выбирают ее из окружающих объектов потому, что она обладает рядом таких особенностей, как движение, размер, определенный контраст и, возможно, определенный цвет. Эта способность лягушек опознавать свою жертву и ловить ее не меняется при некоторых изменениях в окружающей среде, например при изменении освещения.

Указанные выше авторы исследовали зрительную систему лягушки с целью установления тех свойств

---

*) «What the Frog's eye tells the Frog's brain», Proc. IRE, **47** (1959), 1940—1951. [Русский перевод: Леттвин Дж., Матурана Г., Маккаллок У. и Питтс У., Что сообщает глаз лягушки мозгу лягушки. В сб. «Электроника и кибернетика в биологии и медицине», М., ИЛ, 1963.]

«Physiology and anatomy of vision in the Frog», J. Gen. Physiol. **43** (Suppl.) (1960), 129—175; «Two remarks on the visual system of the Frog», in W. Rosenblith (ed.), «Sensory Communication», The M. I. T. Press, Cambridge, Mass., 1961, 757—776.

этой системы, которые позволяют лягушке различать понятия (universals *)): *враг* и *жертва*.

На рис. 2.1 изображены: глаза лягушки (E), идущие от них зрительные нервы (O) (которые затем

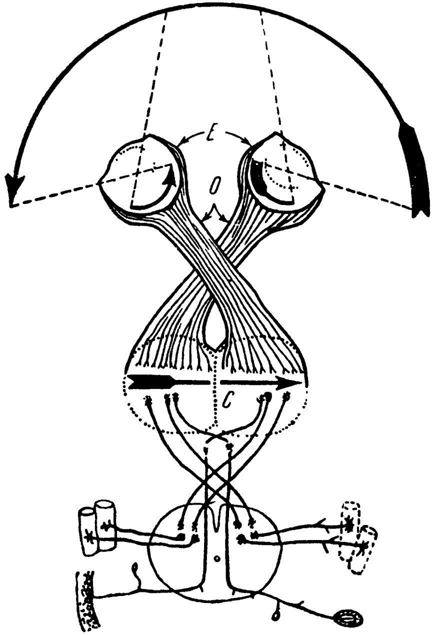


Рис. 2.1. Схематическое изображение зрительной системы лягушки.

пересекаются) и отдел мозга (C), в котором они заканчиваются. Этот отдел мозга называется *верхним перешейком* (superior colliculus) или *зрительным бугром* (optic tectum). Остальные указанные на рисунке отделы мозга здесь не обсуждаются. Стрелки дают

*) В Оксфордском словаре говорится, что «universal» — это то, что утверждается обо всех индивидуумах или разновидностях класса или вида; абстракция или основное понятие, рассматриваемые как имеющее абсолютное, образное или номинальное существование; основной термин или понятие.

некоторое представление о том, как зрительные стимулы отображаются в зрительном бугре.

Луч света, проникая в глаз (рис. 2.2), достигает дна, где он попадает на сетчатку, проходит через прозрачные ганглии и промежуточные клетки и, наконец, достигает палочек и колбочек. Возбужденные в достаточной степени светом палочки и колбочки создают непрерывно меняющуюся разность потенциалов,

Рис. 2.2. Прохождение луча света через сетчатку лягушки.

которая через промежуточные нейроны вызывает импульсы. Аксоны этих промежуточных клеток в свою очередь соприкасаются с дендритами ганглиозных клеток (нейронов). Волокна зрительных нервов — это аксоны ганглиозных клеток. Эти волокна проходят через сетчатку и, собравшись вместе, образуют идущий к мозгу пучок, называемый *зрительным нервом*. Место выхода нерва из глаза называется *слепым пятном*.

У лягушки имеется около миллиона рецепторов (палочек и колбочек), от 2,5 до 3,5 миллионов промежуточных нейронов и полмиллиона ганглиозных клеток. Ввиду такого большого числа клеток кажется маловероятным, чтобы сложная структура сетчатки действовала просто как передатчик, который передает в мозг изображение света и тени, образованное на рецепторах. Скорее можно предположить, что сетчатка анализирует зрительный образ и передает полученную информацию в зрительные центры.

В 1938 г. Хартлайн (Hartline) показал*), что ганглиозные клетки можно разбить на три класса в соответствии с их реакцией в ответ на ограниченные локальные световые раздражения их поля восприятия. *Поле восприятия* ганглиозной клетки определяется как та часть поля зрения, которая отображается на множестве всех тех рецепторов, возбуждение которых влияет на данную ганглиозную клетку. Эти классы таковы:

1. Клетки, которые длительно, но с некоторым запозданием реагируют на появление пятна света.
2. Клетки, которые реагируют на появление и исчезновение света кратковременной серией частых импульсов.
3. Клетки, которые сразу и в течение длительного времени реагируют на исчезновение пятна света.

Эти наблюдения явились первыми, показавшими, что ганглиозные клетки выполняют несколько сложных операций над зрительным образом. Но пятна света являются для лягушки не настолько естественными раздражителями, как муха или червь. Это позволяет предположить, что ганглиозные клетки до некоторой степени повторяют, но более грубым и непостоянным образом, первоначальную конфигурацию зрительного образа с некоторыми локальными различиями. Но восприятие понятий, что следует из поведения лягушки, по-видимому, требует наличия некоторых функциональных инвариантов деятельности составных частей ее зрительной системы. Таким образом, мы предполагаем, что сетчатка, возможно, выполняет вычисления над зрительным образом, которые способствуют выявлению таких характерных свойств, как наличие врага или жертвы. Леттвин и другие пытались найти подходящие функциональные инварианты и, следовательно, аналитические функции ганглиозных клеток, применяя естественно-научный подход и изучая их в терминах реакций на реальные объекты окружающей среды.

---

*) Hartline H. K., The response of single optic nerve fibres of the vertebrate eye to illumination of the retina, Amer. J. Physiol. 121 (1938), 400—415. (*Прим. ред.*)

В своих исследованиях они использовали «обычную американскую лягушку» (*Rana pipiens*). Лягушку помещали таким образом (рис. 2.3), чтобы один глаз был в центре полусферы диаметром 14 дюймов, которая служила экспериментальным полем зрения. Электрод вводили в лягушку так, что он мог

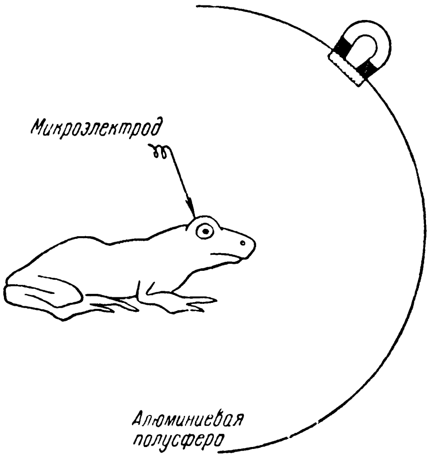


Рис. 2.3. Установка для изучения зрительной системы лягушки.

реагировать либо на активность одной ганглиозной клетки (конец электрода помещали на аксон в зрительном нерве), либо на активность клетки в зрительном бугре. Алюминиевая полусфера составляла около двух третих поля зрения одного глаза лягушки. Подходящей ориентацией животного можно было охватить любую желаемую часть поля зрения и полностью контролировать поле восприятия изучаемых клеток. Стимулирующие предметы перемещались по внутренней поверхности с помощью магнита, который двигался по внешней поверхности. Было использовано много различных по форме и виду предметов, например темные диски, темные полоски и темные квадраты (*заметьте, что стимулы не предполагали реакции на цвет* *)).

Было обнаружено, что по своей реакции ганглиозные клетки распадаются на 5 групп.

Группа I. Детекторы границы. Эти волокна имеют поля восприятия от 2 до 4 градусов в диаметре. Они реагируют на любую границу между двумя оттенками серого в поле восприятия, если она является резкой. Измеряется, по-видимому, скорее резкость границы, чем степень контраста. Реакция усиливается, если граница движется, и не меняется, если освещение данного контраста изменяется в очень широких пределах. Если в поле нет никакой границы, то реакцию нельзя вызвать путем изменения яркости, каким бы резким оно ни было. Другим свойством клеток первой группы является то, что если в полной темноте в поле восприятия возникает граница и затем включается свет, то после короткой паузы наступает длительная реакция.

Аксоны первой группы — это волокна первого класса в классификации Хартлайна — малое, хорошо сфокусированное пятно света определяется как резкая граница.

Группа II. Детекторы движущейся границы и границы выпуклой темной области. Эти волокна имеют поля восприятия от 3 до 5 градусов. Они также реагируют только на резкие границы между двумя оттенками серого, но только в том случае, если граница является кривой и движется, а темная область выпукла. Реакция опять же не меняется в широких пределах освещенности, грубо говоря, между тусклыми сумерками и ярким полднем. Они не принадлежат ни к одному из классов Хартлайна и не реагируют на световые пятна и раздражители, у которых светлые области хотя и выпуклы, но неподвижны.

Группа III. Детекторы меняющейся или движущейся контрастности. Эти волокна имеют поля вос-

---

*) Реакция на цвет обсуждается в статье: W. R. A. Muntz, Effectiveness of different colors of light releasing the positive photoactic behavior of Frogs ... , J. Neurophysiol. 25 (1962), 712—720.

приятия от 7 до 11 градусов в диаметре. Они, в сущности, относятся ко второму классу классификации Хартлайна. Однако их реакция не меняется в широком диапазоне освещенности, если один и тот же силуэт движется с постоянной скоростью на одном и том же фоне. Хотя их реакция непродолжительна, они включаются только тогда, когда происходит смещение или изменение контрастности. Реакция тем интенсивнее (выше по частоте), чем резче граница и чем быстрее она перемещается.

Группа IV. «Сумеречные» детекторы. Эти детекторы относятся к третьему классу классификации Хартлайна. Их поле восприятия достигает 15 градусов. Они реагируют на любое потускнение всего поля восприятия с учетом расстояния до центра этого поля. При реакции границы не играют никакой роли. Одна и та же степень потускнения дает ту же реакцию, более или менее независимую от уровня первоначальной освещенности.

Группа V. Эта группа малочисленна. Мы не можем даже сказать, имеет ли она поле восприятия в обычном смысле. Клетки этой группы возбуждаются с частотой, обратно пропорциональной интенсивности средней освещенности. Когда освещенность меняется, частота медленно изменяется до своего нового уровня.

Каждая ганглиозная клетка принадлежит только к одной из этих групп, а клетки каждого класса равномерно распределены по сетчатке. В любой небольшой области сетчатки можно обнаружить представителей всех групп пропорционально их общему соотношению.

Аксоны клеток каждой группы оканчиваются в разных слоях зрительного бугра. Однако два из них совмещены (нервные окончания пятой группы лежат в слое нервных окончаний третьей группы), так что в действительности они образуют четыре основных слоя нервных окончаний. Каждый из этих четырех слоев нервных окончаний формирует в зрительном бугре «непрерывный» образ сетчатки в соответствии с операцией, выполняемой соответствующими ганглиозными клетками. Эти четыре слоя производят регистрацию, и в любую достаточно малую область зрительного бугра приходят нервные окончания всех слоев одного и того же места сетчатки.

Таким образом, функция сетчатки лягушки не сводится к простой передаче всей мозаичной картины светлых и темных пятен образовавшегося на ней образа. Напротив, она заключается главным образом в анализе каждой точки этого изображения по четырем признакам (граница, движущиеся кривые, меняющаяся контрастность и локальное потускнение), степени освещенности и передаче этой информации в зрительный бугор, где эти функции разделяются в четырех конгруэнтных слоях*) нервных окончаний.

Сетчатка преобразует зрительный образ из мозаики освещенных точек в систему частично перекрывающихся качественных областей, в которой любая точка описывается в терминах ее связи с окружающей окрестностью. Так как преобразование является основной функцией сетчатки, то особое значение имеет интегрирующая способность ганглиозных клеток. Эти соображения побудили Леттвина и его коллег анатомически исследовать способность ганглиозных клеток комбинировать поступающую к ним информацию, пытаясь определить связь между различными морфологическими типами ганглиозных клеток (типы отличаются структурой их древовидных дендритов) и операциями, которые они выполняют. Обратимся к их исследованиям в этом направлении.

Существует пять анатомически различных ганглиозных клеток (рис. 2.4). Вы можете заметить, что показанная на рисунке классификация дендритов позволяет сделать предположение о существовании двух основных подразделений внутреннего сетчатого слоя (inner plexiform layer**)), хотя, несомненно, каждое подразделение можно в свою очередь разбить на несколько слоев. За обсуждением значения этого факта

---

*) Точки различных слоев, которые перемешаны между собой, сходятся в одной небольшой области сетчатки.

**) Слой сетчатки, занимаемый дендритами ганглий, называется *внутренним сетчатым слоем*.

мы отсылаем читателя к третьей из оригинальных работ, упомянутых в начале этой главы.

I. *Узкое однослойное поле.* Это мельчайшие ганглиозные клетки. Главные дендриты простираются только до внутренних слоев внутреннего сетчатого слоя и там ветвятся плотным и узким плоским кустом.

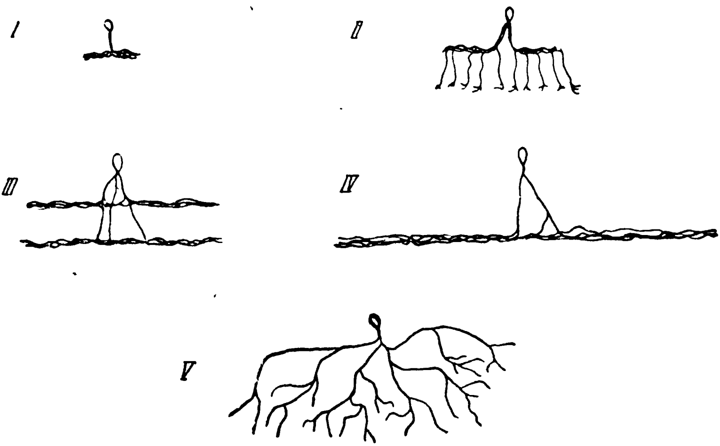


Рис. 2.4. Пять типов ганглиозных дендритовых клеток.

II. *Многослойное Е-распределение.* Это следующие по величине ганглиозные клетки. Главные дендриты простираются только до внутренних слоев внутреннего сетчатого слоя и там ветвятся плоским образом. Однако каждая ветвь по всей своей длине имеет отростки; некоторые из них тянутся до внешних слоев внутреннего сетчатого слоя, а другие остаются во внутренних слоях.

III. *Многослойное Н-распределение.* Следующие по величине ганглиозные клетки. Главные дендриты простираются до внешних слоев внутреннего сетчатого слоя. Однако от них отходят два широко распростершихся дерева, одно — во внутренних, а другое — во внешних слоях.

IV. *Широкое однослойное поле.* Это самые большие по величине ганглиозные клетки. Главные дендриты простираются до внешних слоев внутреннего сетевидного слоя и там широко ветвятся над значительной областью.

V. *Диффузные деревья.* Существует несколько размеров этих клеток. Дендриты ветвятся хаотически по всему внутреннему сетчатому слою и не имеют плоского расположения, как это было в четырех предыдущих видах клеток.

Если предположить, что размер поля дендритов до некоторой степени определяет размер поля восприятия, то появится довольно тесная связь между типами клеток и операциями.

I. Узкое однослойное поле	Обнаружение границы.
II. Многослойное Е-образное поле	Детекторы движущейся темной выпуклой границы.
III. Многослойное Н-образное поле	Детекторы движущегося и меняющегося контраста.
IV. Широкое однослойное поле	Сумеречные детекторы.

Редкий диффузный тип мы более или менее произвольно сопоставляем группе, измеряющей средний уровень освещенности.

Вышеуказанные сопоставления подтверждаются следующими фактами. Диаметры полей дендритов совпадают с угловыми диаметрами полей восприятия; тела клеток распределены по размерам таким же образом, как поля дендритов. Если диаметры аксонов отражают размер тела, то наибольшие аксоны должны иметь наибольшие поля восприятия, а наименьшие аксоны — наименьшие поля; по-видимому, это так и есть. Далее, поля восприятия часто бывают не круглые, а эллиптические или кардиоидные. На рисунках ганглиозных клеток, выполненных Матураном, имеются как эллиптические, так и кардиоидные дендритовые поля.

Переходя от ганглиозных клеток к клеткам зрительного бугра, Леттвин и его коллеги обнаружили несколько типов клеток, хотя и не смогли выделить подгруппы. Однако имеются две совокупности, которые они назвали «нейронами новизны» и «нейронами того же самого». Первые, по-видимому, связаны с обнаружением новизны и видимых событий, а вторые — с непрерывностью во времени интересных объектов в поле зрения.

2.1.1. Сравнения

В эмбриологии установлено, что сетчатка по существу является частью мозга. Работа Леттвина, Матурана, Маккаллока и Питтса обнаружила, далее, некоторую фундаментальную структуру мозга лягушки. Следует, конечно, подчеркнуть применимость этой работы только к лягушке — ни анатомия, ни действие поля восприятия лягушек *не обязательно* являются такими же, как у млекопитающих или других земноводных. Подробная работа о зрительной системе кошки выполнена Д. Х. Хюбелем и Т. Н. Визелем *). В зрительной коре кошки они обнаружили клетки, сравнимые с теми, которые Леттвин и другие нашли в зрительном бугре лягушки. Они говорят: «На первый взгляд может показаться удивительным, что сложность нейронов третьего порядка зрительной системы лягушки должна быть сравнима только со сложностью нейронов шестого порядка в коленчато-корковом пути (geniculo-cortical pathway) кошки. Однако это становится менее удивительным, если у обоих животных обнаруживаются большие анатомические различия, особенно отсутствие у лягушки какой бы то ни было коры или дорзально-латерального коленчатого тела. Несомненно, существует параллельное различие в том, как каждое животное использует свою зрительную систему: зрительный аппарат лягушки, по-видимому, специализирован для распознавания более ограниченного числа стереотипных конфигураций или ситуаций по сравнению с высокой остротой и многогранностью той же системы у кошки.

---

*) D. H. Hubel, T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. **160** (1962), 106—154.

Вероятно было бы целесообразно выяснить, что у кошки специализация клеток для сложных операций отодвигается до более высокого уровня и осуществляется в мельчайших деталях огромным числом клеток».

Чего же можно ждать от высокоорганизованной структуры человеческого мозга?

§ 2.2. Персептрон

В первой главе мы кратко остановились на нейрофизиологии и построили абстрактную модель мозга — модульную сеть, или сеть из нейронов Маккаллока — Питтса. Персептрон, построенный сотрудниками Корнуэльского университета, работает на несколько отличной основе, чем модель нейронной сети. Главное различие заключается в том, что его создатели не стали предполагать, что функции нейронов жестко фиксированы. Вместо этого они допустили, что веса каждого нейрона могут меняться со временем. Цель этого заключается в том, чтобы дать возможность нейронной сети изменяться со временем таким же образом, как при «обучении».

Персептрон можно представлять себе как прибор для распознавания образов, который построен не для того, чтобы узнавать специфическое множество образов, а скорее для того, чтобы «научиться» после конечного числа испытаний узнавать образы из некоторого множества.

В персептрон образ поступает через *сетчатку*, состоящую из *чувствительных элементов* (например, фотоэлементов). Хотя любое воспринимаемое изображение может быть закодировано в форме, подходящей для входа, состоящего из набора чувствительных элементов, однако наиболее естественно (о чем и говорит термин «сетчатка») представлять образ как наблюдаемый вход, состоящий из света и тени. Фотоэлемент, который воспринимает относительно светлую часть образа, возбуждается, а фотоэлемент, воспринимающий относительно темную часть образа, не возбуждается.

После детального рассмотрения в § 2.1 сетчатки лягушки становится ясным, что эта двоичная реакция, связанная с освещением сетчатки, имеет мало сходства с физиологической деятельностью. Чувствительные элементы соединены с *промежуточными элементами* (формальными нейронами), которые в свою

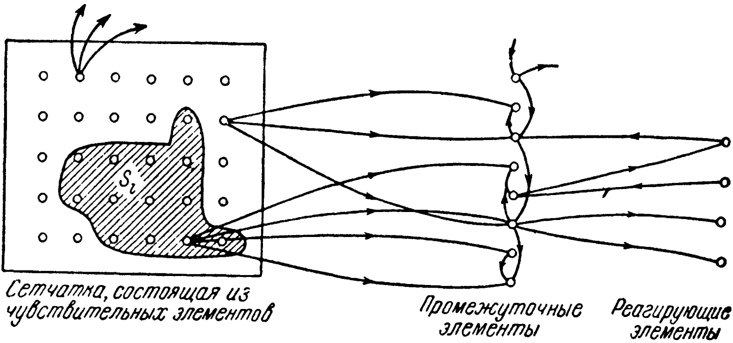


Рис. 2.5. Схематическое изображение персептрона.

очередь могут быть соединены с *реагирующими элементами* (рис. 2.5).

В терминах нашего первоначального нейрофизиологического рассмотрения нервной системы как трехступенчатой системы сетчатка является рецептором персептрона; промежуточные элементы охватывают собственно нервную систему, а реагирующие элементы являются эффекторами или, скорее, соответствуют нейронам, выходы которых управляют эффекторами. С нашим первоначальным рассмотрением совместимо также и то, что когда на сетчатку персептрона действует *изображение*, то импульсы передаются от возбужденных чувствительных элементов к промежуточным элементам. Если сигнал, поступающий на промежуточный элемент, превосходит его порог, то этот элемент становится *активным* и посылает импульсы к элементам, с которыми он связан.

До сих пор персептрон являлся просто другим воплощением сильно упрощенных нейрофизиологических данных о нервной системе, снабженной зрительными рецепторами. Однако конструкторы персептрона пошли дальше этого, так что здесь стоит обсудить дополнительные свойства сети.

1. Правила поощрения

Существует, по-видимому, много подтверждений того, что люди имеют два вида памяти: «кратковременную» и «долговременную». Кроме того, кажется, что мы сначала должны запомнить мысль в кратковременной памяти, прежде чем передать ее в долговременную память. Время, необходимое для этой передачи, оценивается по-разному; в частности, одна из таких оценок равна 20 минутам. По-видимому, если кто-то впадает в бессознательное состояние, то его воспоминания о только что прошедших 20 минутах потеряны навсегда, так как они не были переданы в долговременную память. В настоящее время общепринято считать, что кратковременная память представляет собой в точности тот же тип памяти, которой мы снабдили нашу модульную сеть, т. е. является прохождением по сети сложных конфигураций электрических импульсов. Далее, если такая временная активность продолжается довольно долго, то она

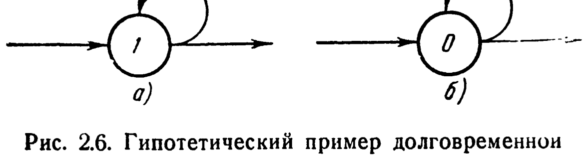


Рис. 2.6. Гипотетический пример долговременной памяти.

фактически изменяет сеть. Это хорошо иллюстрируется простым примером (рис. 2.6).

Модуль показанный на рис. 2.6, *а*, хранит в кратковременной памяти информацию о том, был ли когда-нибудь возбужден его вход. Эта память сохраняется за счет образовавшегося в петле импульса. Модуль имел бы долговременную память, если бы кратковременная память могла, например, понизить его порог с 1 до 0, ибо в этом случае память сохранилась бы даже после исчезновения в петле импульса.

Один из возможных механизмов долговременной памяти заключается в образовании в нейронах специальных протеинов, которые изменяют их пороги в ответ на конфигурации кратковременной памяти. Другой механизм основан на росте концевых луковиц при повторном использовании и, таким образом, увеличивается вес соответствующего синапсического входа и, вероятно, облегчается восстановление конфигураций импульсов, использующих этот синапс. Это облегчает воспоминание прежней информации! Но истинный механизм все еще неизвестен. Поэтому физиология, гистология и анатомия, поддерживающие пока многие гипотезы, должны еще установить это *).

Персептрон снабжен долговременной памятью. Это достигается изменением весов входов нейронов, или — на языке конструкторов персептрона — изменением величины импульса, поступающего от раздражителя. Эти изменения зависят от прошлой активности окончаний. Закон этой зависимости называется *законом поощрения*, так как он служит для поощрения правильных реакций персептрона на действие раздражителя. Сведения из физиологии настолько расплывчаты, что конструкторы персептрона выбрали такие правила поощрения, которые были удобны в теоретическом и экспериментальном отношениях. Их выбор дает персептрону возможность демонстрировать некоторые виды обучения.

2. Случайные соединения

Конструкторы персептрона должны были решить, каким образом надо соединить чувствительные элементы с промежуточными элементами. Первый

---

*) Краткую дискуссию о биологической памяти и прекрасную библиографию можно найти в работе: Hans-Lucas Teuber, Perspectives in the problems of biological memory — a psychologist's view. In F. O. Schmitt (ed.), Macromolecular specificity and biological memory, The M. I. T. Press, Cambridge, Mass., 1962.

подход заключался в том, чтобы соединения были много-многозначными и случайными. Позднее они включили в схему некоторые подсети, например подсети для линейного распознавания, подобные тем, которые нашли в кошке Хюбель и Визель.

Но подобно тому, как наша весьма упрощенная модульная сеть все же была способна хранить информацию и производить вычисления, так и первые персептроны, несмотря на все ограничения, проявляли простые свойства обучения. Эти свойства обсуждаются в оставшейся части этого параграфа.

Конструкторы персептрона провели три основных вида исследований: математический анализ, моделирование на цифровой вычислительной машине и построение действующей машины. Каждый метод имеет свои преимущества. Одним из важных результатов использования действующей машины является выяснение того, что *ни точность, ни надежность компонент не имеют большого значения, а соединения не обязаны быть точными.*

Другой интересный результат заключается в том, что персептрон может «учиться», *несмотря на ошибки учителя.*

Простым персептроном является такой персептрон, в котором промежуточные элементы не соединены между собой. *Это означает, что у него нет кратковременной памяти!*

Пусть имеются чувствительные элементы s_k, промежуточные элементы a_m и n образов S_i. В этой простой модели соединения между s_k и a_m не меняются (рис. 2.7). Следовательно, множество промежуточных элементов, возбужденных образом S_i, остается неизменным.

Пусть $e_{mi}=1$ тогда и только тогда, когда промежуточный элемент a_m не возбужден образом S_i. Если V_m есть вес входа реагирующего элемента, подключенного к промежуточному элементу a_m, когда перед сетчаткой находится образ S_i, то вес входа реагирующего элемента равен

$$u_i=\sum_m V_m e_{mi}.$$

Предположим, что мы разбили имеющиеся образы на два класса: $+1$ и -1. Пусть r_i — один из классов, содержащий образ S_i. Мы хотим, чтобы реагирующий элемент возбуждался тогда и только тогда, когда данный образ принадлежит классу $+1$.

Одним из правил поощрения является *процедура исправления ошибки*, которая заключается в следующем. Персептрону показывается образ S_i он дает

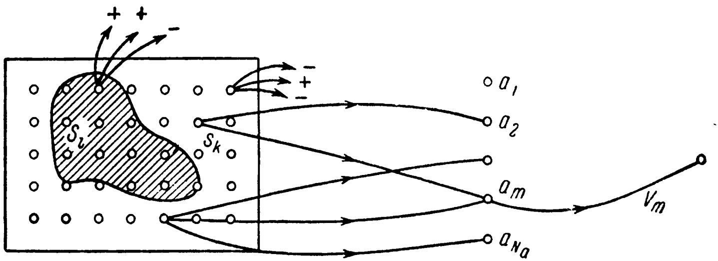


Рис. 2.7. Схематическое изображение простого персептрона.

ответ. В случае правильного ответа никакого подкрепления не делается. Если ответ неправильный, то V_m для *возбужденных* промежуточных элементов увеличивается на ηr_i (η — фиксированное положительное число). Невозбужденные промежуточные элементы остаются неизменными. Начальные значения произвольны, скажем,

$$V_1^0, \ldots, V_m^0, \ldots, V_N^0.$$

Предположим, что образы показываются в произвольном порядке так, что каждый образ повторяется достаточно часто. Блок *) показал, что после некоторого конечного числа шагов машина будет давать правильные ответы на все образы, так что никаких дальнейших изменений не произойдет. Далее он показал,

*) H. D. Block, The perceptron: A model for Brain functioning. I. Rev. Mod. Phys. **34**, № 1 (1962), 123—135.

что порядок числа исправлений равен числу различных образов, подаваемых на сетчатку. Этот результат показал, что машина может выучивать наизусть, но (как мы далее увидим) использование простого персептрона в качестве прибора для распознавания образов нежелательно.

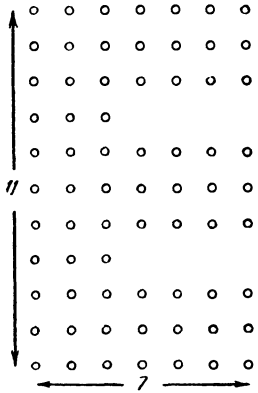


Рис. 2.8. Изображение буквы E, которое может восприниматься чувствительными элементами персептрона.

Персептрон «Марк-1» является простым персептроном. Он имеет сетчатку размером 20×20 фотоэлементов. Рассмотрим случай, когда в качестве образов используется 26 букв (английского) алфавита (*каждая буква задается в стандартной форме*), а выход состоит из пяти двоичных реагирующих элементов ($2^5 = 32 >$ > 26). В проведенном эксперименте, как сообщалось, машина научилась правильно распознавать буквы после 15 показов каждой буквы, т. е. общее число показов было равно 390. Мы же требуем от системы для распознавания нечто большее, чем только выделение образов в стандартной форме. Нам нужна машина, которая может распознавать каждую букву, помещенную на сетчатку, даже если она напечатана слегка искаженно и показана на пестром фоне. Но если мы допускаем это, то число образов катастрофически возрастает. В этом можно убедиться при рассмотрении различных вариантов стандартной буквы E, образ которой представлен в виде точек, как показано на рис. 2.8.

Следовательно, сделанный выше вывод о том, что для заучивания наизусть достаточно произвести некоторое число опытов, по порядку равное числу различных образов, является не столь значительным. Простой персептрон оказывается слишком простым. Этот простой персептрон мы обсуждаем из методических соображений; он полезен при рассмотрении вопроса о памяти нейронных сетей. На самом деле конструкторы персептрона проделали более сложную работу. Так как для своих сетей они ставили более трудные задачи, то и строили их с более сложной структурой. Их работа даже включает нейрохимическое исследование памяти и описание сопутствующих моделей.

Прежде чем закончить обсуждение этой темы, надо честно предупредить читателя, что связанная с персептроном работа подвергалась резкой критике. Конечно, в первых статьях конструкторов персептрона имелись большие претензии, и читать их следует с большой осторожностью. Кроме того, многих отталкивают довольно неудачные сообщения в широкой печати. Однако не следует реагировать так, чтобы отвергать всю последующую работу о сетях, которые могут «учиться».

§ 2.3. Структура против случайности

Анатомия и физиология той части нашего мозга, работа которой относится к высшей нервной деятельности, мало изучена. Хотя макроскопическая анатомия мозга обнаружила сложную структуру, микроскопическая анатомия дает смутную картину, по-видимому, случайных взаимосвязей. Кажется невозможным, чтобы наши гены определяли точную структуру нашего мозга. Гораздо более вероятно, что они определяют некоторые схемы роста, более или менее подверженные меняющимся влияниям опыта. Кроме того, если бы связи были даже строго детерминированы, мы все равно не знаем механизма, посредством которого мозг может распознавать понятия, например узнавать букву А в разных видах и при некоторых искажениях (теория этого вопроса все-таки имеется, см. § 4.4). Поэтому конструкторы персептрона допускают, чтобы первоначальная структура была случайной, а та структура, которая необходима для распознавания образов, возникала в результате изменений, вызванных правилами поощрения. Их подход представляется интересным, но я все же чувствую, что в нем чего-то не хватает.

В этой главе мы рассмотрели зрительную систему лягушки и обнаружили подтверждение того, что по крайней мере в лягушке некоторая очень важная структура предопределена генетически. В § 3.5 мы изложим теорию Винограда — Кована (Winograd — Cowan). Там мы убедимся, что если модульную сеть, предназначенную для определенной цели, преобразовать таким образом, чтобы уменьшить ее чувствительность к ошибкам, появляющимся вследствие неправильного функционирования или гибели нейронов, то в результате получится сеть, имеющая микроскопическую случайность, очень похожую на ту, которая обнаружена микроскопической анатомией мозга. В качестве последнего аргумента против использования полной случайности укажем на то, что имеются умственные акты, доступные ребенку, но совершенно недоступные для гориллы. Это происходит, возможно, вследствие генетически детерминированных различий в структуре. Дарвиновской эволюции понадобились тысячелетия, чтобы сделать наш мозг способным узнавать образы. Было бы крайне удивительно, если бы случайная сеть приобрела такую способность за несколько часов обучения.

Однако я должен признаться, что все вышеприведенные аргументы сводятся к следующему утверждению: адекватная модель человеческого мозга должна быть богата различными специфическими структурами, вроде той, которая обеспечивает восприятие прямых линий. Но это не помогает ответить на вопрос: необходима ли структура для обучения? Другими словами, если допустить, что человеческий мозг обладает структурой, то мы не достигнем еще никакого прогресса в выяснении следующих двух противоположных точек зрения:

а) Человек разумен потому, что эволюция снабдила его мозгом с очень сложной структурой. Эта структура выполняет различные функции и, в частности, дает ему возможность учиться. Требуется некоторая критическая степень сложности структуры, чтобы сеть могла изменять себя (независимо от сложности ее правил поощрения) таким способом, который мы могли бы считать разумным.

б) Человек разумен потому, что эволюция снабдила его мозгом с очень сложными взаимосвязями. Схема взаимосвязей не имеет отношения к истинно разумному обучению, которое является результатом действия правил поощрения на достаточно большую, но существенно случайную сеть.

Возможно, что истина заключается в искусном соединении этих точек зрения. Поиск такого соединения будет одним из самых захватывающих этапов в будущем этого предмета.

Глава 3

Исправление ошибок при передаче и вычислениях

Индивидуальные нейроны мозга могут работать неправильно. Кроме того, несколько тысяч нейронов мозга человека ежедневно погибают и никогда не замещаются. Поэтому очень интересно знать, каким образом такой мозг может функционировать с высокой устойчивостью и надежностью.

Основная цель этой главы заключается в ознакомлении читателя с теорией Винограда—Кована по функционально устойчивым автоматам — модульным сетям, которые надежно функционируют независимо от неисправностей модулей. Эта теория основана на хорошо известной шенноновской теории связи при наличии шумов. Поэтому большую часть этой главы мы посвящаем рассмотрению теории Шеннона — теории, имеющей большое значение и представляющей самостоятельный интерес.

§ 3.1. Надежный мозг из ненадежных нейронов

Этот параграф посвящается краткому обзору некоторых нейрологических данных, которые показывают, что мозг функционирует точно и устойчиво, несмотря на гибель и неправильное функционирование нейронов. Мы широко цитируем обзорную работу У. С. Маккаллока, М. А. Арбиба и Дж. Д. Кована *).

В этой книге мы будем строго придерживаться электрической гипотезы центрального торможения и возбуждения. Она требует, чтобы электрические свой-

*) Neurological models and integrative processes, in M. C. Yovits, G. T. Jacobi and G. D. Goldstein (eds.), Self-organizing Systems, 1962, Spartan Books, Washington, D. C. (1962), 49—59.

ства передачи и геометрия структур определяли выход как специфическую функцию входа каждой ткани.

Возможно, самым простым примером такого однозначного соответствия между структурой и ее функционированием является верхняя олива (superior olive) (рис. 3.1), крупные клетки которой получают сигналы от обоих ушей. Анатомия этих клеток и их

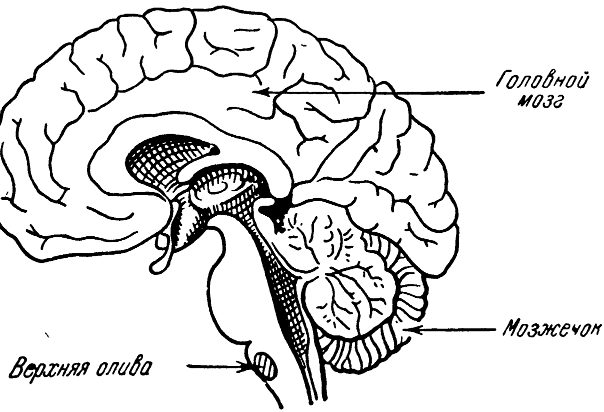


Рис. 3.1. Мозг человека в поперечном разрезе.

синапсов такова, что они возбуждаются от сигналов, идущих только от одного уха, тогда как сигналы, идущие одновременно от обоих ушей, гасят друг друга, и клетка остается невозбужденной. С помощью этих клеток мы определяем направление звука и слышим сигнал в одном ухе, несмотря на значительный общий шумовой фон в обоих ушах.

Вторым примером может служить мозжечок (рис. 3.1), который состоит из свода крупных клеток, каждая из которых получает сигналы от вестибулярных органов обоих ушей и передает информацию об ускорениях головы. На вход поступают сигналы почти от всех частей тела и остальной части мозга. Они поступают на множество мелких клеток, тонкие и однотипные аксоны которых делятся, пересекая ответвления гигантских клеток, как провода пересекают параллельные ряды регулярно расположенных телефонных столбов; многие тысячи проводов присоединены к каждому столбу своего ряда. Мозжечок выполняет роль часов во всех двигательных актах, таких, как письмо, и других, в которых ускорения и замедления должны быть точно отмерены и зарегистрированы в промежутки времени, слишком короткие, чтобы в этом процессе могли участвовать рефлексы.

Основанием для рассмотрения верхних оливов и мозжечка является то, что функция ткани принадлежит не какому-либо отдельному нейрону, а распределена по всем нейронам. Анатомически это проявляется в многократном повторении сходных компонентов и перекрывающихся входов. Физиологически это обнаруживается во временном запаздывании, так как оба органа срабатывают, один выполняя функцию обнаружения, а другой — засечения времени, с точностью до 1—2 микросекунд, тогда как их компоненты как при обнаружении совпадения импульсов, так и при их испускании имеют стандартное отклонение, большее 1/3 миллисекунды. Из данных патологии следует, что потеря точности незначительна, несмотря на гибель одной десятой доли числа всех клеток. *Мы подчеркиваем, что как обе верхние оливы, так и мозжечок функционируют более устойчиво, чем их компоненты, даже если каждая компонента выполняет одну и ту же функцию, но в разное время от различного множества входов.*

Прежде чем продолжить наши рассуждения, мы должны ввести некоторые новые понятия. Входы нейрона называются *афферентами*. Под *взаимодействием афферентов* мы понимаем ситуацию, схематически показанную на рис. 3.2, когда афферент не возбуждает или тормозит сам нейрон, а блокирует импульс в другом афференте. Таким образом, если на этом рисунке импульс, идущий по A, достигает X одновременно с импульсом, идущим по B, то его прохожде-

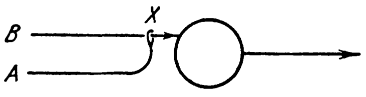


Рис. 3.2. Взаимодействие эффектов.

ние прекращается; если же он не встречается с другим импульсом, то его прохождение продолжается и приводит к возбуждению или торможению этого нейрона.

Под булевой функцией мы понимаем функцию $f(x_1, \ldots, x_n)$ *), аргументами и значениями которой являются только два числа 0 и 1. Таким образом, мы можем считать, что отношение «вход — выход» нейрона Маккаллока — Питтса описывается булевой функцией. Нетрудно показать, что существует много булевых функций, которые не могут вычисляться нейронами Маккаллока — Питтса (например, $f(0, 0) = f(1,1) = 1$; $f(0,1) = f(1,0) = 0$). Если же допустить, что наши формальные нейроны имеют взаимодействие афферентов, то по *любой* данной булевой функции можно построить формальный нейрон, который ее вычисляет.

Имеются основания считать, что точное расположение и последовательность соединений дендритов и тел клеток, а также взаимодействие афферентов, которое определяется их расположением, задают ту функцию, которую в зависимости от порога вычисляет каждая клетка, когда получаемые ею сигналы имеют определенную силу и длительность. Каждому нейрофизиологу-экспериментатору хорошо известно, что, к сожалению, нет никаких оснований считать, что что-либо из этого точно определяется нашими генами, адаптацией или обучением.

Начнем с деталей синапсов наших 10^{10} нейронов, которые при фиксированных условиях окружающей среды вряд ли определяются нашими генами. Нейроанатомы вообще считают, что достаточно большое число аксонов обычно растет. Различие в структурах оценивают, как правило, не менее чем одним процентом. Соседние сходные столбики клеток в коре мозга никогда не имеют полного анатомического сходства и, по-видимому, разумно предполагать, что хотя их синапсы в общем определены, детали в значительной

*) Названа по имени Джорджа Буля, который впервые в 1840 г. изучал применение функций этого вида в математической логике.

мере остаются на долю случая. Даже если бы синапсы были точно определены в деталях, случайная гибель хотя бы одного нейрона в минуту вскоре привела бы к их дезорганизации. Миелинизация *) тончайших аксонов хотя бы во внешних слоях коры мозга продолжается по крайней мере до 50-летнего возраста. Более того, мозг обнаруживает явные пульсации, связанные с дыханием, биением сердца, а также высокой частотой, недавно приписываемые ритмическим сокращениям глии. Живой мозг, конечно, не имеет фиксированной и жесткой геометрии.

Если мы обратимся к другим параметрам, влияющим на нервную деятельность, то заметим, что многие из обычных химических изменений действуют на все компоненты сходным образом. Например, гомеостатические механизмы **) стараются поддерживать pH мозга ***) приблизительно на уровне 7,2. Если это значение повышается до 7,4, то порог нейронов, тела и аксона уменьшается приблизительно на 50% нормальной величины, и при таком высоком pH появляются судороги кистей рук и ступней ног. Когда pH падает до 7,0, порог повышается на 100%. Это немного больше, чем те неопасные для организма изменения, которые появляются вследствие повышенной гипервентиляции при усиленном дыхании, и ни одно из них не мешает водолазу выполнять сложные задания. Обезболивание во время хирургической операции понижает pH мозга приблизительно до 6,9, но дыхание продолжается автоматически.

Наконец, пороги и сила импульсов чувствительны к температуре. Однако известно, что человеческий мозг работает, хотя и не слишком хорошо, при температуре 42°С, довольно близкой к смертельной температуре. При температуре ниже 26° С (тéла, а не помещения) нервные ткани млекопитающих становятся

---

*) *Миелинизация* — это образование «изолирующей оболочки» вокруг аксона нейрона. Клетки, имеющие такую оболочку, называются *глиальными*.

**) См. также § 4.3.

***) Величина pH является характеристикой концентрации ионов водорода. (*Прим. перев.*)

невозбудимыми, но при температуре на несколько градусов выше этой больные, которых охлаждают для операций на сердце, еще в состоянии думать.

Давно известно, что вспышки высокой частоты в нервных клетках и их аксонах изменяют время их восстановления, порог и силу импульсов. Более того, известно, что во время восстановления их градиенты напряженности действуют на другие нейроны, находящиеся в их окрестностях. Эти эффекты значительны, и их легко обнаружить, когда много компонент возбуждается одновременно. Они должны встречаться до некоторой степени и в нормальных условиях и их можно отнести за счет больших флуктуаций порогов, которые мы обнаруживаем в них. Но используются ли эти возбуждения, составляющие, скажем, 10% от порогового значения, при передаче сигналов, или они представляют собой только шум, это еще не установлено.

Надо признать возможность многих локальных случайных изменений в силу причин, подобных тем, что изложены выше, и из-за многих незначительных случайностей. Наконец, так как каждая точка переключения представляет собой небольшую область высокого характерного сопротивления, которая и должна работать при температуре тела, то ее порог должен меняться.

В любом случае адекватная нейрологическая модель должна быть построена так, чтобы она правильно функционировала, несмотря на локальные случайные изменения.

При изучении адекватной нейрологической модели мы должны учесть следующие два нарушения в модели. Первое нарушение состоит в гибели нейронов: часто во время болезни эти нейроны испускают длинные серии импульсов, когда их не должно быть, или после гибели не испускают импульсов, хотя в нормальном состоянии эти нейроны должны были испускать их. Второе нарушение состоит в рассеянном распределении во времени и пространстве импульсов, спонтанно возникающих в аксонах, или в виде невозможности их распространения.

§ 3.2. Многократные схемы фон Неймана

Познакомившись с содержанием предыдущего параграфа, мы вплотную подошли к проблеме: каким образом модульная сеть может надежно функционировать, несмотря на неправильную работу отдельных нейронов? Необходимость решения этой проблемы подтверждается следующим простым подсчетом. Рассмотрим цепочку, состоящую из n модулей, и предположим, что вероятность сбоя *) каждого нейрона равна p. Тогда вероятность того, что выход цепочки является правильным, как нетрудно подсчитать **), равна $(1-p)^n$. Если n взять достаточно большим ***), то даже при малом p величина $(1-p)^n$ может принять значение 1/2. И нас, конечно, не устраивает, если выход с одинаковой вероятностью может быть как правильным, так и ошибочным! Мы снова убеждаемся, что при рассмотрении модульных сетей (лучшим образом приближающихся к нашей модели мозга), в которых при работе нейронов допускаются сбои,

*) Под сбоем здесь понимается выдача нейроном неправильного ответа, т. е. когда нейрон вместо 0 выдает 1 или вместо 1 выдает 0. (*Прим. ред.*)

**) На самом деле этот подсчет неверен. Производя подсчет по формуле, предложенной автором, можно прийти к следующему парадоксу. Пусть, для определенности, выход элемента двоичный {0, 1} и вероятность его сбоя $p=1/2$. Соединим n таких элементов в последовательную цепочку. Тогда, согласно приведенному выше подсчету, вероятность правильного выхода равна $p=1/2^n$. При достаточно большом n вероятность правильного выхода может быть сделана сколь угодно малой. Вероятность того, что выход неверен, равна $1-p=1-1/2^n$ и в этом случае может быть сколь угодно близкой к единице. Но это означает, что, получив на выходе 0, надо рассматривать его как 1 (с вероятностью $1-1/2^n$) и 1 на выходе — как 0. Таким образом, соединив последовательно достаточно много элементов, работающих ненадежно, мы могли бы получить выход с любой заданной степенью надежности. Интуитивно ясно, что если только вероятность правильной работы элемента $0<p<1$, то при увеличении длины последовательной цепочки вероятность ее правильного выхода стремится к $1/2$. Приведенная автором формула в действительности дает вероятность того, что все элементы цепочки сработали правильно. (*Прим. ред.*)

***) Если p равно 0,01 (1 процент), то n примерно равно 70, т. е. не слишком велико для мозга, состоящего из 10^{10} элементов.

необходимо исправлять некоторые виды ошибок. По-видимому, модель такого типа была впервые рассмотрена Дж. фон Нейманом в его статье «Вероятностная логика и синтез надежных организмов из ненадежных компонент» *).

Фон Нейман рассматривает два способа построения надежных автоматов из ненадежных компонент. В каждом случае он использует только один тип компонент. Здесь мы обсудим тот способ, когда основным компонентом является модуль «штрих Шеффера» **) $(a|b)$. Он имеет два входа a и b, и $(a|b)(t+1)=0$ в том и только в том случае, если $a(t)=1$ и $b(t)=1$. То есть $(a|b)$ эквивалентно *«не a или не b»*.

Каждому такому модулю сопоставляется вероятность p ошибки, которая не зависит от его входов. Для уменьшения эффекта этих ошибок фон Нейман использует следующий метод, известный как метод «кратных линий»:

а) каждая линия автомата заменяется на пучок из n линий;

б) каждый модуль заменяется на n модулей;

в) каждый пучок связывается с восстанавливающим органом.

Сигналы, передаваемые по каждому пучку ***), кодируются следующим образом: каждая линия передает 0 или 1. Пусть уровень возбуждения (т. е. число единиц) в некотором пучке равен k. Выберем $\Delta<1/2$ и если $k\geqslant(1-\Delta)n$, то сигнал принимаем равным 1; в случае $k<\Delta n$ сигнал принимаем равным 0, а при любом промежуточном значении k сигнал считается

*) Von Neumann J., Probabilistic logics and the synthesis of reliable organisms from unreliable components; «Automata Studies», ed. Shannon C., McCarthy J., Princeton University Press, 1956, 43—98. [Русский перевод: Нейман Дж., Вероятностная логика и синтез надежных организмов из ненадежных компонент. В сб. «Автоматы», ИЛ, 1956.]

**) Модуль «штрих Шеффера» является *универсальным*, т. е. поведение любого конечного автомата может быть реализовано, в смысле теоремы 1.3.1, в некоторой сети из таких модулей.

***) В этой книге такие сигналы называются еще *кратными* сигналами. (*Прим. ред.*)

ошибочным. Такое Δ назовем *критическим уровнем*, т. е. он фиксирован как основа для сравнения.

Восстанавливающий орган работает следующим образом: для любого кратного сигнала, у которого уровень возбуждения k превосходит некоторую критическую величину k_c, восстанавливающий орган повышает уровень возбуждения; напротив, для любого сигнала, у которого уровень возбуждения меньше k_c, восстанавливающий орган уменьшает уровень возбуждения. Функция восстанавливающего органа иллюстрируется на рис. 3.3.

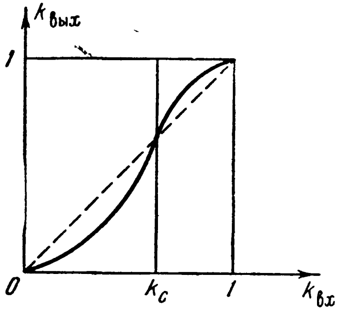


Рис. 3.3. Функция восстанавливающего органа $k_{вх}$ = доле возбужденных входов; $k_{вых}$ = доле возбужденных выходов.

Чтобы уменьшить возможность систематического накопления ошибок, все пучки между модулями перемешиваются случайным образом. Например, автомат, показанный на рис. 3.4, заменяется автоматом, показанным на рис. 3.5. Сам восстанавливающий орган также состоит из $2n$ модулей «штрих Шеффера», работающих со сбоями.

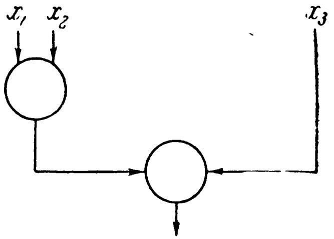


Рис. 3.4. «Исходная» сеть.

Предположив, что ошибки не могут возникать за счет расщепления линий, фон Нейман показывает, что при $p < 0{,}0107$ вероятность неправильного срабатывания автомата равна

$$p_n \sim an^{-\frac{1}{2}} \cdot 10^{-bn}, \qquad (3.2.1)$$

где a и b являются константами, и, таким образом, при достаточно больших n может быть очень малой.

Однако эти константы являются такими, что для получения приемлемых p_n, при вероятности неправильной

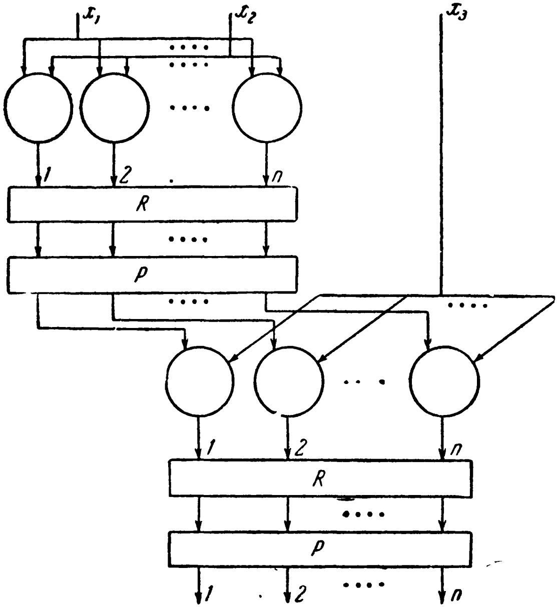


Рис. 3.5. Избыточная сеть, полученная из сети рис. 3.4 с помощью кратных линий фон Неймана. R — восстанавливающий орган, P — смеситель.

работы одной компоненты $p = 1/200$, требуется брать очень большие значения n:

n	p_n
1 000	$2{,}7 \cdot 10^{-2}$
2 000	$2{,}6 \cdot 10^{-3}$
3 000	$2{,}5 \cdot 10^{-4}$
5 000	$4 \cdot 10^{-6}$
10 000	$1{,}6 \cdot 10^{-10}$
20 000	$2{,}8 \cdot 10^{-10}$
25 000	$1{,}2 \cdot 10^{-23}$

Таким образом, решение фон Неймана имеет следующий недостаток: для построения надежного автомата требуется чрезвычайно большая избыточность. Кроме того, результирующий автомат имеет такую структуру, которая не соответствует ни одной известной биологической структуре. Для модели мозга это серьезный недостаток.

Сам фон Нейман был не очень доволен своим решением и был уверен, что ошибку следует рассматривать при помощи термодинамических методов, как это делается с информацией в работах Л. Сциларда и К. Шеннона.

Учитывая это, мы в § 3.3 сосредоточим внимание на теории Шеннона, после чего сможем изучить более приемлемое решение проблемы надежности модели мозга.

§ 3.3. Шенноновская теория связи *)

Предметом этой теории является проблема передачи сообщений при наличии статистических, возмущающих или других видов «шумов». Мы будем рассматривать системы связи, имеющие вид, показанный на рис. 3.6.

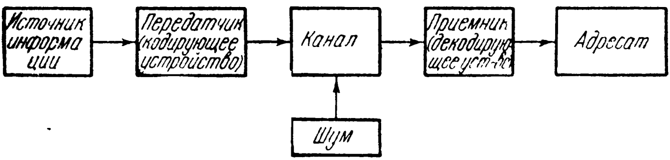


Рис. 3.6. Схематическое изображение системы связи.

Источник информации создает сообщение, которое должно быть передано на приемный конец.

Передатчик (*кодирующее устройство*) перерабатывает некоторым образом сообщение в сигналы, которые можно передавать по данному каналу.

*) Часто называется *теорией информации*.

Канал — это среда, используемая для передачи сигнала от передатчика к приемнику. При передаче сигнал может быть искажен помехами («шумом»).

Приемник (*декодирующее устройство*) — это обратный преобразователь (восстановитель сообщения).

Адресат — это лицо или аппарат, для которого предназначено сообщение.

Здесь мы изучаем только «дискретный» случай, т. е. тот случай, при котором как сообщения, так и сигналы рассматриваются как последовательности символов. Основная цель этого параграфа состоит в доказательстве теоремы Шеннона. Эта теорема показывает, при каких условиях мы можем выбрать наш передатчик так, чтобы, несмотря на наличие шума, принятые сообщения почти всегда совпадали с переданными.

Уивер (Weaver)в своем превосходном и стимулирующем обсуждении расчленяет проблемы связи на три класса:

Технические проблемы. Как точно можно передавать символы связи?

Семантические проблемы. Как точно передаваемые символы выражают желаемый смысл?

Проблемы эффективности. Как эффективно принятая мысль влияет на поведение в желаемом направлении?

Теория Шеннона имеет дело только с техническими проблемами. Для этой цели сообщения будут характеризоваться вероятностями их появления без учета ценности и смысла. Возможны возражения, что исследование, при котором не рассматриваются две последние проблемы, является нереальным и даже ошибочным. На это я могу ответить, что изучение технической проблемы является только первым этапом и он далеко не последний. Мы не можем передавать нашу мысль с любой точностью по системе, которая передает символы не точно. Поэтому сначала целесообразней овладеть технической проблемой.

Довольно общее изложение теории Шеннона влечет за собой использование таких понятий, как теория меры, эргодическая теория и мартингалы. Такое изложение осуществлено А. Я. Хинчиным*), но оно является слишком сложным и длинным, чтобы включить его в эту книгу. Вместо этого, следуя Шеннону, мы рассматриваем источник как марковский процесс, предполагаем, что канал не имеет памяти, и опускаем детали в наших доказательствах**).

В п. 3.3.1 мы дадим определение математической меры информации, которая зависит только от вероятностей создаваемых сообщений.

В п. 3.3.2 мы дадим математическую модель источника и канала. С источником мы свяжем энтропию H, которая понимается как среднее количество информации, содержащейся в каждом создаваемом источником символе. Для обозначения этого количества Шеннон использует слово «энтропия», так как формула для него аналогична формуле для энтропии в статистической механике. О том, насколько далеко простирается эта формальная аналогия, можно спорить. Во всяком случае, читатель предупрежден против неосторожного пользования этой аналогией.

Символы, последовательно создаваемые источником, передаются по каналу, в котором они независимо искажаются шумом. Так как выход канала в данный момент зависит от входа в этот же момент и не зависит от входов в предыдущие моменты, то мы называем его каналом без памяти. Искажения, очевидно,

---

*) Хинчин А. Я., Понятие энтропии в теории вероятностей. Успехи матем. наук, т. 8, вып. 3, 1953.

Хинчин А. Я., Об основных теоремах теории информации. Успехи матем. наук, т. 11, вып. 1, 1956, стр. 17—75. См. также A. Feinstein, Foundations of information theory, McGraw-Hill Book Company, Inc., New York, 1958. [Русский перевод: А. Файнстейн, Основы теории информации, М., ИЛ, 1960.]

) Этот параграф в большой степени написан по работе: C. E. Shannon, The mathematical theory of communication, Bell System Techn. J. **27, № 3 (1948), 379—423; **27**, № 4 (1948), 623—656. [Русский перевод: Шеннон К. Э., Работы по теории информации и кибернетике, ИЛ, 1963, стр. 243—332.] Эта работа была также опубликована в книге: C. E. Shannon and W. Weaver, The mathematical theory of communication, The University of Illinois Press, Urbana, Ill., 1949, в которой также содержится работа Уивера «Recent contributions to the mathematical theory of communication».

приводят к потере информации в канале. В п. 3.3.3 мы изучим среднее количество информации, приходящееся на один канальный символ, называемое *скоростью передачи* R. Очевидно, что $R \leqslant H$, так как мы не можем принять больше информации, чем посылается. Величина $H - R$ называется *ненадежностью.* Мы можем рассматривать ее как количество неопределенности, которое содержится в передаваемом сообщении. Для различных источников R различна. *Пропускную способность* C *канала* определим как максимальное значение R, которое можно получить при подходящем выборе источника. Таким образом, C равно максимальному количеству информации на один символ, на которое мы можем рассчитывать при передаче по нашему каналу.

Можно удивляться тому, что мы дадим определение пропускной способности C для канала с шумом, хотя возможно, мы не будем знать, какая информация в этих случаях передается. Однако ясно, что передача информации в избыточной форме может понизить вероятность ошибки. Например, если сообщение повторять много раз, то, осуществляя статистическое изучение различных принятых вариантов сообщения, вероятность ошибки можно сделать очень малой. Вместе с тем, можно ожидать, что для приближения к нулю вероятности ошибки потребуется очень большая избыточность кодирования и, следовательно, скорость передачи должна стремиться к нулю. На самом деле это не так.

Фактически введенная выше пропускная способность может быть определена многими способами. Вероятно, что для передачи информации со скоростью C и *малой частотой ошибок, т. е. ненадежностью*, требуется надлежащее кодирование. Такое высказывание является неверным для любой скорости, большей чем C (рис. 3.7). Эти результаты являются основным оправданием для определения C и составляют *замечательную* фундаментальную теорему Шеннона для дискретного канала с шумом.

Пусть дискретный канал имеет пропускную способность C, а дискретный источник — энтропию H. Если

$H \leqslant C$, то существует такой способ кодирования, что сообщения источника могут быть переданы по каналу с произвольно малой частотой ошибок (или со сколь угодно малой ненадежностью). Если $H > C$, то можно применить такое кодирование источника, что ненадежность будет меньше чем $H - C + \varepsilon$, где ε сколь угодно мало. Не существует способа кодирования, обеспечивающего ненадежность, меньшую чем $H - C$.

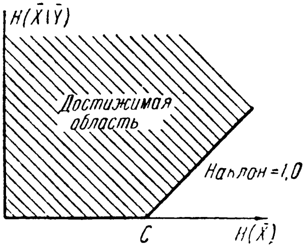


Рис. 3.7. Графическое представление основной теоремы Шеннона для дискретного канала с шумом. Любая точка выше жирной линии в заштрихованной области может быть получена; точки же ниже линии получены быть не могут. Точки самой линии, вообще говоря, получены быть не могут.

Доказательство этой основной теоремы приводится в п. 3.3.4. Я не могу не обратить внимание на то, что эта теорема поистине замечательна. Возможно, с первого взгляда она и не выглядит такой, но тогда читателю надо внимательно прочесть весь текст еще раз.

Точные определения и доказательство теоремы Шеннона требуют знания некоторых понятий теории вероятностей, может быть, незнакомых некоторым читателям. Те читатели, которые плохо знакомы с теорией вероятностей, но не желают вникать в детали, могут непосредственно переходить к п. 3.3.5. Читатель, который желает убедиться в истинности теоремы Шеннона, должен читать подряд.

3.3.1. Мера информации

Пусть x — сообщение, имеющее вероятность появления p. Мы хотим найти числовую меру $I(x)$ *количества уменьшения неопределенности* (=*количеству прироста информации*), когда принято сообщение x. Так как мы ограничились рассмотрением только технической проблемы, то будем предполагать, что $I(x)$ является непрерывной функцией $A(p)$ только от p.

Чем меньше вероятность сообщения x, тем больше уменьшается неопределенность при получении его; поэтому мы требуем, чтобы $A(p)$ увеличилось при уменьшении p. Если x_1 и x_2 являются независимыми сообщениями с вероятностями p_1 и p_2 соответственно, то мы хотим, чтобы информация, содержащаяся в x_1 и x_2 (совместное получение x_1 и x_2 является событием с вероятностью p_1p_2), равнялась сумме количеств информации, содержащихся отдельно в x_1 и x_2. Это приводит нас к следующему.

Мы постулируем, что для сообщения x, появляющегося с вероятностью p, информация $I(x)=A(p)$ удовлетворяет требованиям:

(а) $A(p)$ является непрерывной убывающей функцией от p, $0<p\leqslant 1$.

(б) $A(p_1\cdot p_2)=A(p_1)+A(p_2)$.

Теорема 3.3.1. *Если* $A(p)$ *удовлетворяет* (а) *и* (б), *то*

$$A(p)=-K\log_2 p,$$

где K — *некоторая положительная константа.*

Доказательство. Двумя индукциями (одна по числителю, а другая — по знаменателю) показывается, что $A(t^r)=rA(t)$ всякий раз, когда r является любым положительным рациональным числом. Непрерывность позволяет получить тот же результат, когда r является любым положительным действительным числом. Полагая $t=1/2$, получаем

$$A(2^{-r})=rA(2^{-1}),$$

или другими словами

$$A(p)=A(^1/_2)(-\log_2 p),$$

что и требовалось доказать.

Мы принимаем обычное соглашение, что $K=1$ (выбор единицы измерения) *), так что исход

---

*) Полученная таким образом единица называется «бит» — сокращение от «binary digit», поскольку 0 или 1, встречающиеся с равной вероятностью, а именно $^1/_2$, несут один бит информации.

случайного бросания нефальшивой монеты дает одну единицу информации ($-\log_2 1/2 = 1$).

Определение 3.3.1. *Информация*, содержащаяся в сообщении x, появляющемся с вероятностью p, равна

$$I(x) = -\log_2 p.$$

Если $p=0$, то мы полагаем, что $I(x)=0$. Предположим, что мы теперь имеем совокупность сообщений

$$\overline{X} = \begin{pmatrix} x_1 & \ldots & x_n \\ p_1 & \ldots & p_n \end{pmatrix},$$

где i-е сообщение x_i имеет вероятность p_i $\left(p_i \geqslant 0, \sum p_i = 1\right)$. Из определения 3.3.1 следует, что среднее количество информации, полученной приемником от источника,

$$H(\overline{X})_{\text{def}} = -\sum_i p_i \log_2 p_i,$$

где $0 \log_2 0$ полагается равным 0. Величину $H(\overline{X})$ назовем *энтропией* этой совокупности.

Мы стремились к тому, чтобы приведенное выше определение было правдоподобным. Однако настоящим оправданием для этого определения является то, что функция $-\sum p_i \log_2 p_i$ появляется во многих основных задачах кодирования и связи.

3.3.2. Модель источника и канала

Мы договорились характеризовать сообщение вероятностью его появления. Поэтому мы должны характеризовать наш источник распределением вероятностей сообщений, которые он создает. Воспользуемся следующей характеристикой марковских процессов*): источник имеет конечное число возможных «состояний» $S_1, S_2, \ldots, S_n$. Кроме того, имеется совокупность вероятностей переходов p_{ij}, т. е. вероятностей того, что система, находящаяся в состоянии S_i, перейдет затем

*) Это основное понятие теории вероятностей носит имя русского математика А. А. Маркова, который впервые ввел его.

в состояние S_j. Чтобы превратить этот марковский процесс в источник сообщений, нужно только предположить, что при каждом переходе из одного состояния в другое создается одна буква. Состояния будут соответствовать «остатку влияния» предыдущих букв*).

Среди всевозможных дискретных марковских процессов имеется одна группа процессов с особо важными для теории связи свойствами. Этот специальный

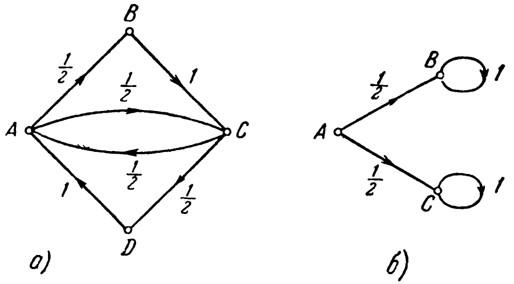


Рис. 3.8. Два марковских источника: *а*) эргодический; *б*) неэргодический.

класс состоит из «эргодических» процессов, и соответствующие источники мы будем называть *эргодическими источниками*. Хотя строгое определение эргодического процесса несколько сложно, общая идея проста. В случае эргодического процесса каждая бесконечная последовательность, создаваемая этим процессом, имеет одни и те же статистические свойства. Поэтому частота букв, пар букв и т. п., полученных на начальных кусках этой последовательности, с увеличением их длины будет стремиться к определенным пределам, не зависящим от этих начальных кусков. В действительности это верно не для всякой последовательности, но множество последовательностей, для которых это неверно, имеет вероятность, равную нулю. Грубо говоря, свойство эргодичности означает статистическую однородность.

*) Ради развлечения читатель может попытаться определить «марковский конечный автомат» (см. § 1.3).

На рис. 3.8 показаны графы двух источников, первый из которых эргодический, а второй — неэргодический (вершины соответствуют состояниям схемы, а вероятности переходов указаны около соответствующих дуг), причем предполагается, что различные переходы задаются различными буквами.

Теперь свяжем с нашим дискретным источником марковского типа понятие энтропии. Для каждого возможного состояния S_i имеется некоторое множество вероятностей $p_i(j)$ выработки различных возможных символов j. Следовательно, для каждого состояния S_i существует энтропия

$$H_i = -\sum_j p_i(j) \log_2 p_i(j).$$

Энтропия источника H определяется как среднее значение величин H_i, которые осреднены в соответствии с вероятностями p_i осуществления соответствующих событий

$$H = \sum_i P_i H_i = -\sum_{i,\ j} P_i p_i(j) \log_2 p_i(j).$$

Эта величина называется *энтропией источника на символ сообщения*. Ясно, что пару «источник S — кодирующее устройство» теперь можно рассматривать как новый источник S'. Процесс кодирования просто заключается в замене одной последовательности символов на другую — закодированную последовательность. Следовательно, множество сообщений, созданных новым составным источником S', имеет в точности то же самое распределение вероятностей, что и множество сообщений, созданных первоначальным источником S. Из этого следует, что энтропия источника S равна энтропии источника S', т. е. в процессе кодирования H не меняется. Этот факт будет использован в п. 3.3.4.

Если источник является эргодическим (с этого момента мы будем рассматривать только эргодические источники), то число тех случаев, когда данный путь $(i,\ j)$ на графе переходов встречается в последовательности *большой* длины N, приблизительно пропорционально вероятности пребывания в состоянии i, скажем P_i, и с последующим прохождением этого пути, т. е. приблизительно равно $P_i p_{ij} N$. Если N достаточно велико, то вероятность ошибки, не превосходящая $\pm\delta'$, в этом случае меньше чем ε, так что все значения, за исключением множества малой вероятности, заключены в пределах

$$(P_i p_{ij} \pm \delta') N.$$

Следовательно, почти все последовательности имеют вероятность

$$p = \prod_{i,\,j} p_{ij}^{(P_i p_{ij} \pm \delta')\,N}$$

и $\log_2 p/N$ ограничен соотношениями

$$\frac{\log_2 p}{N} = \sum_{i,\,j} (P_i p_{ij} \pm \delta') \log_2 p_{ij}$$

или

$$\left| \frac{\log_2 p}{N} - \sum_{i,\,j} P_i p_{ij} \log_2 p_{ij} \right| < \delta,$$

где δ стремится к нулю при $\delta' \to 0$. Таким образом, доказана следующая теорема.

Теорема 3.3.2. *Для любых заданных* $\varepsilon > 0$ *и* $\delta > 0$ *можно найти такое* N_0, *что последовательности любой длины* $N \geqslant N_0$ *распадаются на два класса*:

а) множество последовательностей, суммарная вероятность которых меньше чем ε;

б) остаток, все члены которого обладают вероятностями, удовлетворяющими неравенству

$$\left| \frac{\log_2 p^{-1}}{N} - H \right| < \delta.$$

Другими словами, *почти достоверно, что* $\log_2 p^{-1}/N$ *весьма близко к* H, *когда* N *велико.*

Перейдем теперь к характеристике нашего канала. Здесь рассматривается только самая простая модель канала, а именно модель *дискретного канала без памяти*, в котором последовательные символы искажаются шумом независимо. Поэтому такой канал однозначно описывается множеством переходных вероятностей $p_i(j)$ *).

Если канал с шумом питается от источника, то взаимодействуют два статистических процесса: источник и шум. Имеется также несколько энтропий, которые можно вычислить. Во-первых, имеется энтропия (на символ) $H(\overline{X})$ входа в канал. Во-вторых, имеется энтропия выхода канала, обозначаемая через $H(\overline{Y})$. Таким образом, если s_i есть вероятность получения i-го символа на выходе канала, то

$H(\overline{Y})=$ *среднему количеству информации, извлекаемой из каждого получаемого на выходе символа* $= -\sum_i s_i \log_2 s_i$.

Наконец, имеется две условные энтропии $H(\overline{Y}|\overline{X})$ и $H(\overline{X}|\overline{Y})$, первая из которых является энтропией выхода канала при условии, что вход известен, а вторая — энтропией входа канала, когда известен его выход.

Пусть r_i есть вероятность появления i-го символа на входе канала, s_i — вероятность появления этого символа на выходе; $p_i(j)$ есть вероятность того, что если на вход канала послан символ i, то на его выходе будет принят символ j; наконец, $q_i(j)$ — вероятность того, что если на выходе принят символ i, то на его вход был послан символ j. Тогда

$H(\overline{X}|\overline{Y})=$ *энтропии на символ входной последовательности* $\overline{X}$ *при условии, что принята выходная последовательность* $\overline{Y}=$ *среднему количеству информации на символ неизвестного входа, когда известен выход* $=$

$$= \sum_i s_i \left[-\sum_j q_i(j) \log_2 q_i(j)\right] = -\sum_{i,\,j} s_i q_i(j) \log_2 q_i(j);$$

---

*) Переходную вероятность не надо смешивать с вероятностью $p_i(j)$ источника, где $p_i(j)$ есть вероятность того, что если послан символ i, то будет принят символ j.

аналогично имеем

$$H(\bar{X}\,|\,\bar{Y}) = -\sum_{i,\,j} r_i p_i(j) \log_2 p_i(j).$$

3.3.3. Ненадежность и пропускная способность канала

При наличии шума, вообще говоря, невозможно восстановить исходное сообщение с *полной достоверностью*, какую бы операцию мы ни применяли к принятому сигналу. Но тем не менее мы сейчас рассмотрим некоторые способы передачи информации, которые являются *оптимальными* в отношении борьбы с шумом.

Предположим, что мы имеем два символа 0 и 1, и ведем передачу со скоростью 1000 символов в секунду с вероятностями $p_0 = p_1 = 1/2$. Таким образом, наш источник создает информацию со скоростью 1000 *бит* в секунду. В процессе передачи шум вносит ошибки, так что в среднем один из 100 символов принимается неправильно (0 вместо 1 или 1 вместо 0). Какова скорость передачи информации? Ясно, что она меньше 1000 *бит* в секунду, так как около 1% принятых сигналов неправильны.

Сначала хочется сказать, что эта скорость составляет 990 *бит* в секунду, т. е. хочется просто вычесть среднее число ошибок. Однако это не совсем правильно, так как при этом не учитывается, что адресат не знает, где именно произошла ошибка. Можно рассмотреть крайний случай и предположить шум настолько большим, что принятые символы полностью не зависят от переданных. Вероятность принять как 1, так и 0 равна 1/2, какой бы символ ни был передан (1 или 0). Тогда благодаря одной только случайности около половины принятых символов будут правильными, и нам пришлось бы считать, что система способна передавать 500 *бит* в секунду. Но в действительности никакой передачи информации в этом случае нет. Можно было бы получить столь же «хорошую» передачу, обойдясь вообще без канала и подбрасывая монету на месте приема.

Очевидно, истинная поправка к количеству переданной информации равна той части информации, которая отсутствует в принятом сигнале, или иначе, той неопределенности относительно переданного сигнала, которая имеет место, когда известен принятый сигнал. Исходя из приведенного нами обсуждения понятия энтропии как меры неопределенности, представляется разумным использовать условную энтропию $H(\overline{X}|\overline{Y})$ сообщения (при известном сигнале) как меру этой недостающей части информации. Следуя этой идее, можно получить *скорость действительной передачи информации* R, вычитая из скорости создания информации (т. е. энтропии источника) среднюю скорость условной энтропии

$$R = H(\overline{X}) - H(\overline{X}\,|\,\overline{Y}).$$

Условная энтропия $H(\overline{X}|\overline{Y})$ для удобства будет называться *ненадежностью*. Она является мерой средней неопределенности принятого сигнала.

Если в рассмотренном выше примере принят символ 0, то апостериорная вероятность того, что был передан символ 0, равна 0,99, а того, что была передана единица, — 0,01. Если же была принята единица, то эти цифры поменяются местами. Следовательно,

$$H(\overline{X}\,|\,\overline{Y}) = -[0{,}99 \log_2 0{,}99 + 0{,}01 \log_2 0{,}01] = 0{,}081 \text{ *бит/символ*},$$

или 81 *бит* в секунду. Мы можем сказать, что система передает информацию со скоростью 1000—81 = =919 *бит* в секунду. В том крайнем случае, когда при передаче символа 0 равновероятен прием как 0, так и 1 (аналогично для символа 1), апостериорные вероятности равны 1/2, и тогда

$$H(\overline{X}\,|\,\overline{Y}) = -[1/2 \log_2 1/2 + 1/2 \log_2 1/2] = 1 \text{ *бит/символ*},$$

или 1000 *бит* в секунду. В этом случае скорость передачи, как и следовало ожидать, равна 0.

Пропускная способность C канала с шумом должна совпадать с максимально возможной скоростью передачи. Определим поэтому *пропускную способность канала* как

$$C = \max(H(\bar{X}) - H(\bar{X} \mid \bar{Y})),$$

где максимум берется по всем возможным источникам информации, используемым в качестве входа канала. В дальнейшем будет показано, что $H(\bar{X})$ зависит только от источника информации, тогда как $H(\bar{X} \mid \bar{Y})$ зависит как от источника информации, так и от канала. Если зафиксировать канал и варьировать источник информации S, то число

$R(S) = H(\bar{X}) - H(\bar{X} \mid \bar{Y}) =$

$=$ *среднему количеству информации на один символ, переданный по каналу, когда информация создана источником* S,

изменяется вместе с S. Величина C равна максимально возможному значению скорости $R(S)$ и, таким образом, зависит только от канала.

Прежде чем переходить к доказательству теоремы Шеннона, я советую читателю еще раз внимательно прочесть материал п. 3.3.1, где обсуждался смысл понятия пропускной способности канала.

3.3.4. Основная теорема Шеннона для дискретного канала с шумом

Теперь можно вновь сформулировать и доказать теорему Шеннона.

Теорема 3.3.3. *Пусть дискретный канал имеет пропускную способность C, а дискретный источник — энтропию H. Если $H \leqslant C$, то существует такой способ кодирования, что сообщения источника могут быть переданы по каналу с произвольно малой частотой ошибок (или со сколь угодно малой ненадежностью). Если $H > C$, то можно применить такое кодирование источника, что ненадежность будет меньше чем $H - C + \varepsilon$,* где ε *сколь угодно мало. Не существует способа кодирования, обеспечивающего ненадежность, меньшую чем $H - C$.*

Доказательство*). Метод доказательства первой части этой теоремы состоит не в указании способа кодирования, имеющего желаемые свойства, а в доказательстве того, что *искомый код должен существовать в определенной группе кодов.* Будем усреднять по этой группе частоту ошибок и покажем, что полученное среднее может быть сделано меньше чем ε. Но если среднее некоторого множества чисел меньше чем ε, то в этом множестве должно существовать по крайней мере одно число, меньшее ε. Это и даст желаемый результат.

Пропускная способность C канала с шумом была определена как

$$C \doteq \max(H(\overline{X}) - H(\overline{X} \mid \overline{Y})),$$

где $\overline{X}$ — значение входа, а $\overline{Y}$ — значение выхода канала. Максимум берется по всем источникам, которые могут быть использованы на входе такого канала. Следовательно, если рассматриваемый источник S имеет энтропию $H<C$, то можно взять другой источник S_0, для которого $H(\overline{X}) - H(\overline{X} \mid \overline{Y}) > H$. Пусть S_0 находится на входе канала. Рассмотрим возможные передаваемые и принимаемые последовательности длины T. Из теоремы 3.3.2 следует, что T можно выбрать настолько большим, что выполняются следующие четыре условия.

(а) Передаваемые последовательности распадаются на два класса: высоковероятная группа, содержащая примерно $2^{TH(\overline{X})}$ членов, и остальные последовательности с малой суммарной вероятностью.

(б) Аналогично имеется высоковероятное множество принимаемых последовательностей, содержащее примерно $2^{TH(\overline{Y})}$ членов, и маловероятное множество оставшихся последовательностей.

---

*) Полное доказательство этой теоремы иным методом можно найти в книге: А. Файнстейн, Основы теории информации, ИЛ, 1960, а в более общей постановке и с использованием предложенного здесь метода — в статье: Добрушин Р. Л., Общая формулировка основной теоремы Шеннона в теории информации. Успехи матем. наук, т. 14, вып. 6, 1959, стр. 3—104. *(Прим. перев.)*

(в) Каждая из высоковероятных выходных последовательностей может быть создана примерно $2^{TH(\overline{X}|\overline{Y})}$ входными последовательностями. Суммарная вероятность всех других случаев мала.

(г) Каждая из высоковероятных входных последовательностей может создать примерно $2^{TH(\overline{Y}|\overline{X})}$

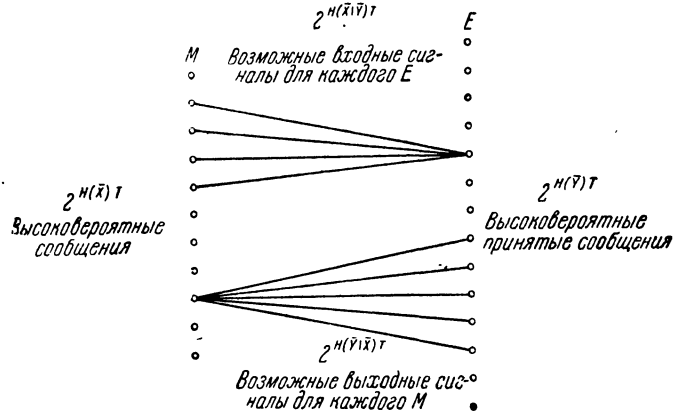


Рис. 3.9. Схематическое представление соотношений между входами и выходами в канале.

выходных последовательностей. Суммарная вероятность всех других исходов мала.

Все «ε» и «δ», подразумеваемые в словах «малый» и «примерно», стремятся к нулю, когда T возрастает. Сказанное выше иллюстрирует рис. 3.9, где входные последовательности показаны точками слева, а выходные последовательности — точками справа. Верхний веер сходящихся линий изображает ситуацию для типичного выхода. Нижний веер изображает ситуацию для типичного входа. В обоих случаях не учитываются «маловероятные» множества.

Вернемся снова к рассмотрению источника S, создающего информацию со скоростью $H<C$. За время T этот источник будет создавать 2^{TH} высоковероятных сообщений. Мы хотим связать их с некоторым

выбранным множеством возможных входных сигналов канала таким образом, чтобы получить малую частоту ошибок. Будем устанавливать эту связь всеми возможными способами (используя, однако, только ту высоковероятную группу входных сигналов, которая определяется источником S) и усредним частоту ошибок по этому большому классу возможных систем кодирования. Это равносильно вычислению частоты ошибок при случайном связывании сообщений и входных сигналов канала длительности T.

Предположим, что наблюдается некоторый выходной сигнал y. Какова вероятность того, что в множество возможных причин получения этого y войдет более чем одно сообщение от источника S. Имеется 2^{TH} сообщений, распределенных случайно по $2^{TH(\overline{X})}$ точкам. Вероятность того, что некоторая конкретная точка будет сообщением, равна поэтому $2^{T(H-H(\overline{X}))}$.

Вероятность того, что никакая другая точка веера (кроме действительного исходного сообщения) не будет сообщением, равна

$$P=\left[1-2^{T(H-H(\overline{X}))}\right]^{2^{TH(\overline{X}|\overline{Y})}-1}.$$

Но $H<H(\overline{X})-H(\overline{X}|\overline{Y})$, по предположению S_0, так что

$$H-H(\overline{X})=-H(\overline{X}|\overline{Y})-k,$$

где $k>0$. Следовательно,

$$P=\left[1-2^{-TH(\overline{X}|\overline{Y})-Tk}\right]^{2^{TH(\overline{X}|\overline{Y})}-1}$$

стремится (при $T\to\infty$) (с учетом асимптотического равенства $(1-a)^b\sim 1-ab$) к $1-2^{-Tk}$. Таким образом, вероятность ошибки стремится к нулю, и первая часть теоремы доказана.

Вторую часть теоремы легко доказать, исходя из того, что можно просто передавать C бит на символ от источника, совсем пренебрегая остатком создаваемой информации. В приемнике эта неучитываемая часть создает ненадежность $H(\overline{X})-C$, а передаваемая часть может добавить только ε.

Последнее утверждение теоремы является простым следствием нашего определения C. Предположим, что можно закодировать некоторый источник с энтропией $H(\overline{X}) = C + a$ таким образом, чтобы получить ненадежность $H(\overline{X} | \overline{Y}) = a - \varepsilon$, где $\varepsilon > 0$. Тогда

$$H(\overline{X}) - H(\overline{X} | \overline{Y}) = C + \varepsilon,$$

где $\varepsilon > 0$. Это противоречит определению C как максимума $H(\overline{X}) - H(\overline{X} | \overline{Y})$, что и требовалось доказать.

В действительности здесь доказано больше, чем утверждалось в теореме. Если среднее значение множества положительных чисел заключено между нулем и ε, то только часть из них, доля которой не превышает $\sqrt{\varepsilon}$, может быть больше ε. Так как ε произвольно мало, то можно сказать, что почти все системы кодирования предложенного выше типа сколь угодно близки к идеальной.

3.3.5. Кодирование

Теорема Шеннона показывает, что существуют коды для передачи информации со сколь угодно малой вероятностью ошибки, но не указывает, каким образом строить такие коды. Имеется много работ, посвященных фактическому построению таких кодов. Мы рассмотрим тот случай, когда сообщения являются последовательностями нулей и единиц.

Предположим, что эти символы порождаются независимо, т. е. 0 и 1 появляются с одной и той же вероятностью. Тогда скорость создания информации на один символ равна

$$-[1/2 \log_2 1/2 + 1/2 \log_2 1/2] = 1,$$

т. е. 1 *бит* на символ. Предположим теперь, что мы желаем передавать эту последовательность по каналу с пропускной способностью $C < 1$. В этом случае мы должны уменьшить скорость передачи так, чтобы она была меньше C. Тогда кодирование и декодирование заключается в преобразовании одних последовательностей двоичных символов в другие последовательности тех же символов. При этом наши 2^k возможных передаваемых последовательностей сообщений длины k (т. е. каждая последовательность состоит из k символов) мы заменяем на $2^k=2^{nR}$ сигнальных последовательностей длины n, где $R=k/n<C$, так что величина k/n, являющаяся *скоростью передачи на один сигнальный символ*, в самом деле меньше C.

Процесс кодирования заключается в следующем (рис. 3.10). Входные символы объединяются в *блоки*

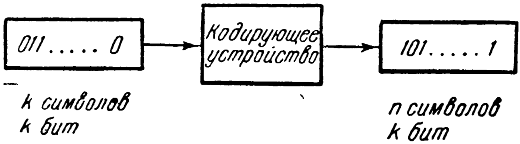


Рис. 3.10. Блочное кодирование. Замечания: 1) блок длины k задерживается так, чтобы он поступил в кодирующее устройство в тот момент, когда его покидает блок длины n; 2) при кодировании осуществляется задержка на k тактов (по входной шкале), так как блок длины n может быть получен лишь после того, как будет закодирован весь блок длины k.

по k символов, а затем каждый блок преобразуется в блок сигнальных символов длины n. Под *временной избыточностью* мы просто понимаем *увеличение времени*, необходимое для передачи соответствующего кодированного сообщения. Блоки длины n выбираются такими, чтобы при декодировании можно было исправлять допустимые ошибки. Один из таких способов заключается в передаче блоков длины n, состоящих из k символов исходного сообщения и $n-k$ *проверочных символов*, которые впоследствии позволяют обнаруживать и исправлять возникшие в процессе передачи ошибки. Любой код, в котором блоки из k символов кодируются блоками из n символов, называется (n, k)-кодом. Скорость передачи (n, k)-кодом на один сигнальный символ равна k/n.

В качестве примеров кодов с исправлением ошибок приведем предложенные Р. В. Хэммингом три кода с исправлением одиночных ошибок, скорости передачи которых соответственно равны 1/3, 2/5 и 4/7. Пусть a_i $(i=1, 2, \ldots, n)$ и b_j $(j=1, \ldots, n)$ — переданные и принятые сигнальные символы.

x	$\rightarrow$	a	$\rightarrow$	b	$\rightarrow$	y
сообщение		кодированное сообщение		принятое сообщение		декодированное сообщение

I. Кодирующие функции — **Декодирующая функция**

$$a_1 = x_1,\quad a_2 = x_1,\quad a_3 = x_1 \qquad\qquad y = b_1b_2 + b_2b_3 + b_3b_1\ {}^{*)}$$

(3,1)-код.

II. Кодирующие функции — **Декодирующие функции**

$$\begin{aligned} a_1 &= x_1 \\ a_2 &= x_1 \\ a_3 &= x_1 + x_2 \\ a_4 &= x_2 \\ a_5 &= x_2 \end{aligned} \qquad\qquad \begin{aligned} y_1 &= b_1 + (b_1 + b_2)\,b' \\ y_2 &= b_5 + (b_5 + b_4)\,b', \end{aligned}$$

где

$$b' = b_1 + b_3 + b_5.$$

(5,2)-код.

III. Кодирующие функции — **Декодирующие функции**

$$\begin{aligned} a_1 &= x_1 \\ a_2 &= x_2 \\ a_3 &= x_1 + x_3 + x_4 \\ a_4 &= x_1 + x_2 + x_4 \\ a_5 &= x_2 + x_3 + x_4 \\ a_6 &= x_3 \\ a_7 &= x_4 \end{aligned} \qquad\qquad \begin{aligned} y_1 &= b_1 + b'' \\ y_2 &= b_2 + b'' \\ y_3 &= b_3 + b'' \\ y_4 &= b_4 + b'', \end{aligned}$$

где

$$b'' = (b_1 + b_2 + b_4 + b_7)\times \times(b_2 + b_5 + b_6 + b_7)\times \times(b_1 + b_3 + b_6 + b_7)$$

(7,4)-код.

---

*) Числа всюду двоичные: 0 или 1. Умножение обычное

$$0\cdot 1 = 1\cdot 0 = 0\cdot 0 = 0;\quad 1\cdot 1 = 1.$$

Сложение является сложением по модулю 2 (т. е. сумма равна остатку от деления обычной суммы на 2)

$$1 + 0 = 0 + 1 = 1;\quad 0 + 0 = 1 + 1 = 0.$$

Способ построения кода Хэмминга очень прост. Проиллюстрируем его на примере построения (3,1)-кода. Имеем 0 → 000, 1 → 111. Сообщения 000 и 111 рассматриваются как точки (0, 0, 0) и (1, 1, 1) трехмерного пространства. Они являются двумя вершинами единичного куба (рис. 3.11). Если при передаче

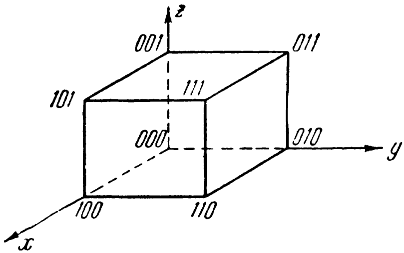


Рис. 3.11. Трехмерный куб.

произойдет не более одной ошибки, то принятая последовательность будет соответствовать либо переданной вершине, либо одной из смежных с ней вершин. Таким образом,

$$0\,0\,0 \to 0\,0\,0,\ 1\,0\,0,\ 0\,1\,0 \ \text{или}\ 0\,0\,1,$$

$$1\,1\,1 \to 1\,1\,1,\ 0\,1\,1,\ 1\,0\,1 \ \text{или}\ 1\,1\,0.$$

Заметим теперь, что обе эти «смежные системы» не пересекаются *). Это и позволяет исправлять ошибки в коде Хэмминга. Декодирующая функция является логическим выражением для «стягивания обратно» вершин куба к «переданной вершине». Коды (5,2) и (7, 4) построены аналогично, но на единичных кубах в пяти- и семимерном пространствах соответственно.

*) Если мы допускаем, что при передаче сообщения произошла не более чем одна ошибка, то принятые сообщения (000, 010, 100, 001) должны расшифровываться как 000, а (111, 110, 011, 101) как 111. Если бы множества принимаемых сообщений, соответствующие 000 и 111, имели непустое пересечение, то такая расшифровка была бы невозможна, ибо, приняв сообщение, принадлежащее пересечению, мы не смогли бы установить, какому передаваемому сообщению оно соответствует. (*Прим. ред.*)

§ 3.4. Теория связи и автоматы

Переведем временну́ю избыточность, обычно связанную с передающим каналом, в избыточность оборудования. Вместо одного канала, по которому последовательно передаются n символов блока, мы будем использовать n каналов для одновременной передачи n символов так, что каждый канал используется для передачи только одного символа блока (рис. 3.12).

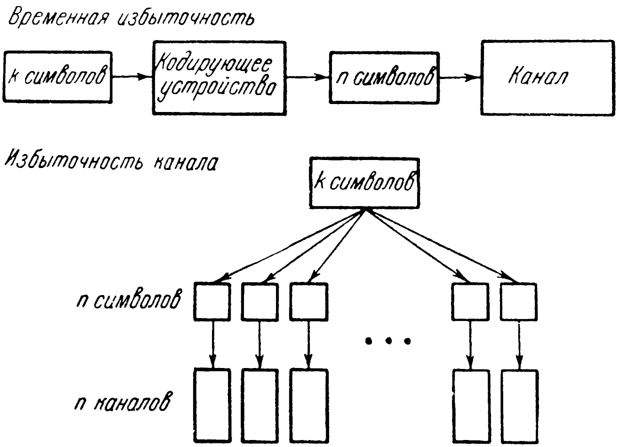


Рис. 3.12. Временная избыточность в сравнении с канальной избыточностью.

При таком подходе вместо скорости передачи k/n рассматривается коэффициент эффективности (обозначаемый через R), равный отношению K/N, где K — число элементов, которое должно быть использовано, если бы не было шума (т. е. число элементов в неизбыточной системе), а N — число элементов, которое на самом деле используется (т. е. число элементов в избыточной системе).

Вернемся теперь к используемой фон Нейманом *покомпонентной избыточности* в вычислительных системах (являющихся *непростыми* системами связи), чтобы сделать их более надежными (см. § 3.2). Мы видели, что при коэффициенте эффективности $R=1/3n$ (т. е. каждая компонента заменяется на n компонент плюс $(2n)$-компонентный восстанавливающий орган) вероятность неправильной работы системы

$$p_n \sim an^{-1/2} \cdot 10^{-bn},$$

где a и b — подходящие константы. Это выражение мы можем переписать следующим образом:

$$p_n = dR^{1/2} \cdot 2^{-c/R}, \tag{3.2.1}$$

где c и d — константы. Графически это изображено на рис. 3.13. По фон Нейману можно получить сколь угодно малое p_n только за счет стремления к нулю величины R, тогда как Шенноном для систем связи установлено существование кодов со сколь угодно малым p_n всякий раз, когда $R<C$. Это означает, что избыточность автоматов по фон Нейману не позволяет получить безошибочное поведение при условии, что пропускная способность больше нуля. Причиной этого является то, что замещение каждого модуля совокупностью модулей и каждого соединения пучком соответствует использованию $(n, 1)$-кода. Только использование (n, k)-кодов, в которых отношение $R=k/n$ остается равным константе при росте k и n, позволяет получить положительные скорости надежной передачи информации и, следовательно, систему с положительной пропускной способностью. Поэтому естественно рассмотреть возможность построения различных избыточных автоматов, которые эффективно используют (n, k)-коды и выполняют вычисления с положительной пропускной способностью.

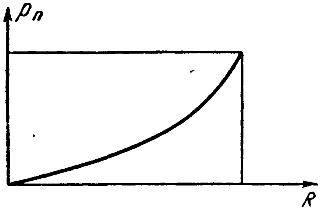


Рис. 3.13. Вероятность ошибки как функция коэффициента эффективности для многократной схемы фон Неймана (ср. рис. 3.7).

Один из подходов к решению этой проблемы заключается в том, чтобы попытаться применить имеющиеся методы теории кодирования, которые, по существу, подходящим образом «подгоняют» источник к каналу, а затем декодируют выходы канала с шумом таким способом, который уже нами обсуждался. Заметим, что Шеннон в своей теории связи предполагал, что кодирующее и декодирующее устройства работают безошибочно. В рассматриваемом им случае это было оправданным, ибо количество шума, возникающего, например, при посылке и приеме радиосигналов, как правило, незначительно по сравнению с тем шумом, который возникает в используемом канале. Однако в нашем случае предположение о безошибочном кодировании и декодировании в некоторой степени является сомнительным, ибо при кодировании и декодировании производятся вычисления и слишком трудно рассчитывать, что элементы кодирующих и декодирующих сетей имеют меньше шума по сравнению с тем, который возникает при основных вычислениях.

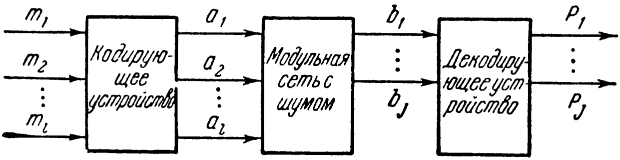


Рис. 3.14. Вычислительная система.

Но, давайте, на минуту предположим, что шум в кодирующем и декодирующем устройствах является незначительным и рассмотрим выводы, вытекающие из работы П. Элайса *).

Модульная сеть (рис. 3.14) должна реализовывать множество событий $P_1, \ldots, P_j$, которые являются логическими функциями входов $m_1, \ldots, m_i$. Рассматриваемая сеть состоит из модулей, реализующих булевы функции от одной или двух переменных с небольшими, но не нулевыми вероятностями ошибок. Эти

*) P. Elias, Computation in the presence of noise, IBM Journ. Res. and Develop. 2, № 4 (1958), 346—353.

вероятности статистически не зависят от общего состояния всей сети и других неисправностей, т. е. мы допускаем, что ошибки такие же, как у фон Неймана. Однако мы не будем следовать за фон Нейманом в смысле исправления ошибок посредством реконструкции самой сети. Скорее избыточность, получаемая в сигнальной последовательности, является такой же, как при передаче сообщений. Последовательности входных букв длины k кодируются последовательностями сигнальных букв длины n, которые затем поступают в модульную сеть. Выход этой сети декодирует блоки длины n в блоки длины k, которые, как мы надеемся, соответствуют реализуемым событиям. Важно заметить, однако, что эта сеть не имеет памяти и *что переработка осуществляется символ на символ.*

Элайс требовал, чтобы входы $m_1, \ldots, m_i$ кодировались отдельно и чтобы декодирование при отсутствии шума было одно-однозначным. Эти предположения гарантируют, что вычисление предписанного множества событий имеет место только в данной модульной сети. В противном случае этот анализ был бы не очень осмысленным.

Элайс получил следующие результаты. Из 16 возможных булевых функций с двумя входами m_1 и m_2 только 8 являются такими, что для них можно использовать (n, k)-коды для вычислений с положительными скоростями и сколь угодно малой частотой ошибок. Из этих 8 функций 6 функций являются мало интересными. К ним относятся тождественно истинная и тождественно ложная функции m_1, $\overline{m}_1$, m_2 и $\overline{m}_2$. Оставшиеся две функции m_1+m_2 и $m_1 \equiv m_2$ являются такими, что для них можно использовать коды Хэмминга. Однако приведенное выше множество из 8 функций не является «универсальным» (т. е. сети, которые реализуют произвольные булевы функции, вообще говоря, нельзя построить только из элементов этого множества).

Для остальных 8 функций $m_1 \& m_2$, $m_1 \& \overline{m}_2$, $\overline{m}_1 \& m_2$, $\overline{m}_1 \& \overline{m}_2$, $m_1 \vee m_2$, $m_1 \vee \overline{m}_2$, $\overline{m}_1 \vee m_2$ и $\overline{m}_1 \vee \overline{m}_2$ можно применять только $(n, 1)$-коды. Так как в этом множестве содержатся обе универсальные функции (т. е. функция «штрих Шеффера» $\overline{m}_1 \vee \overline{m}_2$ и функция $\overline{m}_1 \& \overline{m}_2$), то это означает, что хотя сети, которые реализуют произвольные регулярные события, могут быть построены только из элементов этого множества, однако их нельзя использовать для реализации этих событий с положительной скоростью передачи информации, если требуется, чтобы частота ошибок была сколь угодно малой. Этот результат по существу аналогичен результату фон Неймана, и Элайс высказал гипотезу, что пропускная способность вычислений при надежной обработке в произвольных модульных сетях будет нулевой, если аналогичные ограничения будут наложены на вспомогательное оборудование. Тем не менее в следующем параграфе мы приведем совершенно другую кодирующую схему, которая позволяет получить ненулевую пропускную способность.

Заметим, что если ослабить наше требование, заключающееся в том, чтобы в декодирующем устройстве не производились вычисления, то очень просто (вернее, совсем тривиально!) можно получить ненулевую пропускную способность, например, следующим образом (рис. 3.15). Рассмотрим сеть как канал передачи

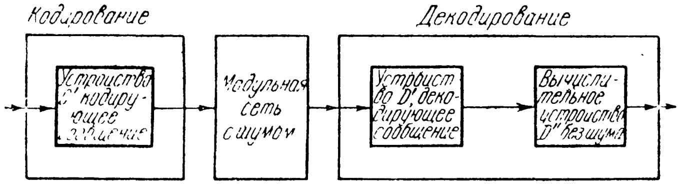


Рис. 3.15. Тривиально! Абсурдно! Так нельзя строить надежную вычислительную систему.

с очень большим шумом и определим ее пропускную способность обычным способом (заметьте, что это определение относится к прохождению информации через канал). Тогда для любой скорости, меньшей этой пропускной способности и любой заданной вероятности ошибки выберем кодирующее устройство C' и декодирующее устройство D' для передачи (через всю систему с большой точностью и без вычислений) исходного сообщения, равного аргументам функции. Затем этот выход поступает на вычислительное устройство без шума D'', реализующее ту же функцию, которую должна была выполнять первоначальная модульная сеть с шумом. Если теперь кодирующее устройство C' поместить перед нашей сетью с шумом, а декодирующее устройство $D' \to D''$ — после этой сети, то мы получим безошибочное вычисление с любой скоростью, меньшей чем пропускная способность. Тем не менее это, конечно, не является настоящим решением нашей проблемы.

§ 3.5. Теория надежных автоматов Винограда — Кована

В этом параграфе мы обсудим успешное применение (n, k)-кодов, изученных нами в § 3.3.5. Изучаемая нами работа издана монографией С. Винограда и Дж. Д. Кована «Надежные вычисления при наличии шумов» *).

В § 3.4 мы видели, что при параллельном вычислении (канальной избыточности) можно использовать временну́ю избыточность и так же как в теории связи для все более и более надежной передачи при заданной скорости необходимо увеличивать задержку, так и для получения надежных вычислений при заданном коэффициенте эффективности необходимо иметь все больше и больше параллельных вычислений.

Предположим, что нам дан автомат A_i, который вычисляет некоторую дефинитную функцию **). Чтобы убедиться в том, как много требуется параллельных вычислений, рассмотрим (рис. 3.16, *а*) m копий этого автомата $A_1, A_2, \ldots, A_m$, каждая из которых вычисляет одну и ту же функцию, но не обязательно от одних и тех же входов. Вместо надежного автомата, заменяющего автомат A_1, мы построим новый автомат

---

*) S. Winograd, J. D. Cowan, Reliable computation in the presence of noise, MIT, Press, Cambridge, Mass., 1963. [Русский перевод этой книги выпускается издательством «Наука».]

**) То есть функцию, которая может быть вычислена подходящей модульной сетью без циклов (см. ниже). (*Прим. ред.*)

A (рис. 3.16, *б*), заменяющий систему параллельно работающих автоматов $A_1, \ldots, A_m$. Коэффициент

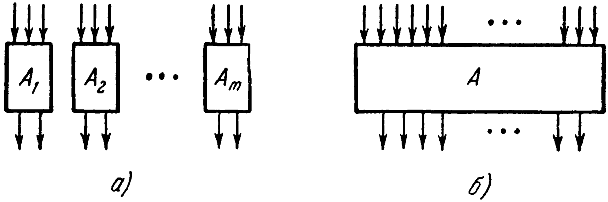


Рис. 3.16. *а*) m неизбыточных параллельно работающих автоматов; *б*) один неизбыточный автомат, заменяющий $A_1, A_2, \ldots, A_m$.

эффективности, связанный с этим замещением, будет равен

$$R = \frac{K}{N} = \frac{\textit{число элементов в } m \textit{ экземплярах автомата } A_1}{\textit{число элементов в } A}.$$

Перед изложением основного результата теории Винограда — Кована мы должны ввести некоторые вспомогательные понятия. Мы будем считать, что выходами сети A являются линии $O_1, O_2, \ldots, O_m$. Ясно, что любая ошибка, возникшая в элементе, выходом которого является линия O_j, не может быть исправлена. Множество элементов, с которых берутся выходы сети, назовем *последним рядом*. Поэтому Виноград и Кован говорят, что автомат может иметь *сколь угодно большую надежность*, если его вероятность ошибки может быть сделана сколь угодно близкой к вероятности ошибки в его последнем ряду.

Если будем рассматривать модуль как *канал связи*, то мы можем определить *пропускную способность* этого модуля точно так же, как в п. 3.3.3. При вычислениях вместо *пропускной способности канала* мы будем говорить о *пропускной способности модуля*.

Дефинитным событием называется такое событие, которое можно реализовать (см. §§ 1.7—1.8) модульной сетью, не содержащей петель. Таким образом, результат вычисления дефинитного события может зависеть только от конечного числа предыдущих значений входов. *Недефинитным событием* называется такое событие, которое может зависеть от сколь угодно большого числа предыдущих значений входов. Примером такого события, например, может служить событие «на вход когда-то в прошлом поступала 1».

Теперь мы можем сформулировать основной результат, полученный Виноградом и Кованом (теорема 8.1 в их монографии).

Теорема 3.5.1. *Пусть дано некоторое дефинитное событие. Если в нашем распоряжении имеются все возможные формальные нейроны с взаимодействием афферентов (и таким образом все возможные булевы функции), каждый из которых имеет одну и ту же пропускную способность C, то для любого $R<C$ можно построить «произвольно надежный» автомат A, который вычисляет данное дефинитное событие с коэффициентом эффективности, большим или равным R.*

По-видимому, требование, чтобы A вычислял *дефинитное* событие, является необходимым. М. О. Рабин показал, что не существует автомата с положительными вероятностями перехода из *любого* внутреннего состояния в любое другое, который мог бы вычислять регулярное недефинитное событие. Он, по существу, показывает, что если некоторая ошибка когда-нибудь возникнет в одной из петель обратной связи, то, вообще говоря, все последующие выходы будут ошибочными. Однако ввиду того, что мы интересуемся поведением автомата (или мозга) только в течение конечного промежутка времени (т. е. в период жизни), то наш автомат вычисляет только дефинитные события и, таким образом, ограничения в приведенной выше теореме являются скорее всего ограничениями математической, а не физической природы.

Доказательство теоремы сводится к построению соответствующего автомата и основано на существовании удовлетворительных кодов для каналов связи. Таким образом, существует определенная взаимосвязь между избыточными схемами, которые используются в доказательстве, и кодированием в линиях связи. Здесь мы не будем приводить полного доказательства, а лишь бегло проследим за основными моментами. Наша исходная модульная сеть состоит из m параллельно расположенных неизбыточных автоматов A_1, A_2, ..., A_m. Ради простоты мы будем предполагать, что эта исходная сеть (т. е. неизбыточная сеть с шумом, с которой мы начинаем) имеет очень специальный вид, указанный на рис. 3.17. Это есть модульная сеть «ширины» l и «глубины» a, модули которой вычисляют различные функции f_{jk} (или реализуют события η_{jk}), где $j=1, \ldots, a$; $k=1, \ldots, l$. Мы предполагаем, что в среднем каждый модуль имеет s входов и один выход, который может ветвиться на s линий. Входами модуля $(r+1)$-го ряда являются либо входы сети, либо выходы модулей r-го ряда.

Мы хотим заменить эту совокупность модулей такой схемой, в которой ошибки исправляются сразу же, как только они возникают, а вычисления осуществляются сразу же после исправления ошибок. Другими словами, мы хотим избежать, насколько это возможно, любого распространения ошибок. С этой целью мы будем сочетать кодирование, вычисление и декодирование таким образом, чтобы модули любого заданного ряда декодировали свои входы, вычисляли соответствующие функции, а затем кодировали свои выходы. Заметим, что мы таким образом переносим нашу избыточность кодирования на структуру автомата и переходим от ряда к ряду: выходы r-го ряда могут подключаться только к модулям $(r+1)$-го ряда.

Прежде чем описывать подробности кодирования, рассмотрим надежность исходной модульной сети, изображенной на рис. 3.17. Будем требовать, чтобы вероятность P_{a-1} того, что имеется по крайней мере одна ошибка в первых $a-1$ рядах, была не более δ. Пусть ε — вероятность появления ошибки в произвольном модуле сети, и пусть все ошибки являются независимыми. Тогда вероятность безошибочной работы модулей из первых $(a-1)$ рядов равна

$$1-P_{a-1}=(1-\varepsilon)^{l(a-1)}. \qquad (3.5.1)$$

Поэтому мы требуем, чтобы

$$\delta \geqslant P_{a-1} \geqslant 1-(1-\varepsilon)^{l(a-1)} \sim l(a-1)\varepsilon,$$

так как $(1-\alpha)^{\beta} \sim 1-\alpha\beta$, т. е. приблизительно

$$\varepsilon \leqslant \frac{\delta}{l(a-1)}.$$

При небольшом фиксированном δ это означает, что ε должно уменьшаться с ростом l и a, т. е. *без подходящего кодирования* вероятность ошибки в *исходной*

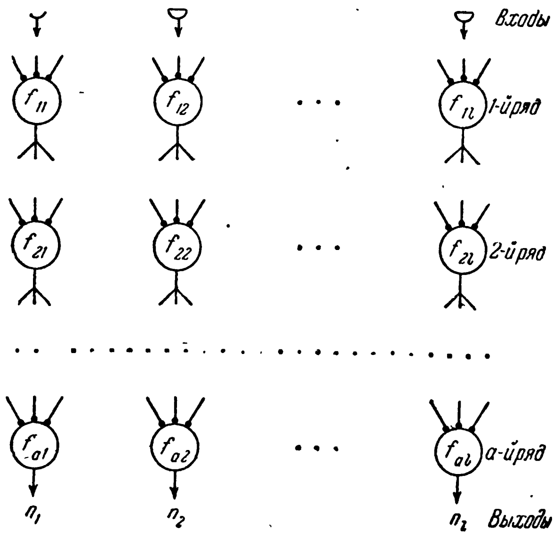


Рис. 3.17. «Исходная» модульная сеть.

сети должна уменьшаться с увеличением размеров сети; в противном случае сеть не может оставаться «произвольно надежной». Чтобы получить сети с произвольной надежностью, состоящие из модулей с фиксированными вероятностями ошибок, мы заменим исходную сеть на другую сеть, в которой каждый модуль вычисляет новую функцию f'_{mn}. В этом избыточном автомате мы намерены использовать (n, k)-код.

Пусть $n_1(t)$, $n_2(t)$, ..., $n_l(t)$ — дефинитные события, которые реализуются в сети (рис. 3.17), вообще говоря, с задержкой a. Опишем теперь, каким образом эти события могут быть вычислены с произвольной

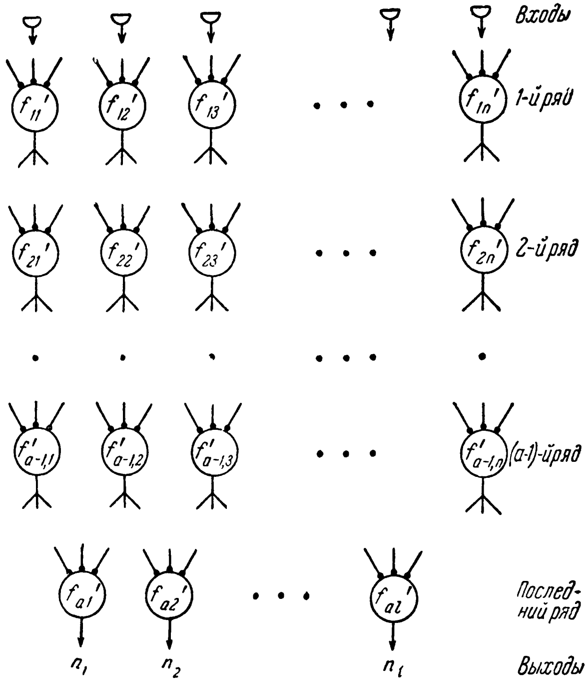


Рис. 3.18. Избыточная сеть, полученная из сети рис. 3.17 по схеме Винограда — Кована.

надежностью в такой модульной сети, которая указана на рис. 3.18. В этой сети каждый модуль вычисляет булеву функцию с единичной задержкой. Каждая из функций f'_{mn} является составной частью (n, k)-кода. Предполагается, что этот код является кодом с исправлением ошибок, который уже обсуждался.

Пусть e_1, e_2, ..., e_n — кодирующие функции и d_1, ..., d_k — декодирующие функции. Тогда назначение элементов первого ряда (с функциями f'_{1k}) заключается в вычислении некоторой подходящей совокупности функций f_{ij} с последующим кодированием

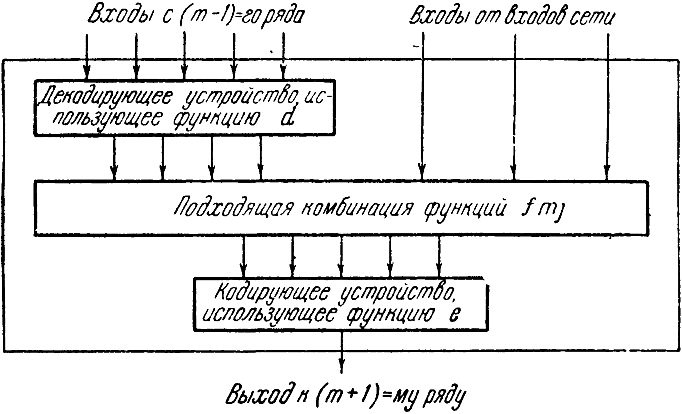


Рис. 3.19. Граф-схема, изображающая f'_{mn}.

результатов посредством кодирующих функций и выдачей их на свои выходы. Назначение элементов любого более позднего ряда (с функциями f'_{ik}, $i>1$) состоит в декодировании выходов, поступающих с предшествующего ряда (посредством декодирующих функций), в вычислении подходящей совокупности функций f_{ij} и кодировании результатов для выходов (с использованием кодирующих функций). В последнем ряду кодирование не производится. На рис. 3.19 мы приводим граф-схему f'_{mn}.

Линии между модулями можно рассматривать как каналы связи с шумом, перед которыми мы хотим кодировать, а после которых мы хотим декодировать. Если некоторая линия пучка поступает с выхода некоторого модуля, то необходимо подходящее декодирование, если же линия поступает со входа сети, то декодирование не требуется. Вычисление функции f_{mn}

для этих значений (мы опускаем детали соответствующих соединений) дает величину, которая затем кодируется посредством кодирующей функции e. В результате получается f'_{mn}. Таким образом, мы получаем избыточную модульную сеть (рис. 3.18).

Любой ряд этого избыточного автомата декодирует выходы предыдущего ряда; при этом предполагается, что «правильный» выход предыдущего ряда имеет некоторую вероятность ошибки P_e. Используя работу Шеннона, Виноград и Кован показывают, что эта ошибка

$$P_e = 2^{k-nC}.$$

Отсюда и из равенства (3.5.1) следует, что вероятность безошибочной работы первых $a-1$ рядов *избыточной* сети равна

$$1 - P_{a-1} = \left(1 - 2^{k-nC}\right)^{a-1},$$

то есть

$$P_{a-1} \sim (a-1)\, 2^{k-nC}.$$

Таким образом, при фиксированном a вероятность ошибки P_{a-1} может быть сделана сколь угодно малой за счет увеличения k и n, если только сохраняется неравенство $k/n < C$. Это означает, что автомат с произвольно высокой надежностью (в смысле нашего определения) может быть построен из модулей с фиксированной пропускной способностью C. Число модулей в избыточном автомате является таким, что коэффициент эффективности превосходит величину R. На этом мы закончим набросок доказательства теоремы 3.5.1.

Заметим, что число входов каждого элемента избыточной схемы увеличивается, если при построении таких схем мы будем использовать все более и более сложные коды. Эта особенность схем иногда может быть препятствием для использования таких кодов, ибо в этом случае требуется, чтобы все элементы имели одно и то же количество шума. С другой стороны, из нашего построения следует, что если мы наложим ограничения сверху на число входов элементов, то единственный способ получения автомата с произвольно высокой надежностью состоит в неограниченном уменьшении величины R. Таким образом, любая схема автомата с произвольно высокой надежностью и ненулевой ограниченной величиной R должна *бесспорно* состоять из сколь угодно сложных элементов.

С увеличением сложности всей сети может случиться, что два элемента не будут соединены в имеющейся схеме автомата, хотя на самом деле они должны быть соединены. В связи с этим возникает следующий вопрос: с какой вероятностью p мы можем пренебречь тем обстоятельством, что некоторые элементы *соединены*, хотя этого и не должно быть, а другие элементы *не соединены*, хотя и предполагалось, что они будут соединенными?

Ответ на этот вопрос дан Виноградом и Кованом. Его можно высказать в виде следующей теоремы (эту теорему мы здесь не будем доказывать; в упомянутой монографии это теорема 9.1).

Теорема 3.5.2. *Пусть A есть проект надежного автомата с коэффициентом эффективности R. Тогда, если в этом проекте вероятность неправильных соединений пропорциональна величине $(C-K)$, то такой автомат будет еще иметь произвольную надежность.*

Грубо говоря, эта теорема указывает на то, что при построении автомата доля допускаемых ошибок должна быть пропорциональна разности между пропускной способностью компонент и коэффициентом эффективности. В определенном смысле эту дополнительную информацию, имеющуюся в нашем распоряжении, мы используем для исправления ошибочных соединений.

Подводя итоги, Кован говорит следующее:

«Мы показали, что к задаче построения надежных сетей из надежных элементов можно применить теорию связи. Полученный надежно функционируемый автомат мало похож на исходный автомат, функционирующий ненадежно.

Надо отметить следующие факты:

(a) Многократное разнообразие сети... каждый элемент может вычислять многие функции и любую исходную функцию вычисляют многие элементы. Для достаточно большого n каждый элемент может вычислять произвольную совокупность исходных функций.

(b) Однородность сети... она непосредственно следует из предыдущего замечания.

(c) Эффективность сети: по сравнению с сетями фон Неймана необходима значительно меньшая избыточность элементов».

Прежде чем оставить этот интересный вопрос об обеспечении высокой надежности функционирования мозга, мы сделаем несколько замечаний, хотя все значение их читатель сможет оценить только после чтения главы 4. Кроме кодирования, обеспечивающего надежность собственно нервной системы, мы имеем несколько встроенных устройств, повышающих нашу надежность (но не обязательно доверие к нам!):

1. Наша нервная система связана с эффекторами. Благодаря самой инерции, скажем, движущейся руки незначительные временные сбои в работе нейронов «сглаживаются» и ошибки в нашем поведении, таким образом, устраняются.

2. Обратная связь (см. § 4.1) с внешним миром дает большие возможности для компенсации некоторых ошибок, которые встречаются в нашем мозгу.

3. Кодирование информации в нервной системе частично происходит в форме частотной модуляции (см. § 4.4) — чем больше интенсивность раздражения или желательной реакции, тем больше частота возбуждения импульсов в соответствующих нервных волокнах. Поэтому отказ некоторых нейронов, составляющих небольшой процент, не может оказать решающего влияния, какое ожидается в чисто цифровой (типа Маккаллока — Питтса) системе, а только слегка искажает интенсивность.

Каково бы ни было значение таких механизмов в улучшении надежности нервной системы, все еще остается необходимость построения теоретических моделей мозга — таких, например, как модель Кована и Винограда, которые устраняли бы проблемы, отмеченные Маккаллоком, Арбибом и Кованом в цитате из § 3.1. Гораздо больше нам надо знать о мозге, и гораздо более сложные модели мозга еще ждут своего построения. Теория Кована — Винограда *) представляет собой важный шаг в этом непрерывном поиске.

---

*) Чтобы получить представление о других теориях функционирования мозга, интересующемуся читателю следует посмотреть раздел о математических теориях в книге: D. A. Sholl, The organization of the cerebral cortex, Methuen & Co., Ltd., London, 1956.

Глава 4

Кибернетика

В 1947 г. Норберт Винер и его коллеги термином «кибернетика» назвали науку, занимающуюся «сравнительным изучением управления и связи в животном и машине». Таким образом, все, что обсуждалось нами до сих пор, может быть объединено под единым заголовком «кибернетика». Тем не менее, настоящая глава будет, в основном, посвящена вопросам, рассматривавшимся Винером в его книге «Кибернетика» *).

§ 4.1. Обратная связь и колебания **)

В § 1.1 мы рассматривали нервную систему человека, в которой были выделены три основные подсистемы (рис. 1.1). Мы сосредоточили внимание на вопросах, касающихся структуры собственно нервной системы (модульные сети нейронов Маккаллока — Питтса, персептрон, теория Кована — Винограда) и рассмотрели вкратце связь зрительных рецепторов с нервной системой у лягушки. Следует отметить, что мы изучали и моделировали изолированный мозг,

---

*) Вышла в свет впервые в 1948 г. [Русский перевод: Н. Винер, Кибернетика, Изд-во «Сов. радио», 1958.] Я, однако, рекомендую второе переработанное издание: The M. I. T. Press, Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1961. [Русский перевод готовится в издательстве «Сов. радио».]

**) Довольно полным учебником по системам с обратными связями является J. C. Gille, M. J. Pélegrin and P. D. Decaulne, Feedback control systems, McGraw-Hill Book Comp., Inc., New York, 1959.

Подборкой хороших статей, написанных для «образованного» начинающего читателя, является книга «Automatic control», Simon and Shuster, Inc., New York, 1955. (Рекомендуем также книгу: А. А. Фельдбаум, А. Д. Дудыкин, А. П. Мановцев, Н. Н. Миролюбов, Теоретические основы связи и управления, Физматгиз, 1963. (*Прим. ред.*))

опустив особенности его взаимодействия с эффекторами. Мы знаем, что конечный автомат превращается в машину Тьюринга, обладающую существенно новыми свойствами (§ 1.7), если ему дать возможность взаимодействовать с окружающей средой, снабдив его лентой и головкой для записи и считывания информации с ленты (см. § 1.7). Поэтому естественно ожидать, что если мы расширим понятие нервной системы, включив взаимодействие эффекторов и рецепторов, мы можем добиться новых успехов.

Учитывая это, обратим теперь внимание на связь эффекторов с рецепторами. Легко заметить, что информация может являться *обратной связью* для собственно нервной системы о том, насколько правильно последняя управляет деятельностью эффекторов.

Поясним это на двух простых примерах:

а) Я хочу взять карандаш и протягиваю к нему руку. Глаза сигнализируют мне, какое расстояние должна пройти рука. В мозг непрерывно поступает информация от рецепторов (глаз), характеризующая положение эффектора (руки). Поэтому мой мозг в состоянии выработать необходимые управляющие сигналы, поступающие затем на мускулы руки и направляющие ее движение так, чтобы уменьшить разницу между действительным и требуемым положением.

б) Когда я иду, я попеременно поднимаю ноги, а затем свободно их опускаю. Я управляю сокращением мышц ноги, меняя положение ног, когда давление ноги на землю превышает некоторое критическое значение. Положение эффектора (ноги) по отношению к земле передается по цепи *обратной связи* в мозг от рецепторов давления ступни. Здесь опять-таки существует разница в положении ноги и поверхности земли, определяющая вид сигналов, посылаемых мозгом мышцам ноги. Существенность этой «обратной связи» от рецепторов давления в ступне хорошо иллюстрируется всем знакомым ощущением: поднявшись после долгого сидения «по-турецки», вы чувствуете себя как бы на булавках и вынуждены двигаться, не используя рассмотренной обратной связи, в результате чего походка становится неуклюжей и плохо координированной.

Отсюда следует вывод: обратная связь от эффекторов к рецепторам играет решающую роль в нашем поведении. Другими словами, понятие *обратной связи* должно занимать существенное место при изучении работы мозга и машины. Мы будем подразумевать в дальнейшем, что в организме или машине присутствует обратная связь, если их деятельность определяется, до некоторой степени, результатом сопоставления действительного состояния с желаемым. В частности, как показывает пример с рукой и карандашом, нас особенно должна интересовать *отрицательная* обратная связь, *уменьшающая* рассогласование между действительным и желаемым положением.

Простым примером отрицательной обратной связи в машинах служит терморегулятор, работа которого заключается в уменьшении разности между действительным состоянием (температура комнаты) и желаемым (температура, установленная на задающей шкале) путем соответствующей корректировки работы источника тепла.

Блок-схема простейшей системы с отрицательной обратной связью представлена на рис. 4.1. На вход

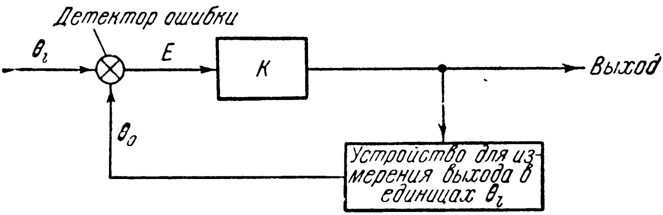


Рис. 4.1. Система с обратной связью.

системы постоянно поступает величина θ_i, определяющая требуемое значение выхода. Рассмотрим теперь действительное значение выхода. Оно измеряется и в виде величины θ_0 подается на детектор ошибки, который определяет разность между требуемым и действительным значением выхода и вычисляет сигнал ошибки

$$E = \theta_i - \theta_0.$$

Именно этот сигнал и управляет, в действительности, системой K, определяя ее выход. Для каждого конкретного случая основной задачей является создание «черного ящика» K, для которого величина E стремилась бы к нулю. Читатель согласится, что наш терморегулятор может быть естественно трактован в терминах рис. 4.1. Система с отрицательной обратной связью, изображенная на рис. 4.1, часто называется *следящей системой* (servomechanism).

Другим примером следящей системы является рулевая машина корабля. Этот пример и лежит в этимологии слова «кибернетика», поскольку χνβερνητηζ — по-гречески «рулевой».

Рассмотрим теперь простейшую математическую модель *линейной* следящей системы, в которой связь между θ_0 и E описывается линейным дифференциальным уравнением с постоянными коэффициентами. Этот пример приводится только для того, чтобы показать, как используется математический аппарат для описания систем с обратной связью. Полученные результаты будут носить качественный характер; ниже они получат интуитивное подтверждение.

Рассмотрим простейшую математическую модель устройства K, описываемого уравнением второго порядка *), где ζ и ω_n — постоянные параметры устройства K,

$$\frac{d^2E}{dt^2}+2\zeta\omega_n\frac{dE}{dt}+\omega_n^2E=\frac{d^2\theta_i}{dt^2}+2\zeta\omega_n\frac{d\theta_i}{dt}.$$

Решим это уравнение относительно $E(t)$, потребовав постоянства выхода, т. е. $\theta_i(t)=k$.

Имеем (см. подробности в любом учебнике по теории автоматического регулирования)**)

$$E(t)=A+Be^{a_1t}+Ce^{a_2t}, \tag{4.1.1}$$

где A, B и C — константы, определяемые через ω_n, ζ и различные значения сигнала в K при $t=0$ (т. е. в ис-

*) Это уравнение весьма просто выводится в любой книге по теории автоматического регулирования.

**) См., например, Айзерман М. А., Теория автоматического регулирования, Изд-во «Наука», 1966. (*Прим. перев.*)

ходный момент); e — основание натуральных логарифмов ($e=2,718281\ldots$), a_1 и a_2 — корни уравнения

$$a^2+2\zeta\omega_n a+\omega_n^2=0.$$

Я рассматриваю данную линейную систему второго порядка не потому, что она имеет особую биологическую подоплеку, а просто для иллюстрации качественной характеристики следящих систем — стабильности. Из физических соображений ζ и ω_n — действительные величины, а следовательно, a_1 и a_2 или обе действительные или комплексно-сопряженные величины. Если a_1 и a_2 обе действительные и отрицательные, то при $t\to\infty$ $e^{a_1t}\to 0$ и $e^{a_2t}\to 0$. Однако если $a_1>0$, то при $t\to\infty\, e^{a_1t}\to\infty$ и в этом случае система нестабильна. Если a_1 и a_2 — комплексно-сопряженные, т. е. $a+ib$ и $a-ib$, тогда выражение $Be^{a_1t}+Ce^{a_2t}$ может быть приведено к виду

$$A+e^{at}(D\cos bt+F\sin bt);$$

его значение стремится к A при $t\to\infty$ и $a<0$; положительное же значение a обусловливает неограниченный рост амплитуды колебаний при $t\to\infty$ (рис. 4.2).

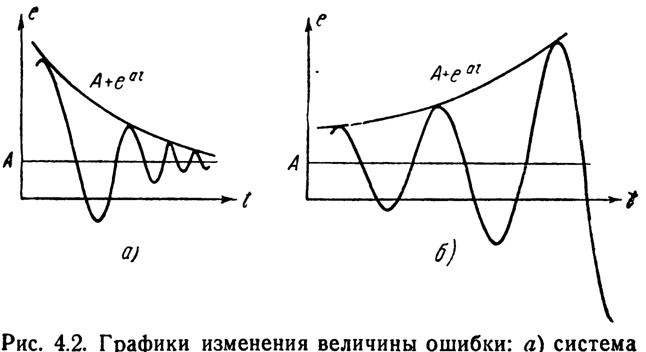


Рис. 4.2. Графики изменения величины ошибки: *а*) система стабильна (*а* отрицательно); *б*) система нестабильна (*а* положительно).

Таким образом, если a_1 или a_2 имеют положительную действительную часть, значение ошибки *неограниченно* *растет*, т. е. система *нестабильна.* Если же и a_1 и a_2 имеют отрицательные действительные части, то значения ошибки ограничены, система *стабильна.*

К выводу о том, что в нестабильной системе могут существовать незатухающие колебания *), можно прийти путем следующих качественных рассуждений. Вернемся к рис. 4.1, из которого ясно, что назначение устройства K состоит в компенсации ошибки. Допустим, что K осуществляет *перекомпенсацию.* Если начальное значение E положительно, то в результате перекомпенсации оно становится отрицательным; дальнейший эффект перекомпенсации ведет к тому, что E снова становится положительным и бóльшим по модулю, и так до бесконечности; система идет «вразнос». Наша задача при создании системы, безусловно, состоит в том, чтобы выбрать такую схему K, при которой следящая система в целом была бы стабильной, т. е. ошибка (4.1.1) уменьшалась бы до своего конечного установившегося значения A (желательно, чтобы это был нуль).

Только что полученные нами краткие сведения относительно обратных связей и природы возникновения колебаний могут пролить свет на болезнь нервной системы, называемую *атаксией.* Нижеследующие строки заимствованы из «Кибернетики» Винера **).

«В неврологическую клинику приходит больной. Он не парализован и, получив приказание, может двигать ногами. Тем не менее он страдает тяжелым недугом. Он идет странной неуверенной походкой и все время смотрит вниз, на землю и на свои ноги. Каждый шаг он начинает с рывка, выбрасывая вперед сначала одну, потом другую ногу. Если ему завязать глаза, он не может стоять, он шатается и падает. Что с ним?

Приходит другой больной. Пока он неподвижно сидит на стуле, кажется, что у него все в порядке. Но

*) Конечно, это не единственная неприятность, которая может встретиться в следящих системах, однако мы на этом не будем останавливаться.

**) Норберт Винер, Кибернетика, или управление и связь в животном и машине, Изд-во «Сов. радио», 1958.

если предложить ему папиросу, то при попытке взять ее рукой он промахнется. Затем он столь же тщетно качнет руку в обратном направлении, потом опять вперед, и, наконец, его рука станет совершать лишь быстрые и бесцельные колебания. Дайте ему стакан воды, и он выплеснет всю воду, прежде чем сумеет поднести стакан ко рту. Что с ним?

Оба больные страдают разными формами так называемой атаксии. Их мышцы достаточно сильны и здоровы, но они не могут управлять своими движениями. Первый больной страдает сухоткой спинного мозга (tabes dorsalis). Часть спинного мозга, обычно воспринимающая ощущения, повреждена или разрушена поздними осложнениями от сифилиса. Поступающие сигналы притуплены или даже полностью пропадают. Рецепторы в суставах, сухожилиях, мышцах и подошвах его ног, обычно сообщавшие ему о положении и движении ног, не посылают сигналов, которые центральная нервная система могла бы принять и передать, и чтобы получить информацию о положении своего тела, больной должен полагаться на глаза и органы равновесия внутреннего уха. Физиолог на своем языке скажет, что больной потерял значительную часть проприоцептивных и кинетических ощущений.

Второй больной не потерял проприоцептивных ощущений — у него повреждение мозжечка, и он болен так называемым мозжечковым, или интенционным тремором. По-видимому, функция мозжечка в этом плане состоит в том, чтобы соразмерять мышечную реакцию с проприоцептивными сигналами, и если эта соразмерность нарушена, одним из следствий может явиться тремор.

Мы видим, таким образом, что для эффективного воздействия на внешний мир необходимо не только иметь хорошие эффекторы (исполнительные органы), но, кроме того, действие эффекторов должно находиться под надлежащим контролем центральной нервной системы; показания же контрольных органов должны сочетаться надлежащим образом с другими сведениями, поступающими от органов чувств, образуя правильно соразмеренные выходные сигналы к эффекторам.»

Итак, первый случай атаксии — утрата обратной связи, а второй, по-видимому, — нестабильность K.

§ 4.2. Резонансные частоты в нервных сетях

Познакомившись со всем изложенным выше, читатель может быть обеспокоен отсутствием наглядного соответствия между механизмами обратной связи в нервной системе и описанием нервной системы в виде сети, состоящей из абстрактных нейронов. Получается, что в сети, состоящей из нейронов, работающих в дискретном времени, присутствует непрерывно меняющийся сигнал обратной связи. Противоречие снимается, если в нервной сети отказаться от дискретной шкалы времени. Мы установили, что нейрон не может быть возбужден более одного раза в течение рефрактерного периода. Тем не менее допустима различная частота импульсов возбуждения, если период повторения превышает период рефрактерности. Таким образом, величину давления на подошву ноги можно закодировать, например, определенной частотой импульсов, передаваемых по аксонам соответствующих нервов. Действительно, нейрофизиологические данные подтверждают это и указывают, что частота, по-видимому, пропорциональна логарифму давления. Наглядный пример можно позаимствовать из описания зрительной системы лягушки (§ 2.1). Мы видели, что ганглиозные клетки, реагирующие на перемещение или изменение контраста, реагируют более интенсивно (растет частота возбуждений), когда граница изображения резко обозначена или объект передвигается быстро по сравнению со случаем нечеткой границы и медленного перемещения объекта.

Замечания, подобные приведенным выше, должны служить предостережением против излишне поспешного отождествления *любой* нашей модели нервной системы и живого мозга. При построении этих моделей использовались далеко не все известные современной науке факты, являющиеся в свою очередь не более чем каплей в океане закономерностей мозга, которые еще предстоит экспериментально выявить.

Чтобы усилить это предостережение, последующую часть этого параграфа мы посвятим модели нервной системы, в корне отличающейся от модели Маккаллока — Питтса. Главная роль теперь отводится частотам. Эта модель была предложена Петером Х. Грином на конференции 1962 г. по самоорганизующимся системам в Чикаго. Его доклад назывался «О представлении информации в моделях нервных сетей» *). Более подробный отчет об этой работе Грин опубликовал в сентябрьском и декабрьском номерах журнала «Bulletin of Mathematical Biophysics» за 1962 г. Статья начинается так:

«Ряд характерных особенностей поведения животных, проявляющихся главным образом при обучении, натолкнул многих исследователей на мысль, что одной из важнейших функций мозга является установление или изменение функциональных связей между нейронами. В настоящей работе гораздо более внушительный перечень особенностей поведения позволяет предположить, что, возможно, существует еще один важный механизм, использующий резонансные явления в линейных цепях мозга.

Чтобы объяснить упомянутый выше расширенный перечень явлений в терминах модели, описываемой только межнейронными соединениями, потребуется принять еще целый ряд допущений; в то же время как те, так и другие факты могут быть немедленно объяснены на предлагаемой модели. Входной элемент может установить функциональную связь с требуемым выходным элементом, если частота его возбуждений близка к резонансной частоте последнего. Информация от некоторой совокупности элементов рассматриваемого типа может быть передана по нервному тракту простейшим частотным кодом. Одно предположение нейрофизиологического характера состоит

*) Peter H. Greene, On the representation of information by neural nets models. [Русский перевод статьи помещен в сборнике «Принципы самоорганизации», Изд-во «Мир», 1966.]

в том, что частотно-кодированные сигналы возбуждают переходные процессы в сети дендритов. Если амплитуда возбуждения (в дендритах) становится достаточно большой, что и должно быть вблизи резонанса, возбуждение передается на аксон по принципу «все или ничего». Действие этого сигнала скажется в возбуждении резонирующих цепей и в подавлении прочих — это является неотъемлемым свойством любой линейной цепи. Показывается, как при помощи этого механизма можно объяснить целый ряд определенных особенностей поведения.

Все отмеченное выше заставляет предположить, что для машины, которой доступны разумные действия, вряд ли имеет ценность модель, построенная на статических элементах, из которых изгнаны различные колебания и нестабильности; гораздо вероятней, что именно подобные колебания должны использоваться системой, побуждая ее к действию».

Далее Грин приводит девять важных характеристик поведения животных.

«...1. Поведение является сложным и описывается совокупностью многих переменных (например, мускульные усилия). Таким образом, входы и выходы часто многомерны.

2. Поведение часто определяется *относительной* интенсивностью сигналов. В качестве примера можно привести результирующее направление движения эффектора, управляемого несколькими моторными элементами, а также такие типы поведения как, например, «брачные игры» у некоторых видов животных, которые, по-видимому, направляются противоречивыми побуждениями. Другими словами, вид реакций может определяться отношением выходных величин, в то время как их абсолютная величина определяет интенсивность.

3. При образовании условного рефлекса афферентные нейроны, соответствующие самым разнообразным стимулам, могут вызвать идентичное распределение реакций на выходе. Так, реакция нервной сети может быть в большой степени инвариантной по отношению к *локализации* входа.

4. Многие реакции на стимул, по-видимому, постепенно формируются из более элементарных составляющих. Примером могут явиться образцы сложного поведения, определяемые инстинктом животного, а также разнообразные эксперименты, связанные с обучением.

5. Часто сложная реакция организма в ответ на комплексный стимул проявляется как единое целое, а не складывается из независимых более элементарных реакций на его составные части.

6. Ответная реакция, однажды возникнув, может продолжиться даже после исчезновения вызвавшего ее стимула.

7. При исследовании инстинктивного поведения оказывается, что существуют определенные «пусковые» стимулы (releasing stimuli), кстати, довольно простые и носящие дискретный характер, в ответ на которые реакция проявляется с максимальной интенсивностью независимо от наличия или отсутствия других, казалось бы необходимых, стимулов. В ряде случаев «пусковые» стимулы обладают свойством аддитивности. Интенсивность реакции на неполный комплексный стимул зависит только от того, какие из его пусковых компонент присутствуют, а какие — отсутствуют. Как бы то ни было, при любой попытке подойти к теории обучения с количественной стороны, приходится делать подобные предположения.

8. Если несколько стимулов, действуя сообща, вызывают некоторую реакцию, моторный центр, со своей стороны, должен быть в состоянии ответить на такое возбуждение достаточно сложным образом.

9. Животное часто в состоянии быстро переключиться с одной реакции на другую, не осуществляя скрупулезного процесса точной подстройки отдельных частных черт новой реакции. Это положение подтверждается нашими наблюдениями — от наблюдения простейших моторных реакций до сложных проявлений инстинкта у чаек и, наконец, до выработки человеком субъективных оценок своим собственным мыслям».

Далее Грин рассматривает линейную модель, описываемую системой дифференциальных уравнений вида

$$\sum_j b_{ij}\left(\frac{d}{dt}\right)y_j = x_i,$$

где $i=1, \ldots, n$; x_i и y_j — соответственно i-й вход и j-й выход, а $b_{ij}(d/dt)$ — некоторый многочлен, включающий оператор дифференцирования по t. Грин отмечает, что подобная система обладает рядом хорошо известных свойств.

«...Если в произвольную точку такой системы подать одиночный импульс, то, как хорошо известно, ее ответная реакция будет суперпозицией *нормальных колебаний* (normal modes), или распределением простых реакций, каждая из которых имеет одну и ту же частоту для всей сети. Эти *естественные*, или *резонансные частоты* могут быть выражены через параметры системы — коэффициенты b_{ij}. Индивидуальное нормальное колебание характеризуется распределением отношений интенсивности возбуждения в различных точках одна относительно другой, в то время как абсолютная интенсивность в целом зависит только от места приложения и интенсивности входа. Нормальные колебания не влияют одно на другое в том смысле, что возбуждение одного нормального колебания ограничивается этим колебанием (не вызывает других нормальных колебаний). Любое стимулирование может быть представлено в виде суммы синусоидальных колебаний (быть может, затухающих или нарастающих), и каждая из этих входных частот вызовет возбуждение всех нормальных колебаний, причем величина возбуждения зависит от частоты и точки приложения входного сигнала. Если частота источника возбуждения приближается к частоте резонанса, амплитуда соответствующего нормального колебания значительно превышает амплитуды остальных нормальных колебаний. Таким образом, к какой бы точке системы ни приложить входное возбуждение, имеющее частоту, близкую к частоте резонанса, в системе будут преобладать колебания соответствующей частоты. Итак, при резонансе выходы принимают большие значения, а распределение отношений ответов инвариантно относительно положения источника».

Грин предлагает следующую модель нервной активности, основанную на свойствах линейных систем. «Предположим, что одной из задач сенсорных входов является возбуждение линейных систем. Конечно, этим системам предшествуют, за ними следуют и ими управляют *нелинейные системы, о которых мы не станем делать никаких предположений.* Например, можно допустить, что входным сигналам соответствует по линейному закону рост активности в цепях дендритов, которая, достигнув определенного значения, может вызвать переход аксона из невозбужденного состояния в возбужденное. Однако мы не будем разбирать подробно эту частную иллюстративную модель; наши выводы всецело будут основываться на свойствах линейных цепей. Мы будем считать, что как сенсорные, так и моторные системы являются линейными цепями. Компонентами выходного вектора будем считать уровни возбуждения в различных точках цепи; эти сигналы и управляют активностью множества нервов и мускулов, что в конечном счете определяет поведение животного. Естественно предположить, что сенсорная и моторная системы построены таким образом, что ряд важных для организма входных воздействий возбуждают собственные колебания в нейронных сетях, которые характеризуют наличие необходимых признаков в поступившей сенсорной информации. Это в свою очередь является необходимым условием возбуждения нормальных колебаний в моторной системе. Колебания, возникающие в моторной системе, и определяют конкретное поведение животного».

Не будем здесь углубляться в анализ частного примера, на котором Грин демонстрирует возможность моделирования различных стилей передвижения «лошади» при помощи возбуждения нормальных колебаний в простой нейронной сети. Ограничимся выводами, к которым он приходит, анализируя свою модель «лошади».

«В заключение мне хотелось бы сделать кое-какие выводы в отношении нашего гипотетического четвероногого, подчеркнув, что эти замечания относятся в основном к элементам, объединенным в системы. Предположим, что мы обнаружили у себя в лаборатории эту модель и, приняв ее за живую лошадь, решили провести нейрофизиологические эксперименты, чтобы выяснить особенности управления для всех четырех стилей ее передвижения. Для каждого стиля мы ожидаем обнаружить систему нейронов, обусловливающую работу четырех мускулов в заданном ритме. Для одного из стилей нужно соединить ноги A и B; обратная связь от B на C должна подавлять движения последней, в то время как C через ряд синапсических задержек подключена к D, — или что-то в этом роде. Затем мы станем разыскивать элемент, осуществляющий переключения разных схем управления в соответствии с нужным в данный момент стилем передвижения. Возможно, эти четыре системы не будут полностью физически независимы: изменение порогов может позволить одной системе управлять более чем одним стилем. Может оказаться очень трудным в сложной системе проследить все структуры, регулировки и, наконец, связать их с определенным внешним поведением, но в принципе это и было бы полным пониманием функционирования нервной системы животного. Однако даже в случае нашей модели мы не добились бы подробного уровня понимания, следуя вышеупомянутым предположениям и воспользовавшись обычными приемами анализа реакций отдельных нейронов. Устанавливая частные зависимости, мы оказались бы погребены под грудами данных, анализ которых либо вообще невозможен, либо должен длиться бесконечно долго, хотя в действительности наша модель состоит из единственной системы четырех нейронов и снабжена четырьмя каналами связи между ними — это не четыре системы с переключателем, и здесь нет связей, которые можно было бы принять за тормозящие, цепи задержки и т. д., исходя из аналогий с внешними признаками, когда медленная походка называется «шаг», более быстрая — «рысь», «галоп» и т. д.».

Прежде чем окончить этот раздел, необходимо сделать пару выводов, которые могли возникнуть после знакомства с работой Грина.

Первый вывод — это не принятие предположения о том, что мозг является линейной цепью (на самом деле, *чем подробней мы знакомимся с характеристиками мозга, тем бесспорней убеждаемся в их нелинейности*), но согласие, что явления резонанса в мозгу *могут* играть значительную роль в формировании поведения животного. Выбор Грином *линейной* цепи диктовался необходимостью упростить математические выкладки при выводе зависимостей для частот резонанса *).

Второй вывод крайне важен, и к нему приводит последняя цитата. А именно: как бы ни был объективен физиолог или психолог, выбор им эксперимента и экспериментальной методики всегда определяется какой-то гипотезой, или моделью, или просто интуицией. Если модель никуда не годится, время будет убито на эксперименты, которые либо вообще не дадут почти никакой информации, либо полученные результаты наверняка будут неверно истолкованы. Математические модели могут так же не соответствовать действительности, как и нематематические. Бесспорным достоинством первых является то, что они часто позволяют количественно и качественно предсказывать поведение системы, а поэтому могут быть легче подвергнуты более строгой проверке, чем нематематические модели. Но одно только жонглирование формулами не придает теории магической силы, поэтому всегда следует опасаться дефектов в наших прежних моделях, а также ложных идей, ведущих к появлению новых моделей, математических или же нематематических.

*) Читателя, который заинтересуется резонансными нервными сетями, можно отослать к статьям: R. L. Beurle, Properties of a mass of cells capable of regenerating pulses, Phil. Trans. Roy Soc., London, Ser. B, 240 (1956) 55. B. G. Farley, Self-organizing models for learned perception, in «Self-organizing systems», Pergamon Press, New York, 1960. [Русский перевод: Фэрли Б. Г., Самоорганизующиеся модели для обучения восприятию. В сб. «Самоорганизующиеся системы», Изд-во «Мир», 1964.]

§ 4.3. Протезирование и гомеостазис

Исследования роли обратной связи в нервной системе человека начинают оказывать заметное влияние на конструирование протезов.

Старейший вид протеза — «деревянная нога» — выполнял единственную функцию утраченной конечности — поддерживание тела. Современные протезы конечностей, как, например, протез руки, который позволяет брать предметы и производить некоторые манипуляции с ними, возвращают своему обладателю уже несколько степеней свободы. Но мы видели в § 4.1, что для полноценного управления конечностями чрезвычайно важна обратная связь (так называемое, кинестетическое ощущение). В настоящее время уделяется много внимания созданию протезов, использующих нечто вроде кинестетического ощущения, например протезы руки, в которых на кончиках пальцев помещаются датчики давления, электрический сигнал от которых подводится к коже уцелевшей части руки. В конце 1961 г. стало известно, что около 35 таких протезов с обратной связью изготовлены и используются в Советском Союзе *). Дальнейшие усилия в этом направлении как в СССР, так и в США, потребуют проведения ряда нейрофизиологических исследований особенностей стыковки кинестетических цепей обратной связи и нервной системы человека. Кроме того, потребуется соответствующий анализ, помогающий изготовлению этих цепей «в металле».

Существует также ряд физиологических процессов, в значительной мере использующих принцип обратной связи и независимых от собственно нервной системы. Можно привести внушительное количество примеров, когда обратная связь не только присутствует в физио-

*) На самом деле такие протезы были изготовлены в Советском Союзе намного раньше: Кобринский А. Е., Брейдо М. Г., Гурфинкель В. С., Славуцкий Я. Л., Сысин А. Я, Полян Е. П., Цетлин М. Л., Якобсон Я. С., Работы ЦНИИПП в области создания биоэлектрической системы управления. Труды 7-й научной сессии Центр. научно-исслед. ин-та полиграф. пром-сти, 1958, 125—132. (*Прим. перев.*)

логическом процессе, но и имеет жизненно важное значение. Это — *гомеостазис*. Говорят, что в (физиологической) системе имеется гомеостазис, если после некоторого внешнего воздействия, вызвавшего отклонение ряда ее параметров от нормальных значений, составные части системы реагируют и взаимодействуют таким образом, что эффект постороннего влияния в большой степени нейтрализуется. Мы встречались с гомеостазисом в § 3.1. Вот еще два примера, которые должны пояснить это определение.

а) Когда человеку холодно, соответствующие сигналы воздействуют на определенный механизм мозга, действие которого вызывает дрожь. В результате работы мускулов выделяется тепло, которое препятствует охлаждению.

б) Внезапное кровотечение вызывает резкое падение кровяного давления. Это влечет сужение просвета артерий, что в свою очередь уменьшает величину падения давления.

До какой степени *математическая теория* обратной связи углубит наше понимание механизма гомеостазиса — вопрос будущего, но бесспорно, что *понятие* обратной связи служит серьезным подспорьем в этом вопросе *).

Гомеостазис проявляется и в процессах самовосстановления. Шведский ученый Ларс Лёфгрен (Lars Lofgren) теоретически исследовал сеть из логических элементов, похожих на наши абстрактные нейроны, обладающую способностью обнаруживать и заменять неисправные элементы **).

---

*) Ряд статей на эту тему помещен в «Homeostatic Mechanisms» Brookhaven Symposia in Biology, № 10, Upton, N. Y., 1957.

**) Self-repair as the limit for automatic error-correction in H. Von Foerster and G. Zopf (eds.), «Proc. symp. on principles of self-organization», Pergamon Press, New York (1962), 181—228. [Русский перевод: Самовосстановление как предел для автоматической коррекции ошибок. Сб. «Принципы самоорганизации», Изд-во «Мир», М., 1966] и «Kinematic and tesselation models of self-repair» in «Biological prototypes and synthetic systems», Plenum Press, N. Y. (1962), 342—369. [Русский перевод помещен в сб. «Проблемы бионики», Изд-во «Мир», 1965, стр. 475—517.]

Изложим выводы, к которым пришел Лёфгрен. Он показал, что максимальная продолжительность жизни полностью локализованного автомата (т. е. автомата, чей рост, если таковой наблюдается, должен быть ограничен постоянным объемом) из ненадежных компонентов достигается в том случае, когда в автомате присутствует система обнаружения ошибок. Далее он рассмотрел вопрос локализации ошибок с учетом того, что сбои возможны в самой системе поиска ошибок, и показал, что максимальная продолжительность жизни такого полностью локализованного автомата конечна. Эта величина пропорциональна квадрату срока службы компонентов автомата. Таким образом нетрудно видеть, что продолжительность жизни всего автомата может значительно превышать срок службы отдельных его компонентов.

В своей второй статье он снимает ограничение полной локализации. Показано, что самовосстанавливающийся неполностью локализованный автомат может иметь неограниченную продолжительность жизни, если он обладает необходимыми возможностями роста и самовосстановления. В этой интерпретации подобный автомат включает в себя самовоспроизводящиеся подавтоматы. Приводятся вероятности возникновения ошибок, сложность самовоспроизводящихся подавтоматов и геометрическая форма всего автомата с самовосстановлением в процессе его роста, обеспечивающие бессмертие автомата.

Лёфгрен подчеркивает, что конечное время жизни самовосстанавливающихся полностью локализованных автоматов соответствует конечному времени жизни любого растения или животного в природе. Неограниченное время жизни самовосстанавливающегося нелокализованного автомата, развивающегося безгранично, соответствует неограниченному времени существования человеческого общества, рассматриваемого как нечто целое.

Результаты Лёфгрена подтверждают мысль, что человечество в борьбе за жизненное пространство должно выйти за пределы Земли.

§ 4.4. Образы и понятия

В этом параграфе мы будем иметь дело с вопросами, рассматривавшимися Винером в VI главе «Кибернетики», а также в работе У. Питтса и У. Маккаллока «Как мы узнаем понятия... *)».

Как мы узнаем черты знакомого нам человека, когда видим его в профиль или в фас? Как мы узнаем квадрат, независимо от того, мал он или велик, далек или близок? Как мы узнаем круг, даже если наблюдаем его под углом в виде эллипса?

Важнейшим фактором при сравнении форм разных предметов является зрительно-мускульная система обратной связи. Иногда эта связь носит чисто гомеостатический характер, например сокращение зрачка, поддерживающее интенсивность света в узких пределах. Когда в поле периферического зрения попадает предмет, выделяющийся своей яркостью, цветом или движением, возникает рефлекторный сигнал обратной связи, вызывающий перемещение изображения в центр поля зрения, т. е. в область сетчатки, обладающую наибольшей чувствительностью к форме и цвету.

Другими словами, человек стремится придать предмету, привлекшему его внимание, стандартное положение и ориентацию, чтобы зрительное впечатление, возникающее от подобных предметов в мозгу, менялось в минимально возможных пределах.

Однако эта система зрительно-мускульной обратной связи недостаточно полно объясняет наше восприятие таких понятий, как стул, круг и т. п. Нам по-прежнему приходится задавать себе вопрос, каким образом мозг узнает, например, букву А независимо от действия на нее ряда преобразований — поворотов, сдвигов в поле зрения, изменения размера и т. д. Питтс и Маккаллок посвятили свою работу рассмотрению возможных нервных механизмов, действием которых

*) W. H. Pitts and W. S. McCullock, How we know universals — the perception of auditory and visual forms, Bull. Math. Biophys, 9 (1947), 124—147.

можно было бы объяснить способность мозга к восприятию. Их конструкции обнаруживают ряд аналогий, возможно, поверхностных, со структурами живого мозга.

Займемся немного математикой. А поэтому вспомним понятие группы.

Определение 4.4.1 *). *Группой преобразований* называется совокупность G таких преобразований, что если $A, B \in G$, то $AB \in G$ (под AB понимается преобразование, полученное выполнением сначала B, а затем A; заметим, что, в общем случае, $AB \neq BA$) и где:

а) тождественное преобразование $I \in G$ (I — тривиальное преобразование, которое оставляет все неизменным, например, смещение на нулевое расстояние).

б) Если $A \in G$, то и его инверсия $A^{-1} \in G$, где $AA^{-1} = A^{-1}A = I$ (A^{-1} — преобразование, результат которого полностью противоположен результату от действия преобразования A).

Заметим, что для преобразований автоматически выполняется свойство ассоциативности:

$(AB)C = A(BC) = C$, после которого следует B, а затем A.

Питтс и Маккаллок утверждают:

«...ряд сетей, объединенных в особые структуры нервной системы, предназначен для классификации информации в соответствии с полезными общими свойствами. В зрительной сфере эти сети устанавливают эквивалентность реализаций (apparitions) за счет подобия и конгруэнтности, сохраняющихся при наблюдении одного и того же предмета из разных точек пространства. В сфере слухового восприятия они распознают тембр и гамму звуков независимо от высоты тона. Эквивалентные реализации во всех случаях обладают общей фигурой и определяют группу преобразований, переводящих эти реализации одна в другую, но сохраняющих фигуру неизменной. Так, например, группа сдвигов переводит квадрат с одного места на

*) См. Приложение 2.

другое, однако фигура квадрата остается неизменной. Эти фигуры являются «геометрическими объектами» по Картану и Вейлю или «образами» по Вертхаймеру (Wertheimer) и Колеру (Kohler).

Мы ищем общие принципы построения нервных сетей, распознающих фигуры так, что для всех реализаций одной и той же фигуры они выдают один и тот же выход».

Итак, все, что мы видим или слышим, отображается в виде своеобразной картины из возбужденных нейронов в коре головного мозга M. Распределение возбуждения в M описывается функцией $\varphi(x, t)$, где $\varphi(x, t)=1$, если нейрон, находящийся в точке x, возбужден в момент t, и $\varphi(x, t)=0$ в противном случае.

Пусть G — группа преобразований, которая переводит функции $\varphi(x, t)$, описывающие реализации, возникающие при предъявлении некоторой фигуры, в эквивалентные им функции. Пусть G имеет N элементов. Питтс и Маккаллок рассматривают случай, когда преобразования T из G могут быть порождены преобразованиями $\tilde{T}$ основного множества M *), так что

$$T\varphi(x)=\varphi(\tilde{T}x).$$

Например, если G — группа сдвигов, то

$$T\varphi(x)=\varphi(x+a_T),$$

где a_T — постоянный вектор, зависящий только от T. Если G — группа изменений масштаба, то $T\varphi(x)=\varphi(a_T x)$, где a_T — положительное действительное число, зависящее только от T.

Все подобные преобразования линейны, то есть

$$T(a\varphi(x)+b\Psi(x))=a\varphi(\tilde{T}x)+b\Psi(\tilde{T}x)=$$
$$=aT\varphi(x)+bT\Psi(x).$$

Простейший способ получения инвариантов данного распределения $\varphi(x, t)$ возбуждения заключается в том, чтобы усреднить группу G. Так, пусть f —

*) Там, где это не вызовет двусмысленности, мы будем опускать переменную t и записывать $\varphi(x, t)$ как $\varphi(x)$.

произвольный функционал (т. е. функция от функции), ставящий в соответствие каждой функции $\varphi(x, t)$ некоторое число. Осуществляем поочередно все преобразования $T\varphi(x, t)$, вычисляем $f(T\varphi)$ и усредняем результаты по G. Имеем

$$a = \frac{1}{N} \sum_{\substack{\text{по всем} \\ T \in G}} f(T\varphi).$$

Если мы начнем с $S\varphi (S \in G)$ вместо φ, то получим

$$\frac{1}{N} \sum_{\substack{\text{по всем} \\ T \in G}} f(TS\varphi) = \frac{1}{N} \sum_{\substack{\text{по всем} \\ R \in G \text{ таким, что } RS^{-1} \in G}} f(R\varphi) = a,$$

поскольку G — группа. Чтобы полностью характеризовать фигуру, соответствующую функции $\varphi(x, t)$ по G (т. е. понятие, соответствующее той реализации, которая вызывает возбуждение нейронов $\varphi(x, t)$), нам нужен полный набор таких значений a для разных функционалов f (один функционал, например, характеризует «округлость», другой — «прямоугольность» и т. п.). Функционалы можно отличить один от другого, приписав им индексы ξ, принимающие значения из некоторого множества Ξ. Таким образом, мы получим различные усредненные значения

$$f_\xi(\varphi) = \frac{1}{N} \sum_{T \in G} f_\xi(T\varphi).$$

Введем новую совокупность нейронов, обозначим ее также Ξ и поставим в соответствие каждому ξ один нейрон. Ξ может быть многомерным; в этом случае ξ может быть задан своими координатами $(\xi_1, \ldots, \xi_m)$. Если система нейронов не нуждается в полной информации для правильного распознавания очертаний, то множество Ξ может оказаться значительно меньше M, иметь меньше измерений и вообще выродиться в изолированные точки. Время t и некоторые из x_j, описывающие положение в M, могут служить координатами в Ξ.

Пусть Ξ многомерно. Тогда простейшая нервная сеть, реализующая описанный выше формальный процесс, может быть получена следующим способом.

Повторим исходное множество M в $N-1$ слоях, для каждого $T \in G$ — слой M_T, затем подключим их к M или к его сенсорным путям таким образом, чтобы при наличии в M функции возбуждения $\varphi(x)$ в M_T возникала преобразованная функция $T\varphi(x)$. Далее, отдельно для каждого значения ξ и для каждого M_T соответствующей сетью нейронов вычисляется значение $f_\xi(T\varphi)$. Результаты вычислений от всех M_T суммируются, поступая на нейрон, находящийся в точке ξ мозаики Ξ.

Но если мы будем действовать подобным образом, нам потребуется неправдоподобно много нейронов. Все множества M_T в совокупности имеют размерность в N раз большую, чем M, и во столько же раз больше степеней свободы, чем в группе G. Еще более важно число нейронов и связей, которые необходимы для вычисления значений $f_\xi(T\varphi)$, зависящих в принципе, от всего распределения $T\varphi$ и поэтому требующих отдельного вычислителя для каждого ξ и каждого $T \in G$. Особенно это сказывается, если схема для вычисления f_ξ выполнена отдельно от M_T: количество соединений в этом случае растет еще в большей степени.

Можно существенно упростить эту модель.

Пусть слои M_T остаются соединенными, как и ранее, но повысим их пороги до такой величины, чтобы они не могли сами по себе возбуждаться от своих специфических афферентов. Введем дополнительные управляющие нервные волокна разветвляющиеся в пределах каждого M_T таким образом, чтобы при их возбуждении компенсировалось действие повышенных порогов, и на M_T могла бы снова возникнуть картина $T\varphi(x)$.

Свяжем все нейроны, имеющие одну и ту же координату x во всех N слоях M_T, аксонами с нейроном x, находящимся в специальном слое, который также является копией M, скажем, Q. Пусть любой аксон может возбудить этот нейрон. Если дополнительные управляющие нервные волокна, введенные выше, последовательно возбуждать так, чтобы выдавался каждый слой M_T, причем только один слой в

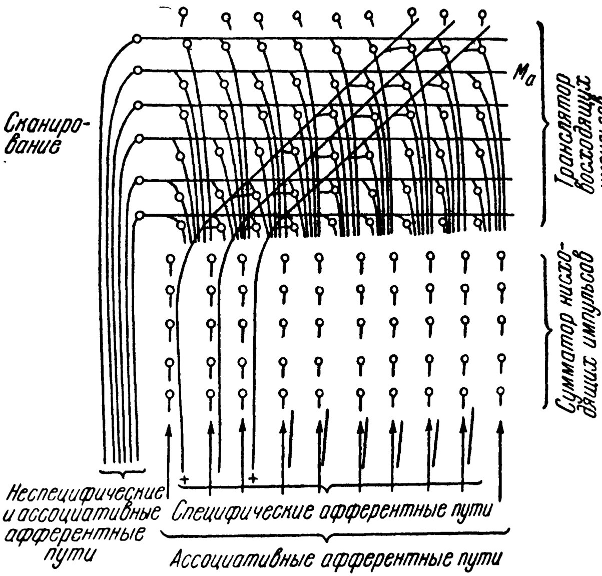


Рис. 4.3. Входная информация поступает по специальным афферентным путям (отмечено знаком +) и проходит до уровня в рецептивном поле, активированном в данный момент неспецифическими афферентами. Это позволяет импульсам поступать только в один слой.

данный момент времени, то на слое Q будут последовательно возникать все преобразования $T\varphi$ функции $\varphi(x)$ (рис. 4.3).

Для каждого ξ имеется единственный вычислитель f_ξ, который теперь воспринимает входные значения из Q, а не из M_T, как раньше. Все значения $f_\xi(T\varphi)$ будут получены последовательно согласно сигналу сканирования, вызывающему периодическое проектирование на Q картин $T\varphi$. Эти значения $f_\xi(T\varphi)$ могут накапливаться в течение цикла на конечном Ξ-нейроне ξ любым способом.

Описанное устройство иллюстрирует использование общего полезного приема, который можно было бы назвать *обменом времени на оборудование.* Другими словами, переходя от параллельного выполнения операций к последовательному, мы, экономя на оборудовании, расплачиваемся соответствующим снижением быстродействия.

Насколько мне известно, до сих пор нельзя доказать нейрофизиологическими приемами соответствие рассмотренной модели механизмам человеческого мозга. Тем не менее внешнее сходство настолько велико, что, по словам Винера, когда нейрофизиолог фон Бонин увидел схему, подобную нашей (рис. 4.3), он спросил: «Не схема ли это четвертого слоя зрительной области коры головного мозга?»

§ 4.5. Некоторые другие направления

В этом параграфе вкратце изложены вопросы в основном из сферы кибернетической психологии, касающиеся «разумных машин», которых мы в подробностях рассматривать не будем.

В 1958—1959 гг. трое американских ученых встретились в Центре современных исследований в области наук о поведении (Center for Advanced Study in the Behavioral Sciencies) и задумались над вопросом: «Имеет ли отношение кибернетика к психологии?». Заинтересованный читатель сможет найти прекрасно аргументированный утвердительный ответ в книге: G. A. Miller, E. Galanter and K. H. Pribram, Plans and the structure of behavior, Publ. by Holt, Rinehart and Winston, Inc., 1960).

Их работа в области «искусственного интеллекта» была весьма плодотворной — они занимались программированием на ЦВМ процессов поиска решения задач. По этому вопросу читателю будет полезно ознакомиться с обзорной статьей: M. L. Minsky, Steps toward artifical intelligence, Proc. IRE, **49**, № 1 (1961), 8—30.

В кибернетической психологии фундаментальное понятие обратной связи используется в двух смыслах:

а) в буквальном смысле — как в следящих системах, б) в переносном смысле — подразумевая обратную связь между науками о мозге и машинах (имеется в виду обратная связь между психологией и искусственным интеллектом).

Те, кто интересуется буквальным смыслом этого термина, могут обратиться к упоминавшейся выше книге. Мы же коротко обсудим обратную связь между изучением мозга и вычислительных машин. Это могло бы вылиться в форму беседы по *нейрофизиологии*, как и при рассмотрении модели мозга Маккаллока и Питтса, но здесь мы сосредоточим наше внимание на использовании вычислительных машин как автоматов с тем, чтобы с их помощью объяснить различные психологические теории. Работы Тьюринга (см. обсуждение гипотезы Тьюринга в § 1.6) показали, что устройства типа вычислительных машин являются как раз теми машинами, на которых удобно моделировать человеческое поведение. Заметим, что в задачу машины не входит буквальное моделирование (мы не будем требовать от машины, играющей в шахматы, чтобы она имела пальцы); они предназначены для имитации тех аспектов поведения, на которых в данное время мы хотим сосредоточить внимание. Так, психолог-теоретик, следуя рассматриваемому принципу обратной связи, может воплотить свои предположения в программу для вычислительной машины, которая будет занесена в ее «память». При предъявлении машине входного «стимула», она, как и живой организм, «отреагирует» на него, действуя в строгом соответствии с заложенной в нее программой.

Следует различать при этом так называемый *«искусственный интеллект»* и *«моделирование»*. Первый термин подразумевает, что машина решает задачу любым способом, а второй — что машина копирует действия человека в подобной ситуации. На этот счет Минский отмечает:

«...кажется бесспорным, что, по крайней мере, в течение ближайших нескольких лет будет наблюдаться плодотворное содружество наук, изучающих поведение человека и попыток повысить «интеллектуальность» машин. Но в перспективе мы должны быть готовы к тому, что возникнут более эффективные приемы программирования, и необходимость в копировании поведения человека отпадет».

Но что касается психолога, то он бесспорно заинтересован в том, чтобы промоделировать поведение игрока в шахматы или специалиста-логика, но совсем не в том, чтобы заменить их машиной, или создать что-либо более совершенное. Для иллюстрации упомянем вкратце о двух программах для ЦВМ, относящихся к области «искусственного интеллекта».

1) Г. Гелернтер и Н. Рочестер*) составили программу для доказательства на ЦВМ теорем планиметрии путем применения заданных правил вывода к данной совокупности аксиом. Так же, как и учащийся, машина рассматривала чертежи. Для повышения надежности в каждом случае анализировалось несколько фигур, что позволяло избежать неверного заключения из-за случайного совпадения несущественных свойств. Рассмотрение чертежей в громадной степени упростило процесс поиска доказательства.

2) А. Л. Сэмюэль**) подводит следующий итог своим исследованиям.

«Довольно детально были рассмотрены две программы по обучению машины игре в шашки. Проделанной работы достаточно для подтверждения возможности составления такой программы для ЦВМ, по которой машина будет играть в шашки лучше составителя программы. Более того, она сможет сделать это в исключительно короткий период времени (8—10 часов машинной игры), будучи снабжена лишь правилами игры, «ощущением» направления, а также приблизительным и неполным перечнем параметров, которые, как предполагается, имеют какое-то значение, но правильные знаки и относительные веса

---

*) H. Gelernter and N. Rochster, Intelligent behavior in problem-solving machines, IBM J. Res. Develop. 2, № 3 (1958), 336—345.

**) A. L. Samuel, Some studies in machine learning, using the game of checkers, IBM J. Res. Develop. 3, № 3 (1959), 211—229.

которых неизвестны. Принципы обучения машины, подтвержденные этими экспериментами, бесспорно, применимы и ко многим другим ситуациям *).

Восстановим в памяти содержание этой главы и укажем на ряд вопросов, заслуживающих внимания при дальнейших исследованиях. Мы убедились в исключительной важности понятия обратной связи и в том, как оно помогает разобраться в ряде нервных заболеваний, в явлении гомеостазиса, а также позволяет совершенствовать протезы. Все эти области применения еще требуют длительных исследований.

Мы убедились в необходимости более сложных моделей нервных сетей для сложных систем с обратными связями и для частотного кодирования сенсорной информации. Нам известно, что зрительно-мускульная обратная связь играет существенную роль при узнавании видимых форм, известно, каким образом эта обратная связь может быть воплощена в нервной системе. Тем не менее нам еще очень далеко до понимания нервных процессов, лежащих в основе нашего восприятия понятий. Исследование механизмов обратной связи в психологическом аспекте и связанное с этим мастерство составления программ «искусственного интеллекта» по-прежнему являются источником многих исследовательских работ.

Мы в состоянии лишь упомянуть о таких исключительно интересных проблемах, как, например, теория клеточных автоматов (molecular automata theory), генетические коды и математическая лингвистика Чомского (Chomsky). Читателю, желающему основательнее познакомиться с работами в этой области, можно рекомендовать просмотреть следующие журналы: The Bulletin of Mathematical Biophysics, Information and Control, Journal of the Association for Computing Machinery, Journal of Theoretical Biology, Cybernetik, а также труды специализированных секций IRE (IEEE) по электронным вычислительным машинам, теории информации и т. п. Механизмам дей-

---

*) См. также М. М. Ботвинник, Алгоритм игры в шахматы, Изд-во «Наука», 1968. (*Прим. перев.*)

ствия обратной связи внутри живой клетки посвящена книга B. C. Goodwin, Temporal organisation in cells, Academic Press, Inc., London, 1963 *).

Мы подходим к концу этой главы с пониманием того, что перед кибернетикой еще неисчерпаемое море увлекательных проблем. Быть может, наиболее важными из них в области психологии и физиологии будут работы по установлению функций координирования и связи множества подсистем, которые и составляют в совокупности нервную систему человека, о чем пока мы имеем весьма смутное представление.

---

*) Из отечественных журналов мы посоветуем читателю журналы «Биофизика», «Кибернетика», «Техническая кибернетика», «Проблемы передачи информации», «Автоматика и телемеханика» и периодический сборник «Проблемы кибернетики». (*Прим. ред.*)

Глава 5

Теорема Гёделя о неполноте

В этой, нашей последней главе мы сосредоточим свое внимание прежде всего на основаниях математики *). В § 5.1 мы дадим краткий исторический обзор формального подхода к основаниям математики; там же покажем, почему теорема Гёделя полностью опровергает всю программу формалистов; в § 5.2 вернемся к рассмотрению рекурсивных множеств, которые изучались в первой главе; в § 5.3 обсудим некоторые общие свойства рекурсивных логических систем (рекурсивных логик). В § 5.4 мы ограничимся так называемыми арифметическими логиками, которые позволят нам (в § 5.5) доказать теорему Гёделя о неполноте. Наконец, мы вернемся к главной теме этой книги и обсудим философский спор о том, что можно извлечь из теоремы Гёделя для решения вопроса: «Имеет ли мозг существенные преимущества перед машиной?».

Я включил в эту книгу доказательство теоремы Гёделя главным образом по следующим трем причинам:

1) На нее часто ссылаются при споре о возможностях человека и машины.

2) Это очень важная теорема для оснований математики.

3) Я полагаю, что доказательство, данное в этой книге, достаточно короткое и простое для того, чтобы с его помощью можно было опровергнуть миф (су-

---

*) В качестве полного и очень интересного введения в этот предмет см. E. Nagel and J. R. Newman, Gödel's proff, New York Univ. Press, 1958. Для радикального изучения всех технических подробностей см. E. W. Beth, The Foundations of mathematics, North Holland Publishing Co., Amsterdam, 1959.

щественную часть современного математического фольклора) о том, что доказательство теоремы Гёделя является настолько трудным, что оно доступно только специалистам по математической логике.

§ 5.1. Основания математики

Кант утверждал, что аксиомы евклидовой геометрии априорно заданы человеческой интуиции. Это утверждение было в духе определения аксиомы как очевидной истины, определения, которое не вызывало никаких сомнений на протяжении 2000 лет.

Эта позиция была поколеблена в девятнадцатом веке благодаря работам Больяи, Лобачевского и Римана. Они постулировали геометрические системы, которые не были евклидовыми, более точно, системы, которые отвергали истинность следующей аксиомы Евклида: «Если даны прямая и точка, не лежащая на ней, то существует только одна прямая, которая проходит через данную точку параллельно данной прямой». И мы с вами живем в такое время, когда общепринятые взгляды на Вселенную опираются на теорию относительности Эйнштейна, а она имеет дело с пространством, геометрия которого лучше описывается неевклидовой римановой схемой, в которой параллельных линий просто не существует.

Другими словами, наши современные взгляды не только противоречат взглядам Канта, что аксиомы Евклида даны a priori человеческой интуиции, более того, они утверждают, что одна из аксиом Евклида, так называемая аксиома о параллельных, в действительности неверно описывает Вселенную. В настоящее время мы полагаем, что геометрия Евклида достаточно точно описывает пространственные отношения в нашей повседневной жизни, но не космическое пространство Вселенной. Теперь у нас может возникнуть искренний и важный вопрос, от которого точка зрения Канта позволяла легко отделаться: «Откуда мы знаем, что евклидова геометрия непротиворечива (т. е. свободна от противоречий)?»

Как будто бы для того, чтобы еще более усугубить важность этого вопроса, Риман показал, что *если евклидова геометрия непротиворечива, то и его (Римана) неевклидова геометрия также непротиворечива.* Вот прекрасный пример противоречия: не только непротиворечивость евклидовой геометрии перестает быть a priori очевидной, но оказывается, что она влечет непротиворечивость соперничающей с ней системы! Из этого можно сделать вывод, что тот, кто занимается аксиоматическими системами, должен ставить вопрос об их непротиворечивости независимо от того, как бы явно «истинно» система не описывала «реальный мир».

Между тем развитие теории множеств показало, что непротиворечивость системы нельзя установить просто при помощи общих рассуждений. В самом деле, теория множеств, развитая Кантором, казалась совершено непротиворечивой до тех пор, пока Рассел, а за ним и другие, не показали, что эта, казавшаяся такой «надежной» система содержит парадокс, к которому можно прийти следующим образом. Рассмотрим множество математиков — оно не является математиком, и, следовательно, это множество не является своим собственным элементом. А теперь рассмотрим множество всех вещей, о которых говорится в этой главе. Здесь, кроме всего прочего, говорится о множестве всех вещей, о которых говорится в настоящей главе, и следовательно, оно является элементом самого себя. Определим множество N как множество всех тех множеств, которые не являются своим элементом. Таким образом, *M принадлежит N тогда и только тогда, когда M не принадлежит M.*

Поэтому множество математиков принадлежит N, а множество всех вещей, о которых говорится в этой главе, не принадлежит N.

А само N принадлежит N? Согласно данному выше определению N принадлежит N тогда и только тогда, когда N не принадлежит N.

Парадокс! И следовательно, наивная теория множеств является противоречивой. Рассел устранил такие парадоксы введя свою теорию типов, но нам важно здесь подчеркнуть, что непротиворечивость не является очевидным свойством логической системы.

Риман показал, что его геометрия непротиворечива, если непротиворечива евклидова геометрия, а Гильберт показал, что евклидова геометрия непротиворечива, если непротиворечива арифметика, т. е. элементарная теория положительных целых чисел *).

Логическую систему можно рассматривать как некоторую совокупность аксиом, из которых можно получать теоремы повторным применением определенного числа правил вывода. Представители формальной школы, основанной Гильбертом, полагали, что при поисках доказательства непротиворечивости они могут полностью игнорировать все вопросы об истинности и смысле теорем и аксиом, а вместо этого могут рассматривать аксиомы просто как строчки символов, а правила вывода — как способы получения новых строчек. Более того, они решили потребовать, чтобы правила вывода были финитными и полностью детерминированными, т. е. в некоторой степени были похожи на работу машин Тьюринга, которые мы изучали в первой главе. Мы назовем логические системы, удовлетворяющие таким условиям, *рекурсивными логиками.* Если мы хотим использовать рекурсивную логику для описания теории положительных целых чисел, то мы должны снабдить ее символами, которые бы соответствовали основным количественным понятиям элементарной теории чисел.

Для доказательства непротиворечивости формалисты искали арифметическую логику, которая была бы полной, т. е. такую, в которой можно было бы вывести (как теоремы) все истинные утверждения о целых числах. Кроме того, они требовали, чтобы непротиворечивость системы была показана надежным, полностью детерминированным, «финитным» способом.

*) Разъяснение понятия «элементарная теория» и многие факты об элементарных теориях можно найти, например, в обзоре: Ю. Л. Ершов, И. А. Лавров, А. Д. Тайманов, М. А. Тайцлин, Элементарные теории, Успехи матем. наук, т. 20 вып. 4, 1965, стр. 37—108. (*Прим. ред.*)

Эта программа формалистов была разрушена теоремой Гёделя о неполноте, впервые сформулированной в его знаменитой работе о формальной неразрешимости подобных систем *).

Теорема утверждает, что любая адекватная непротиворечивая арифметическая логика неполна, т. е. что существует истинное утверждение о целых числах, которое нельзя доказать в такой логике. Этот важный результат (который мы вскоре докажем) показывает, что поиски формалистами полной непротиворечивой арифметической логики были обречены на неудачу. На самом деле Гёдель показал даже больше, а именно, что невозможно доказать непротиворечивость арифметической логики (пусть даже неполной) теми методами, которые выразимы в самой этой логике.

Впоследствии Генцен (Gentzen) доказал, что элементарная теория чисел непротиворечива, но при этом он использовал «ε_0-индукцию», которая является бесконечным обобщением обычной математической индукции, — метод, не удовлетворяющий формалистов, так как он не является «финитным». На сегодняшний день положение с поисками доказательства непротиворечивости арифметики следующее. Гёдель показал, что не существует финитного доказательства, выразимого в самой арифметике; Генцен дал доказательство, но при этом он пользовался такими методами, что это доказательство можно рассматривать как неудовлетворительное (недостоверное); и, наконец, вопрос о том, существует ли доказательство непротиворечивости, которое, хотя и не выразимо в арифметике, но тем не менее «финитно», остается открытым.

§ 5.2. Некоторые факты из теории рекурсивных функций

Мое изложение логики и теоремы Гёделя в §§ 5.3—5.5 во многом носит нестрогий, неформальный характер. Строгое формальное изложение этих вопросов

*) Kurt Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I (часть II не была опубликована), Monats Math. Phys. 38 (1931), 173—198.

можно найти в гл. 8 книги Мартина Дэвиса *). Кроме того, я старался подражать превосходному стилю ясной статьи Э. Поста **). Позвольте мне повторить заявление, сделанное мною в § 1.6. Мои рассуждения будут вполне строги, за исключением следующего: если интуитивно ясно, что данная процедура эффективна, то я буду считать, что существует соответствующая машина Тьюринга. Замечу, что во всех таких случаях имеются строгие доказательства, например, в книге М. Дэвиса. Приняв это соглашение, мы ввели следующие определения на языке наших эффективных процедур.

Определение 1.6.1. Функция называется *рекурсивной*, если существует эффективная процедура для ее вычислений.

Определение 1.6.2. Множество называется *рекурсивным*, если существует эффективная процедура для выяснения того, принадлежит или не принадлежит произвольный элемент к этому множеству.

Определение 1.6.3. Множество называется *рекурсивно перечислимым*, если существует эффективная процедура для последовательного порождения (перечисления) его элементов.

Если мне потребуется убедить Вас в том, что существует программа для выполнения определенных вычислений (сколь угодно длинная и работающая достаточно долго), то я просто представлю Вам блок-схему вычислений, а не программу на языке машины. Ниже мы останавливаемся только на существенных вопросах опуская детали, подробное рассмотрение которых могло бы затемнить основное содержание.

Мы доказали три теоремы о рекурсивных и рекурсивно перечислимых множествах.

Теорема 1.6.1. *Если R и S — рекурсивно перечислимые множества, то рекурсивно перечислимы множества $R \cup S$ и $R \cap S$.*

---

*) M. Davis, Computability and unsolvability, McGraw-Hill Book Company, Inc., New York, 1958

) E. Post, Recursively enumerable Sets and their decisions problems, Bull. Amer. Math. Soc. **50, № 5 (1944), 284—316.

Теорема 1.6.2. *Множество положительных целых чисел S рекурсивно тогда и только тогда, когда S и $\bar{S}$ рекурсивно перечислимы.*

Теорема 1.6.3. *Существует рекурсивно перечислимое, но не рекурсивное множество положительных целых чисел.*

Последняя теорема по существу является абстрактной формой теоремы Гёделя. Большая работа, которая проделывается в двух последующих параграфах, заключается в том, чтобы определить и разъяснить понятие адекватной ω-непротиворечивой арифметической логики. После этого уже легко доказывается теорема Гёделя о неполноте: «Каждая адекватная ω-непротиворечивая арифметическая логика неполна».

§ 5.3. Рекурсивные логики

Теперь мы сформулируем в очень общем виде требования, предъявляемые формалистами к логике. Прежде всего, у нас должен быть алфавит $a_0, a_1, a_2, \ldots$, в терминах которого мы могли бы записывать утверждения нашей логики. Конечная последовательность этих символов будет называться «словом» *). Результат приписывания (справа) к слову X слова Y будем записывать XY.

Под *высказыванием* мы понимаем некоторое утверждение, которое является либо истинным, либо ложным. *Предикатом* называется выражение, содержащее несколько символов (переменных), которое становится высказыванием, если вместо этих символов подставить элементы некоторого специального множества. Если это специальное множество является множеством слов, то предикат называется *словарным*. Предикат называется *рекурсивным*, если существует эф-

*) Отметим, что понятие «слово» у нас употребляется несколько необычно. В нашей терминологии фразу Cogito ergo sum можно рассматривать как слово, если алфавит имеет «пустую» букву, означающую пропуск между словами. (*Прим. автора.*) Cogito ergo sum — я мыслю, значит я существую — знаменитое положение Декарта. (*Прим. ред.*)

фективная процедура, позволяющая для каждого высказывания, полученного из данного предиката, узнать, истинно оно или ложно.

Например, высказывание «x — человек» является словарным предикатом, и если мы вместо x поставим слово «Джон», то получим истинное высказывание «Джон — человек». Если же вместо x поставить слово «стул», то получится ложное высказывание «стул — человек». В приведенном примере предикат является одноместным; так как он содержит только одну переменную.

Для рекурсивной логики мы требуем, чтобы была эффективная процедура, позволяющая по любому слову сказать, является ли это слово аксиомой или нет, и требуем существования эффективной процедуры, которая позволяла бы узнать, можно ли по данному правилу вывода получить данное слово из других или нет*). Дадим точное определение.

Определение 5.3.1. Под *рекурсивной логикой* L мы понимаем рекурсивное множество слов, называемых *аксиомами* L, и конечное множество рекурсивных словарных предикатов, называемых *правилами вывода* L (ни один из этих предикатов не является одноместным).

Если $R(Y, X_1, \ldots, X_n)$ — правило вывода L, то будем говорить, что Y *непосредственно следует* из $X_1, \ldots, X_n$ по правилу R.

Конечная последовательность слов $X_1, X_2, \ldots, X_m$ называется *выводом* (слова X_m) в L, если для любого i, $1 \leqslant i \leqslant m$, либо X_i является аксиомой, либо существуют такие $j, \ldots, k < i$, что X_i непосредственно следует из $X_j, \ldots, X_k$ по одному из правил вывода логики L. Каждое X_i, $i = 1, 2, \ldots, m$, называется *шагом вывода*. Будем говорить, что W — *теорема* или что W *доказуемо* в L, и записывать

$$T(L): W,$$

*) Для сравнения различных способов гарантии такой эффективности см. M. A. Arbib, Monogenic Normal Systems are Universal, J. Australian Math. Soc. 3, № 3 (1963), 301—306.

если существует вывод W в L. Читатель найдет пример такого вывода при доказательстве леммы 5.4.1 в следующем параграфе.

Обозначим через T_L множество всех теорем L. Заметим, что каждая аксиома является теоремой L.

Теорема 5.3.1. *Множество T_L рекурсивно перечислимо.*

Доказательство. Укажем эффективный метод порождения всех теорем. Так как множество аксиом рекурсивно, мы можем их эффективно перенумеровать $A_1, A_2, A_3, \ldots$ Для каждого натурального n читатель может проверить, что выводы, которые имеют не более чем n шагов, и в которых использованы только аксиомы $\{A_1, A_2, \ldots, A_n\}$, можно получить эффективно, т. е. эти выводы образуют рекурсивно перечислимое множество. Отсюда легко следует утверждение теоремы.

Если мы хотим в нашей логике говорить о подмножестве натуральных чисел Q, мы должны для каждого натурального n иметь эффективный способ выписывания слова W_n, означающего $n \in Q$.

Если мы можем доказать, что W_n является теоремой L в точности тогда, когда $n \in Q$, то будем говорить, что в L имеется *полуполное* описание Q. Описание Q называется *полным*, если мы также имеем выражение $\overline{W}_n$, которое интерпретируется $n \notin Q$, и можем доказать, что $\overline{W}_n$ является теоремой L в точности тогда, когда $n \notin Q$ *). Другими словами, L содержит полное описание Q, если в L для каждого n мы можем установить, принадлежит оно Q или нет. Введем определение.

Определение 5.3.2. Логика L называется *полуполной* для множества целых чисел Q, если существует рекурсивно перечислимое множество слов W_0, W_1, $W_2, \ldots$ такое, что

$$Q = \{n \mid T(L)\colon\ W_n\}.$$

*) Если Q имеет полуполное описание в L, то его часто называют *перечислимым в* L. Если Q имеет полное описание в L, то оно называется *разрешимым* в L. (*Прим. ред.*)

L называется *полной* для Q, если она полуполная для Q и $\overline{Q}$.

Теорема 5.3.2. *Если L полуполная для Q, то Q рекурсивно перечислимо.*

Доказательство. $Q=\{n \mid T_L \cap \{W_0, W_1, W_2, \ldots\}\}$ является пересечением двух рекурсивно перечислимых множеств и, следовательно, согласно теореме 1.6.1, рекурсивно перечислимо.

Вспомнив теорему 1.6.2, получаем

Следствие 5.3.3. Если Q рекурсивно перечислимое, но не рекурсивное множество, то никакая рекурсивная логика не может быть полной для Q.

От этого следствия до теоремы Гёделя совсем недалеко. Для этого надо средствами некоторой рекурсивной логики развить теорию целых чисел, причем так, чтобы принадлежность чисел данному множеству можно было трактовать адекватно (т. е. число n принадлежит Q тогда и только тогда, когда некоторое эффективно сопоставленное ему слово выводимо в L). Это возможно только тогда, когда Q по крайней мере рекурсивно перечислимо. Следовательно, множества, не являющиеся рекурсивно перечислимыми, в лучшем случае могут быть представлены в рекурсивной логике неполно.

Чтобы объяснить интерес лиц, занимающихся теорией автоматов, к множествам, которые не являются рекурсивно перечислимыми, оставим на минуту теорему Гёделя и обсудим *проблему останова машин Тьюринга.* Напомним (см. § 1.5), что если мы установим машину Тьюринга в начальное состояние q_0 и на ее ленте запишем $\bar{n}$, то она начнет считывать с ленты и записывать на нее, двигать ленту, но у нас нет гарантии, что она остановится (например, машина Тьюринга, которая сдвигает ленту вправо и больше ничего не делает).

Проблема останова для машины Тьюринга Z состоит в том, что для каждого числа n требуется определить, остановится Z или нет, если она начинает работу в состоянии q_0, наблюдая самый левый квадрат ленты $\bar{n}$. Проблема останова называется *разрешимой*, если существует *эффективный метод* для ее решения.

Если через $\overline{R}_Z$ обозначить множество тех n, для которых Z *не* останавливается, то вопрос «Разрешима ли проблема останова для Z?» эквивалентен вопросу «Является ли $\overline{R}_Z$ рекурсивным множеством?». Ясно, что для лиц, занимающихся теорией автоматов, $\overline{R}_Z$ представляет интерес. Я покажу, что существует такая машина Тьюринга Z, для которой проблема останова неразрешима эффективно, и следовательно, $\overline{R}_Z$ не является даже рекурсивно перечислимым.

Отметим, что R_Z (множество тех n, для которых Z *останавливается*) всегда является рекурсивно перечислимым: на n-м шаге, $n=1, 2, 3, \ldots,$ мы включаем в R_Z те из первых n чисел, для которых машина Z останавливается за n тактов. Поэтому, если $\overline{R}_Z$ не рекурсивно, то и R_Z не рекурсивно, а так как оно рекурсивно перечислимо, то $\overline{R}_Z$ не является рекурсивно перечислимым (см. теорему 1.6.2). Значит, чтобы найти Z, у которой $\overline{R}_Z$ не является рекурсивно перечислимым множеством, надо найти Z, для которой R_Z не рекурсивное множество. Но это мы немедленно получим (см. теорему 1.6.3), если сможем показать, что каждое рекурсивно перечислимое множество является множеством R_Z для некоторой машины Z. Используя тезис Тьюринга (см. § 1.6), мы получаем, что для любого данного рекурсивно перечислимого множества R можно найти машину Тьюринга $Z(R)$, которая эффективно выполняет следующую процедуру: для данного числа n машина последовательно порождает элементы $x_1, x_2, \ldots,$ из R до тех пор, пока не найдется x_i, равное n. Тогда машина стирает n и останавливается.

Ясно, что $Z(R)$ останавливается только тогда, когда $n \in R$, т. е. $R=R_{Z(R)}$. Поэтому, если R — рекурсивно перечислимое, но не рекурсивное множество, то $\overline{R}_{Z(R)}=\overline{R}$ не является рекурсивно перечислимым и $Z(R)$ имеет неразрешимую проблему останова.

Значит, существуют множества, которые не являются рекурсивно перечислимыми, но тем не менее представляют интерес для лиц, занимающихся теорией автоматов. Следовательно, мы нашли разные степени неполноты в нашей рекурсивной логике.

§ 5.4. Арифметические логики

Этот параграф посвящен разъяснению и точному определению всех понятий, содержащихся в формулировке теоремы Гёделя о неполноте: *«Каждая ω-непротиворечивая и адекватная арифметическая логика неполна»*.

Рекурсивная логика L называется *арифметической логикой*, если она обладает свойствами, перечисленными ниже в пп. 1, 2, 3 и 4. Каждый из этих пунктов разделен на две части. В части (а) даются содержательные разъяснения, а в части (б) — формальные требования.

1. Правильно построенные формулы

(а) У нас должен быть эффективный способ, позволяющий для любой строчки символов определить, имеет она смысл или нет, прежде чем говорить об ее истинности или ложности. Понятие «правильно построенная формула» (п. п. ф.) родственно понятию грамматически правильно построенного предложения в русском языке. При такой интерпретации логики п. п. ф. содержат предложения, которые представляют высказывания или предикаты. Следовательно, обе фразы «x — человек» и «Сократ — человек» являются п. п. ф. русского языка, а «человек не» — не п. п. ф. русского языка.

(б) Для каждой арифметической логики L имеется непустое рекурсивное множество слов, называемых п. п. ф. логики L. Все теоремы L являются п. п. ф.

2. Пропозициональные связки

(а) Если нам даны п. п. ф. A и B, мы хотим иметь возможность комбинировать их различными способами, для того чтобы получать новые п. п. ф., такие, как

$\sim A$	не A
$A \supset B$	A влечет B
$A \& B$	A и B
$A \vee B$	A или B (или оба)
$A \equiv B$	A тогда и только тогда, когда B.

Логические системы, которые мы рассматриваем, являются *двузначными*, т. е. любое *высказывание* является либо истинным, либо ложным (*предикаты* принимают единственное истинностное значение при замене всех переменных константами). Мы можем рассматривать приведенные выше *пропозициональные связки* как двузначные функции (см. булевы функции, § 3.1), т. е. $A \& B$ истинно, когда A истинно и B истинно; $A \& B$ ложно в остальных случаях.

Мы можем выразить это при помощи «истинностной таблицы», позволяющей определить истинностное значение $A \& B$ по известным истинностным значениям A и B *).

A	B	$A \& B$
и	*и*	*и*
и	*л*	*л*
л	*и*	*л*
л	*л*	*л*

Мы требуем, чтобы $A \supset B$ было истинным, если A и B истинны. Мы также требуем, чтобы если A истинно и $A \supset B$ истинно, было истинным и B. Схема, задающая истинностные значения $A \supset B$, которые удовлетворяют этим требованиям, такова:

A	B	$A \supset B$
и	*и*	*и*
и	*л*	*л*
л	*и*	*и*
л	*л*	*и*

*) См., например, W. V. O. Quine, Mathematical logic, rev. ed. Harvard Univ. Press, Cambridge, Mass., 1958. (*Прим. автора.*) См. также П. С. Новиков, Элементы математической логики, Физматгиз, 1959. (*Прим. ред.*)

Мы также имеем

A	$\sim A$
и	*л*
л	*и*

и

A	B	$A \vee B$	$A \equiv B$
и	*и*	*и*	*и*
и	*л*	*и*	*л*
л	*и*	*и*	*л*
л	*л*	*л*	*и*

Теперь легко проверить, что все наши пропозициональные связки могут быть построены из $\sim$ и $\supset$:

$$A \& B = \sim [A \supset \sim B],$$

$$A \vee B = \sim A \supset B,$$

$$A \equiv B = [A \supset B] \& [B \supset A] = \sim [[A \supset B] \supset \sim [B \supset A]].$$

Следовательно, если мы введем одни только связки $\sim$ и $\supset$ и при помощи подходящих аксиом установим их свойства, то наша логика соответственно может быть дополнена другими пропозициональными связками. При этом мы, конечно, должны потребовать, чтобы эти связки сохраняли свойство правильного построения формул *).

б) Задаются рекурсивные словарные функции: $[A \supset B]$ от A и B и $\sim A$ от A. $\sim A$ есть п. п. ф. тогда и только тогда, когда A есть п. п. ф. $[A \supset B]$ есть п. п. ф. тогда и только тогда, когда и A и B являются п. п. ф. Если A, B и C — п. п. ф. в L, то мы имеем следующие аксиомы:

$$T(L)\colon\ [A \supset [B \supset A]], \tag{5.4.1}$$

$$T(L)\colon\ [[A \supset [B \supset C]] \supset [[A \supset B] \supset [A \supset C]]],$$

$$T(L)\colon\ [[\sim B \supset \sim A] \supset [A \supset B]]. \tag{5.4.2}$$

*) То есть, чтобы формулы с этими связками также были п. п. ф. (*Прим. ред.*)

Кроме того, мы используем следующее правило вывода:

Если $T(L)$: A *и* $T(L)$: $[A \supset B]$, *то* $T(L)$: B.

Это правило называется *modus ponens*.

3. Кванторы

(а) В излагаемой ниже теории чисел мы должны иметь в своем распоряжении числовые переменные. Для этого мы введем последовательность $x_1, x_2, x_3, \ldots$ переменных, которые в качестве значений могут принимать натуральные числа.

Для данной п. п. ф. A мы хотим иметь возможность сказать, что она выполняется, например, при *всех* возможных значениях переменной x_1, или *по крайней мере для одного* ее значения. Тогда нам нужны:

Квантор всеобщности: $(x_1)A$: A истинна для всех значений x_1.

Квантор существования: $(\exists x_1)A$: Существует по крайней мере одно значение x_1, для которого A истинна.

Если переменная в п. п. ф. находится под знаком квантора (всеобщности или существования), то мы говорим, что она *связана*. В противном случае мы говорим, что переменная *свободна*. Например, x связана в (истинном) высказывании $(x)[x+2>x]$, но свободна в предикате $[x+2>x]\&(y)\,[y>0]$.

Введем обозначение $B(M, A)$, которое читается: «M связано в A». Мы скажем, что п. п. ф. *замкнута*, если никакая переменная не является в ней свободной. При будущей интерпретации замкнутые п. п. ф. будут представлять собой высказывания, а не замкнутые п. п. ф. — предикаты. Если мы имеем предикат, то можем получить из него высказывание, связав все его свободные переменные кванторами всеобщности. Таким образом, из $[x>y]$ получаем (ложное) высказывание $(x)\,(y)[x>y]$. Замкнутая п. п. ф. $C(W)$, полученная из данной п. п. ф. W, называется замыканием W. Далее, если X есть п. п. ф. и M есть переменная, свободная в X, мы хотим иметь возможность делать подстановку. Обозначим через $S(X, M, N)$ слово, которое получается в результате замещения переменной M на N во всех вхождениях M в X.

(б) Имеется рекурсивно перечислимая последовательность x_1, x_2, x_3, ... различных слов в L, называемых *переменными.* Никакая переменная не является п. п. ф.

Имеется бинарная словарная рекурсивная функция $M(A)$. $M(A)$ является п. п. ф. тогда и только тогда, когда A — п. п. ф. и M — переменная. Под $(\exists M)A$ мы понимаем $\sim(M)\sim A$.

Имеется бинарный словарный рекурсивный предикат $B(M, A)$. Если $B(M, A)$ истинный, то A есть п. п. ф., M — переменная и M есть часть A (т. е. $A=BMC$ при некоторых B и C). Когда $B(x_1, A)$ истинно, мы говорим, что x_1 *связано* в A. Когда $B(x_1, A)$ ложно и x_1 является частью A, мы говорим, что x_1 *свободно* в A.

Если никакая переменная не является свободной ни в какой п. п. ф. B, которая является частью п. п. ф. A, то мы говорим, что A *замкнута.* Мы пишем *з. п. п. ф.* для замкнутых п. п. ф. $B(x_i, (x_i)A)$ всегда истинна.

Если $B(x_i, A)$ и A есть часть п. п. ф. C, то $B(x_i, C)$. Существует словарная рекурсивная функция $S(X, M, N)$ такая, что $S(X, M, N)$ является словом, полученным в результате замещения M на N во всех его вхождениях в X. Таким образом,

$$S([X\supset Y], M, N) \quad \text{есть} \quad [S(X, M, N)\supset S(Y, M, N)]$$

и

$$S(\sim X, M, N) \quad \text{есть} \quad \sim S(X, M, N).$$

Если $i\neq j$, то

$$S((x_j)X, x_i, N) \quad \text{есть} \quad (x_j)S(X, x_i, N).$$

Имеется еще одна словарная рекурсивная функция $C(W)$, которая называется *замыканием* W. Если W есть п. п. ф., то $C(W)$ есть з. п. п. ф. и получается дописыванием спереди W нескольких кванторов общности (возможно ни одного). Если W есть з. п. п. ф., то $C(W)=W$. Если W не п. п. ф., то и $C(W)$ не п. п. ф.

Отсюда следует, что *класс з. п. п. ф. является рекурсивным*, так как W есть з. п. п. ф. тогда и только тогда, когда $C(W)=W$, и существует рекурсивная процедура для распознавания равенства.

4. *Целые числа*

(а) Наше последнее требование к арифметической логике состоит в том, что мы должны иметь возможность говорить в ней о целых числах. Для этой цели мы хотим иметь в нашей логике слова, называемые *нумералами*, которые бы служили *названиями* целых чисел. Идея введения нумералов преследует цель — разделить *вещи* и *имена этих вещей*. П. п. ф. в нашей логике — это не сами числа как таковые, а нумералы, которые являются именами этих чисел.

(б) С каждым целым числом n рекурсивно связываем слово (обозначаем его n^*), которое мы называем нумералом, связанным с числом n. Если $n \neq m$, то $n^* \neq m^*$.

Если A есть п. п. ф., x_i свободно в A и N — переменная или нумерал, то $S(A, x_i, N)$ есть п. п. ф. Если x_i не связано в A, то

$$T(L)\colon\ [S(A,\ x_i\ m^*) \supset (\exists x_i)\, A]. \qquad (5.4.3)$$

Этим завершено наше определение арифметической логики. Таким образом, *арифметическая логика* является *рекурсивной логикой*, снабженной точными критериями *правильного построения* формул с пропозициональными *связками*, с *кванторами* общности и существования и с *числами.*

Посвятим оставшуюся часть этого параграфа понятиям ω-непротиворечивости, адекватности и полноты.

Определение 5.4.1. П. п. ф. W в L называется *n-арной*, если переменные $x_1, \ldots, x_n$ свободны в W и никакие другие переменные не являются свободными в W. Если W — n-арная п. п. ф. в L и $(y_1, \ldots, y_n)$ — n-ка чисел, то $W(y_1^*, \ldots, y_n^*)$ обозначает п. п. ф., полученную из W замещением x_i на y_i^* во всех вхождениях x_i в W, $i=1, 2, \ldots, n$.

Мы называем логику непротиворечивой, если она свободна от противоречий.

Определение 5.4.2. Арифметическая логика называется *непротиворечивой*, если ни для какого слова A не может быть $T(L): A$ и $T(L): \sim A$ (другими словами, если нет такого выражения, для которого средствами этой логики можно было бы получить его и его отрицание).

Теперь предположим, $T(L): W(m^*)$ для $m=0, 1, 2, 3, \ldots$ Даже если в L невозможно доказать $(x_i)W$, мы, конечно, можем считать это следствием приведенного выше списка теорем. Следовательно, мы могли бы рассматривать $T(L): \sim (x_i)W$ как противоречие. Таким образом, ω-непротиворечивость является так же желательной, как и просто непротиворечивость, где

Определение 5.4.3. Арифметическая логика L называется *ω-непротиворечивой*, если в ней нет такой унарной п. п. ф., что для всех целых чисел m $T(L)$: $W(m^*)$ и $T(L)$: $\sim (x_i)W$.

Лемма 5.4.1. *L противоречива тогда и только тогда, когда все п. п. ф. в L являются теоремами.*

Доказательство. Пусть $T(L): A$, $T(L): \sim A$ и B — произвольная п. п. ф. в L. Покажем, что $T(L): B$.

Из (5.4.1) $T(L): [\sim A \supset [\sim B \supset \sim A]]$.

По модус поненс $T(L): [\sim B \supset \sim A]$.

Из (5.4.2) $T(L): [[\sim B \supset \sim A] \supset [A \supset B]]$.

По модус поненс $T(L): [A \supset B]$.

И наконец $T(L): B$.

Обратное утверждение (т. е. если все п. п. ф. являются теоремами, то L противоречива) тривиально (см. определение).

Следствие 5.4.2. Если L ω-непротиворечива, то она непротиворечива.

Скажем, что предикат P *вполне представим**) в логике, если мы средствами этой логики всегда можем решить, является он истинным или ложным после подстановки констант (на места свободных переменных).

---

*) Нумерически выразим (по Клини). (*Прим. ред.*)

Определение 5.4.4. Пусть $P(x_1, \ldots, x_n)$ — предикат. Скажем, что P *вполне представим* в арифметической логике L, если существует n-арная п. п. ф. W в L такая, что

а) Для каждой n-ки чисел $(y_1, \ldots, y_n)$, для которой $P(y_1, y_2, \ldots, y_n)$ истинный, $T(L)$: $W(y_1^*, \ldots, y_n^*)$.

б) Для каждой n-ки чисел $(y_1, \ldots, y_n)$, для которой $P(y_1, \ldots, y_n)$ ложен, $T(L)$: $\sim W(y_1^*, \ldots, y_n^*)$.

Попробуем теперь определить, какую логику мы будем называть адекватной. Мы хотели бы иметь возможность отвечать на вопросы о рекурсивно перечислимых множествах в такой логике, так как мы знаем, что это такие множества, которые могут быть порождены механическими процессами. Поэтому они представляют интерес для специалистов, строящих различные модели, занимающихся теорией автоматов, для биологов и инженеров. В § 5.3 мы показали, что множество чисел Q может быть описано в рекурсивной логике L только тогда, когда оно рекурсивно перечислимо. В этом случае мы можем надеяться, что она является полуполной для Q, т. е. существует рекурсивно перечислимое множество слов $W_0, W_1, W_2, \ldots$ такое, что

$$Q = \{n \mid (L): W_n\}.$$

Если мы хотим назвать нашу логику адекватной, мы можем потребовать, чтобы она была полуполной для каждого рекурсивно перечислимого множества, т. е. чтобы она представляла каждое множество, которое *может* быть представлено в рекурсивной логике!

При доказательстве теоремы 5.3.1 мы отметили, что выводы рекурсивной логики могут быть эффективно перенумерованы. Помня об этой нумерации, мы можем говорить об y-м выводе (выводе номера y). Возьмем наше множество Q и соответствующее ему множество слов $\{W_0, W_1, W_2, \ldots\}$. Рассмотрим следующий предикат $R(x, y)$: «y — номер вывода W_x». Если $R(x, y)$ истинный, то y является номером вывода о том, что x принадлежит Q. Теперь, взяв x и y, мы можем эффективно построить W_x и эффективно породить y-й вывод, а следовательно, эффективно узнать, является $R(x, y)$ истинным или нет. Следовательно, $R(x, y)$ — такой рекурсивный предикат, что

$$Q = \{x \mid \textit{существует такое } y, \textit{ что } R(x, y) \textit{ истинный}\}.$$

Если мы добавим требование, чтобы каждый *рекурсивный* предикат был *вполне представимым*, то у нас будет достаточно оснований для формального введения следующего определения адекватности (по Дэвису).

Определение 5.4.5. Арифметическая логика L называется *адекватной*, если для каждого рекурсивно перечислимого множества Q существует вполне представимый в L предикат $R(x, y)$ такой, что $Q=\{x \mid$ существует такое y, что $R(x, y)$ истинный$\}$.

И, наконец, последнее определение в этой книге. Логика называется полной, если в ней можно доказать истинность или ложность каждого утверждения. Более формально

Определение 5.4.6. Арифметическая логика L называется *полной*, если для каждой з. п. п. ф. A в L либо

$$T(L)\colon\ A \textit{ либо } T(L)\colon\ \sim A.$$

Простое доказательство теоремы Гёделя о неполноте дано в следующем параграфе.

§ 5.5. Доказательство теоремы Гёделя о неполноте

Теорема 5.5.1 (Гёдель). *Если L является ω-непротиворечивой и адекватной арифметической логикой, то L неполна.*

Доказательство. Согласно теореме 1.6.3 выберем такое Q, которое является рекурсивно перечислимым, но не рекурсивным множеством. Так как L адекватна, то существует предикат $R(x, y)$, вполне представимый в L, такой, что

$$Q = \{x \mid \text{существует } y \text{ такое, что } R(x, y) \text{ истинный}\}. \tag{5.5.1}$$

Значит, существует бинарная п. п. ф. W такая, что когда $R(x, y)$ истинный,

$$T(L):\ W(x^*, y^*).$$

Когда $R(x, y)$ ложный,

$$T(L):\ \sim W(x^*, y^*).$$

Пусть U — унарная п. п. ф. $(\exists x_2) W$. Очевидно, что

$$Q = \{x \mid T(L):\ U(x^*)\}. \tag{5.5.2}$$

Для каждого целого числа n запишем $W(n^*, x_2)$ для

$$S(W, x, n^*).$$

Предположим, что $n_0 \in Q$. Тогда найдется y_0 такое, что $R(n_0, y_0)$ истинный (согласно 5.5.1). Следовательно, $T(L)$: $W(n_0^*, y_0^*)$. Но по (5.4.3)

$$T(L):\ [W(n_0^*, y_0^*) \supset \exists x_2 W(n_0^*, x_2)]$$

и по модус поненс

$$T(L):\ (\exists x_2)\, W(n_0^*, x_2),$$

а следовательно,

$$T(L):\ U(n_0^*).$$

Предположим обратное, т. е. $T(L): U(n_0^*)$. Тогда

$$T(L):\ (\exists x_2)\, W(n_0^*, x_2)$$

или

$$T(L):\ \sim (x_2) \sim W(n_0^*, x_2).$$

Так как L является ω-непротиворечивой, то существует такое целое число m_0, что п. п. ф. $\sim W(n_0^*, m_0^*)$ не является теоремой L. Следовательно, значение $R(n_0, m_0)$ — «истина» (если бы оно было «ложь», то мы бы имели $T(L)$: $\sim W(n_0^*, m_0^*)$, а значит, $n_0 \in Q$). Таким образом, мы доказали

$$Q = \{x \mid T(L):\ U(x^*)\}.$$

Теперь предположим, что $T(L)$: $\sim U(n_0^*)$. Значит, $n \in \overline{Q}$ (если бы $n \in Q$, то мы имели бы $T(L): U(n^*)$

и логика L была бы противоречивой). Таким образом, $\{x \mid T(L): \sim U(x)\}$ является подмножеством множества $\bar{Q}$. Согласно следствию 5.3.3

$$\{x \mid T(L): \sim U(x^*)\} \neq \bar{Q},$$

так как $\bar{Q}$ не есть рекурсивно перечислимое множество. Значит, существует такое число n_0, что $n_0 \in \bar{Q}$, и в то же время неверно, что $T(L)$: $\sim U(n_0^*)$. С другой стороны, неверно также, что $T(L)$: $U(n_0^*)$, ибо из этого следует, что $n_0 \in Q$. Следовательно, $U(n_0^*)$ и $\sim U(n_0^*)$ не выводимы в L и, значит, L не полна, что и требовалось доказать.

Посмотрим теперь на наше высказывание $\sim U(n_0^*)$. Согласно (5.5.1) его можно интерпретировать как $n_0 \in \bar{Q}$ и, следовательно, оно *обязательно является «истинным» высказыванием.* И все же оно *не является теоремой* L, т. е. мы показали, что если L ω-непротиворечива, то существует истинное высказывание, не выводимое в L. Изменить это положение, добавляя $U(n_0^*)$ к списку аксиом, не удается, так как для вновь полученной логики L' опять верны все те обстоятельства, которые выше привели нас к теореме Гёделя. Значит, мы снова можем найти такое число n_1, что высказывание $\sim U(n_1^*)$ истинно, но не является теоремой L', и т. д.

§ 5.6. Мозг и машина

Э. Нагел и Дж. Ньюман в своей книге «Теорема Гёделя» *) утверждают, что теорема Гёделя указывает пределы математических возможностей машин. Приведем несколько цитат:

«В связи с теоремой Гёделя возникает вопрос, можно ли построить вычислительную машину, которая с точки зрения математических способностей равнялась бы человеческому мозгу. Современные вычислительные машины содержат конечное число заложенных

*) E. Nagel and J. Newman, Gödel's proff, New York Univ. Press, New York, 1958.

в них команд, которые соответствуют фиксированным правилам вывода формализованной аксиоматической процедуры. Эти машины дают ответы на вопросы, функционируя согласно заложенным в них программам. Однако, как показал Гёдель в своей теореме о неполноте, существует бесконечное число проблем элементарной теории чисел, решение которых невозможно никаким данным аксиоматическим методом, и на которые машины не смогут дать ответа, как бы совершенны ни были их конструкции и с какой бы скоростью они ни выполняли операции. Человеческий мозг, возможно, также ограничен, но теорема Гёделя указывает, что структура и мощь человеческого ума являются более сложными и тонкими, чем строение и функционирование любой машины, какую мы только можем себе вообразить в настоящее время.»

Процитировав Нагеля и Ньюмана, я думаю, что по справедливости, надо процитировать также и тех авторов, которые с ними не согласны. Обе приведенные ниже цитаты взяты из книги «Возможности ума» *). Хилари Путнам (Hilary Putnam) в статье «Ум и машины» («Minds and machines») пишет:

«Пусть T — машина Тьюринга, которая представляет меня в том смысле, что T может доказать в точности все те утверждения, которые могу доказать я. Тогда утверждают (Нагел и Ньюман не приводят этого аргумента, но я думаю, что именно это они имели в виду), что используя метод Гёделя, я могу обнаружить высказывание, которое T не может доказать, и, более того, я могу доказать это высказывание. Это опровергает утверждение, что T «представляет» меня и, следовательно, я не являюсь машиной Тьюринга. Здесь имеет место элементарная ошибка, заключающаяся в неправильном применении теоремы Гёделя. Для произвольной данной машины T я могу найти такое U, для которого я могу доказать:

если T — непротиворечивая машина, то U истинно, причем U неразрешимо посредством T, если T дей-

*) «Dimensions of Mind», Sydney Hook (ed.), Collier Books, a division of Growell-Collier Publishing Co., New York, 1961.

ствительно непротиворечивая. Однако T также может отлично доказать это, т. е. T может доказать, что U в ней неразрешимо, и если T — непротиворечивая, то U — «истинно» при данной программной интерпретации. И утверждение U, которое T не может доказать (предполагая непротиворечивость), я также не могу доказать (если я не докажу, что T непротиворечива, что мало вероятно, если T достаточно сложная машина)!»

По-моему, это опровержение Нагела и Ньюмана не вполне удовлетворительно. Более удовлетворительное возражение приводит Michael Scriven «The Complet Robot: «A Guide to Androidology». «Нагела и Ньюмана вдохновил тот факт, что какие бы аксиомы и правила вывода мы ни задали вычислительной машине, есть явные математические истины, которые она никогда не сможет «получить» из этих аксиом по данным правилам вывода. Это верно, но их предположение, что это можем сделать мы, имея адекватное данной машине понятие математической истины, ее аксиомы и правила вывода, неверно. В таком случае можно предположить, что формалисты были правы, и они могли показать неправоту Гёделя. Теорема Гёделя не является бóльшим препятствием для машин, чем для нас самих. Мы можем только сказать, что математикам было бы намного легче, если бы формалисты были правы, и в этом случае можно было бы сравнительно просто построить механического математика. Но они неправы, и это невозможно. Точно так же как мы распознаем истинность невыводимой формулы, сравнивая то, что она утверждает, с тем, что мы знаем; это может делать и вычислительная машина».

Любой человек, знакомый с вычислительными машинами, хорошо знает, что при современном уровне программирования их «интеллект» не идет ни в какое сравнение с человеческим. Приведенные выше аргументы нельзя рассматривать как доводы в пользу большого «интеллекта» современных машин. Они лишь утверждают, что *теорема Гёделя не дает доказательства невозможности «интеллектуальной» машины.*

Наше знакомство с персептроном (§ 2.2) и краткие сведения об искусственном интеллекте (§ 4.5) должны навести нас на мысль, что многие ограничения «интеллекта» современных вычислительных машин могут быть преодолены за счет более совершенных конструкций и программирования. Пример персептрона показывает, что машина может приспосабливаться. При обсуждении искусственного интеллекта мы видели, что машина может быть «творческой». Примеры следящих систем показывают, что машина может иметь целенаправленное поведение. Конечно, все эти действия очень примитивны по сравнению с действиями человека, но все же ясно, что многие различия между человеком и машиной, которые до последнего времени казались очень существенными, являются только количественными.

Заключение

В этой книге мы охватили много экспериментальных данных и основательное количество более или менее связанных с ними теорий. Эта область науки огромна, нова и очень быстро развивается. Почти так же быстро, как сама наука, растет число названий, обозначающих саму науку или ее части: кибернетика, бионика, нейродинамика, теория самоорганизующихся систем, «искусственный интеллект» и т. д. Заканчивая книгу, я выделю из всего этого изобилия имен слово «кибернетика» и буду обозначать им все содержание этой книги, а не только те направления, которыми мы интересовались в главе 4. Познания о кибернетике *в этом широком смысле*, которые дает эта книга, легко могут привести к чрезмерной самонадеянности. Поэтому я думаю, что неплохо дать читателю несколько предостережений:

1. Отбор материала, содержащегося в этой книге, в силу самой его природы мог быть только субъективным. Ни одну из рассматриваемых здесь теорий нельзя считать последним словом в этой области. Я могу только надеяться, что дал не слишком предвзятый обзор предмета и что эта книга послужит твердой основой для дальнейшего чтения.

2. Кибернетика — это наука, возникшая на стыке многих дисциплин. Поэтому, в лучшем случае, она дает ученому удовольствие работать одновременно в нескольких областях таких, как техника, психология, математика и физиология. В худшем же случае она позволяет не очень хорошему инженеру уйти от выполнения своих обязанностей, некомпетентно занимаясь теоретической биологией; его гордость при этом не страдает, так как он слишком мало знает

биологию, чтобы понять, каким дураком он выглядит. Чтобы кибернетика могла избежать налета псевдонауки, люди, работающие в ней, должны быть более чем компетентны, по крайней мере в одном из разделов современной науки и должны быть действительно хорошо знакомы с одним или более из соседних разделов. Коллективы узких специалистов, слишком игнорирующие соседние специальности, чтобы эффективно общаться друг с другом, не подходят для этой цели.

3. Читатель, который решит погрузиться в массу специальной литературы (в частности, отчеты о многочисленных симпозиумах), должен быть настроен критически, потому что многие статьи содержат чрезмерные претензии, технические ошибки и незнание смежных областей.

4. В первом порыве возбуждения, вызванном кибернетикой, многие лица слишком смело предсказывали, что к 1960 г. кибернетика достигнет многих захватывающих результатов, таких, как машинный перевод и создание искусственного мозга. Такой сверхоптимизм должен был привести к обратной реакции, самым выдающимся примером которой является, по-видимому, книга Мортимера Таубе «Вычислительные машины и здравый смысл» *). Реакция Таубе настолько сильна, что он сравнивает кибернетику как науку с астрологией! Несомненно, такая позиция дала бы неспециалисту очень искаженное представление о кибернетике. Тем не менее я рекомендую книгу Таубе вашему критическому вниманию. Ибо, когда вы сможете согласиться с его справедливыми критическими высказываниями и *привести собственные аргументы против его несправедливых критических высказываний* (*а таково большинство из них*), тогда вы сможете реалистически увидеть возможности и границы кибернетики.

---

*) Mortimer Taube, Computers and common sense, the myth of Thinking machines, Columbia University Press, New York, 1961. [Русский перевод: М. Таубе, Вычислительные машины и здравый смысл, Изд-во «Прогресс», 1964.]

Несмотря на все предостережения, эта книга подтверждает мою веру в то, что под знаменем кибернетики была проделана большая и хорошая работа. Цитируемая литература содержит основную суть специальных публикаций, действительно добавляющих нечто к нашему пониманию общей основы «мозга, машин и математики», где математика используется, чтобы выявить аналогию между работой мозга и управляющими, вычислительными и информационными аспектами машин. Я верю, что эти публикации образуют ядро новой волнующей и ценной области человеческого познания.

Приложение 1

Основные понятия теории множеств

Под *множеством* мы понимаем совокупность объектов (называемых его *элементами*), которые могут быть точками, числами или состояниями конечного автомата.

Если X состоит из элементов $x_1, x_2, x_3, \ldots$, то иногда пишут $X=\{x_1, x_2, x_3, \ldots\}$.

Через $\varnothing$ обозначается *пустое множество*, т. е. множество, не содержащее ни одного элемента.

Мы используем запись $x \in X$, означающую «x является элементом множества X». Так, например, если X — множество четных чисел, то $2 \in X$, $4 \in X$ и т. д.

Обозначение $\{x \mid y\}$ используется для «множества всех x, для которых истинно утверждение y». Таким образом,

$$X=\{x \mid x \in X\}.$$

Множество четных чисел $=\{x \mid x=2n$ *для некоторого целого* $n\}$.

Если X и Y — два множества, то запись $X \subset Y$ означает, что X является *подмножеством* Y, т. е. каждый элемент из X является элементом Y.

Например, $\varnothing \subset X$ для любого множества X. Если X и Y — два подмножества, то

$X \cap Y =$ *пересечению* X *и* $Y = \{x \mid x \in X$ *и* $x \in Y\} =$
$=$ *множеству элементов, каждый из которых принадлежит как* X, *так и* Y.

$X \cup Y =$ *объединению* X *и* $Y = \{x \mid x \in X$ *или* $x \in Y\} =$
$=$ *множеству элементов, каждый из которых принадлежит по крайней мере к одному из множеств* X *или* Y.

Если X_a — совокупность множеств, то

$$\bigcup \{X_a \mid P(a)\} = \{x \mid \textit{существует } a, \textit{ для которого } P(a) \textit{ истинно и } x \in X_a\} =$$
$$= \textit{объединению тех } X_a, \textit{ для которых } P(a) \textit{ истинно.}$$

Если X и Y — два множества, то *разность* $X \setminus Y = \{x \mid x \in X$ и $x \notin Y\}$, где $x \notin Y$ означает «x не является элементом Y».

Пусть $N = \{0, 1, 2, 3, \ldots\}$ — множество натуральных чисел. Если $X \subset N$, то *дополнение* к X определяется как

$$\overline{X} = N \setminus X =$$
$$= \{x \mid x \textit{ является целым числом, не входящим в } X\}.$$

Обозначение $f: X \to Y$ используется для «функции f, отображающей множество X в множество Y», т. е. «функции f, которая каждому $x \in X$ сопоставляет элемент $f(x) \in Y$». Например, если f определяется как $f(x) = x^2$, то $f: N \to N$.

Если X является подмножеством множества целых чисел, то характеристическая функция $C_X: N \to \{0, 1\}$ множества X определяется так:

$$C_X(x) = \begin{cases} 0, & \text{если } x \notin X, \\ 1, & \text{если } x \in X. \end{cases}$$

Если X и Y — два заданных множества, то *декартово произведение* X и Y

$$X \times Y = \{(x, y) \mid x \in X \text{ и } y \in Y\} =$$
$$= \textit{множеству упорядоченных пар, в которых первый элемент принадлежит } X, \textit{ а второй — } Y.$$

Например, декартова плоскость является прямым произведением прямой x и прямой y. Если $f: X \times Y \to Z$, то пишут $f(x, y)$, а не $f((x, y))$.

Приложение 2

Основные понятия алгебры

Операцией (бинарной), заданной на непустом множестве A, называется функция ω, которая каждой упорядоченной паре $(a, b) \in A \times A$ ставит в соответствие определенный элемент множества A, называемый *результатом* операции на элементах a и b. Очень часто операцию называют умножением и употребляют для нее обычную мультипликативную запись

$$ab = c$$

(«c является результатом операции (умножения) на элементах a и b» или «произведение ab равно c»).

Операция называется *ассоциативной*, если для любых $a, b, c \in A$

$$(ab)c = a(bc).$$

Множество A с заданной на нем ассоциативной операцией называется *полугруппой*. Полугруппа называется *конечной*, если A — конечное множество. Элемент $e \in A$ называется *единичным элементом* (*единицей*) полугруппы, если для каждого элемента $a \in A$

$$ea = ae = a.$$

Пусть A и A' — две полугруппы. Говорят, что полугруппа A *гомоморфно* отображается *на* полугруппу A', если существует отображение $\varphi: A \to A'$ такое, что для каждой пары элементов $a, b \in A$ справедливо равенство

$$\varphi(ab) = \varphi(a)\varphi(b),$$

и при этом каждый элемент A' имеет хотя бы один прообраз в A. В этом случае A' называется *гомоморфным образом* A, а само отображение φ называется *гомоморфизмом*.

Группой называется такая полугруппа, в которой для любой пары элементов a и b уравнения

$$ax = b \quad \text{и} \quad ya = b$$

имеют решения, притом единственные.

Отметим следующие свойства групп (которые легко доказываются из определения группы):

1. В каждой группе существует единичный элемент e.

2. Для каждого элемента a группы G существует один и только один элемент $b \in G$ такой, что

$$ab = e.$$

Для этого элемента справедливо также равенство $ba = e$.

Элемент b называется *обратным элементом* для a и часто обозначается a^{-1}. Ясно, что обратным элементом для a^{-1} служит сам элемент a и что $e^{-1} = e$.

Непустое подмножество H группы G, которое само является группой по отношению к групповой операции G, называется *подгруппой* группы G. Если a — произвольный элемент G, то совокупность всевозможных произведений вида ah, где h пробегает всю подгруппу H, будем обозначать aH. Подгруппа H называется *нормальным делителем* группы G, если для любого $a \in G$ имеет место равенство

$$aH = Ha.$$

Приложение 3

Математика в биологии *)

Эдвард Мур

Биологи используют математический аппарат, но пока сложные системы, изучаемые ими, не поддаются математическому описанию. Теоретический анализ механизмов самовоспроизведения и может привести в будущем к появлению такого описания.

Декарт рассматривал животных и человека, за исключением его души, как механизмы. Однажды, когда он давал уроки шведской королеве Христине, она обратилась к нему с вопросом: «Но как же может машина воспроизводить себя?» Этот же вопрос волнует сегодня и математиков, которые изучают пределы возможностей машин. В своих усилиях создать математическую теорию самовоспроизведения они применяют математические методы для изучения процессов до сих пор считавшихся сугубо биологическими.

Однако это оказалось не так просто. Несмотря на то, что физика и математика бурно развивались последние триста лет, стимулируя и обогащая друг друга, передовая математическая мысль немного сделала для биологии. Я припоминаю несколько исключений. Томас Мальтус создал математическую модель, по которой население увеличивается в геометрической прогрессии, в то время как производство продовольствия возрастает лишь в арифметической, обусловливая «борьбу за существование». Чарльз Дарвин и Альфред Уоллес, ознакомившись с трудами Мальтуса, увидели в этой борьбе проявление закона естественного отбора. Другими словами, чисто математическая

*) Edward Moor, Mathematics in the biological scienses. Scientific America, № 9, 1964.

концепция способствовала развитию центральной идеи биологической эволюции.

И биологии редко удавалось стимулировать развитие математики. В этом смысле стоит отметить изучение генетики популяций. Р. А. Фишер в Англии и Сьюэлл Райт в США создали математические модели, демонстрирующие закономерности совместного действия законов наследственности и случайных факторов, обеспечивающих выживание или гибель определенного гена в популяции. Чрезвычайно сложная модель Райта базировалась на теории диффузии. Это привело Вильяма Феллера из Принстонского университета к изучению новых разделов математики.

Конечно, в своей повседневной работе биологи используют математику. Как и при любых исследованиях, биологи должны подвергать свои результаты определенным статистическим тестам (некоторые из них были разработаны Фишером). Часто им приходится пользоваться и аналитической геометрией, представляя наблюдаемые зависимости в виде кривых. Уравнения термодинамики знакомы биохимикам. Статистика сыграла значительную роль в расшифровке генетического кода и в изучении строения генов. Но все это — традиционная математика. Хотя и было предпринято множество отчаянных попыток создать «математическую биологию», большая их часть, по-видимому, не оправдала первоначальных надежд. Тем не менее, возможно, новейшие математические исследования с использованием вычислительных машин позволят достигнуть бóльшего, чем раньше, когда модели биологических процессов приходилось до крайности упрощать.

Особая ценность математики для биологии заключается не в использовании ее в качестве инструмента, а в ее силе абстракции, позволяющей обнажить фундаментальные проблемы и нащупать взаимосвязи между казалось бы различными сущностями и процессами. Организмы — это машины, хотя и высоко организованные. Мне кажется, что подлинное сотрудничество между математикой и биологией возникнет в результате разработки ряда логических проблем в теории машин, которые, как оказалось, имеют глубокую связь с важнейшими проблемами биологии. Это и будет в основном темой нашего последующего обсуждения.

Клод Е. Шеннон (Массачусетский технологический институт) и Джон Маккарти (Станфордский университет) однажды заметили, что когда люди моделируют человеческий организм, машины, которые они выбирают для этой цели, всегда являются отражением своей эпохи. Декарт сравнивал организм со сложными водяными часами; в начале нашего века мозг рассматривали как коммутатор АТС. Затем наиболее совершенной моделью живого организма стали считать ЭВМ. Возможно, по этой причине большинство ученых, занимающихся сравнительным изучением животных и машин, сосредоточили внимание на центральной нервной системе. Им пришлось задать себе два вопроса: Можно ли рассматривать мозг как своего рода вычислительное устройство? Можно ли создать вычислительную машину, которая бы «мыслила» как мозг?

Возникло несколько направлений исследований, но основой большинства из них является концепция «автомата», под которым понимается идеализированная машина или часть ее. Когда же требуется более полное соответствие с данными нейрофизиологии, за основу берется некоторый идеализированный организм или его часть, например клетка.

В теории автоматов рассматривают не их внутреннее устройство, а наблюдаемые внешние характеристики. По словам покойного Джона фон Неймана элементы машины или организма «...рассматриваются как автоматы, внутреннюю структуру которых не нужно конкретизировать, но которые реагируют совершенно определенным образом на совершенно определенные раздражители».

Одной из самых полезных в этом смысле абстракций является «конечный автомат». Это «черный ящик», имеющий конечное число дискретных внутренних состояний. Обычно он имеет конечное число входов и выходов. Состояние автомата и его выход в любой момент времени T находятся в определенной зависимости от состояния и входа в предыдущий момент времени $T-1$. Имея некоторый конечный

Предыдущее состояние	Текущее состояние	
	0	1
q_1	q_4	q_3
q_2	q_1	q_3
q_3	q_4	q_4
q_4	q_2	q_2

Текущее состояние	Текущий выход
q_1	0
q_2	0
q_3	0
q_4	1

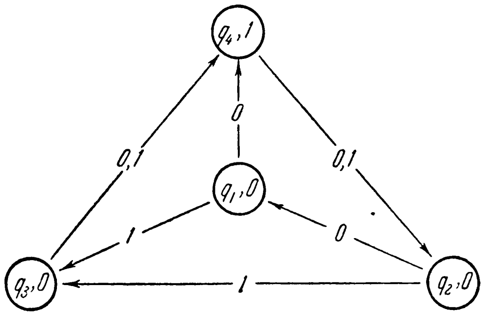


Рис. 1. Конечный автомат это идеализированная машина или часть машины: в теории автоматов это основная составная часть систем, начиная с нервных систем и вычислительных машин и кончая самовоспроизводящимися машинами. Автомат может быть описан двумя таблицами. Первая из них (вверху слева) задает изменение состояния в единицу времени в зависимости от входа. Другая (вверху справа) задает выход по каждому состоянию. Это описание может быть задано диаграммой (внизу). В вершинах указаны состояния и выходы; дуги задают переходы для каждого входа.

автомат и совокупность правил, определяющих его переходы из одного состояния в другое, можно, зная исходное состояние и входную последовательность, определить состояние и выход автомата в любой момент времени. Правила переходов могут быть представлены в виде таблицы или диаграммы (рис. 1).

Имея дело с конечным автоматом, важно ясно представлять себе, что понимается под его состоянием. Состояния конечного автомата, моделирующего работу, например, наборного замка, это не просто «открыт» или «закрыт» — это состояния его различных

Предыдущее состояние	Текущее состояние 0	Текущее состояние 1
q_1	q_2	q_1
q_2	q_1	q_3
q_3	q_4	q_1
q_4	q_1	q_1

Текущее состояние	Текущий выход
q_1	0
q_2	0
q_3	0
q_4	1

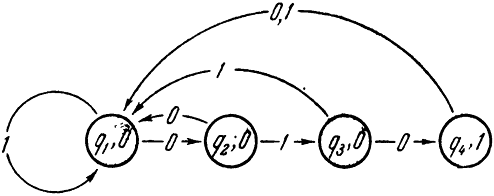


Рис. 2. Наборный замок может быть описан на языке конечных автоматов. Входной последовательностью является комбинация набранных цифр. В этом простом примере такой комбинацией является *0*, *1*, *0*. Выход *1* означает «открыто».

деталей, положения невидимых снаружи шестерен и многочисленных рычажков, меняющиеся с каждым перемещением внешних дисков с цифрами, как показано на рис. 2.

В 1943 г. У. Маккаллок и У. Питтс (Массачусетский технологический институт) создали абстрактную и очень упрощенную модель основного элемента биологической нервной системы — нейрона, или нервной клетки. Фактически это был конечный автомат с двумя возможными состояниями: возбуждения и покоя. Комбинируя эти элементы, или формальные нейроны они строили модель нервной системы. Позже С. К. Клини (Висконсинский университет) доказал общую теорему, дающую возможность предсказать характер поведения, которого можно ожидать от нейронной сети Маккаллока — Питтса.

Теорема Клини справедлива для любых устройств с конечным числом состояний, не обязательно построенных из нейронов. Аппарат, разработанный для анализа процессов в нервной системе, привел в итоге к получению новых теоретических результатов в логике и в электротехнике.

Английский логик А. М. Тьюринг к проблеме «мыслящей машины» подошел по-иному. Не делая каких-либо предположений о физиологии мозга, он поставил задачу: определить в терминах математической логики автомат, который мог бы в принципе выполнять любые строго описанные вычисления. Машина Тьюринга работает, выполняя очень большое число чрезвычайно простых операций. Это — автомат с конечным числом состояний, но снабженный «лентой» бесконечной длины. Его программа записана на конечном участке ленты. Если машине предоставить достаточно времени, то ее «головка» прочитывает записанные команды и печатает результаты на свободных участках той же ленты. Тьюринг показал, что можно сконструировать даже «универсальную» машину для производства любых возможных вычислений. Если ее снабдить описанием задания и машины, которая может выполнить это задание, то «универсальная» машина «разберется», как выполнить задание и выполнит его.

Как модель нервной сети Маккаллока — Питтса, так и более абстрактное понятие машины Тьюринга стимулировали появление интересных гипотез о природе мышления и о колоссальных возможностях машин. Специалисты по теории автоматов, например, пытались моделировать способность биологических систем к самовосстановлению и к надежному функционированию, несмотря на наличие ненадежных компонент. Фон Нейман показал, как можно построить машину, которая бы работала требуемым образом, даже если в процессе работы некоторые ее элементы выходят из строя. Это достигается введением избыточности: например, увеличением количества логических элементов и числа соединений между ними в электронной вычислительной машине. И самовосстановление и избыточность чрезвычайно интересны с точки зрения проектировщика ЭВМ или других автоматических систем; в настоящее время этими проблемами серьезно занимаются.

Биологический нейрон, несомненно, не просто выключатель с двумя состояниями: «включено» и «выключено» в каждый данный момент времени. Бесспорно и то, что ЭВМ не являются «мыслящими машинами». Возможно, удастся построить лучшие машины, не подражая мозгу в том виде, как мы его сейчас понимаем, и, возможно, существуют более короткие пути, ведущие к более полному его пониманию, чем построение сверхупрощенных моделей нейронов. Иногда мне кажется, что усилия, направленные на конструирование этих моделей в качестве первого шага на пути к «мыслящим машинам», похожи на попытки наших предшественников создать искусственные крылья, чтобы научить человека летать: изучение полета птиц способствовало открытию ряда законов аэродинамики, которые справедливы и для аппаратов, создаваемых человеком, но конкретные детали существенно различны. Что касается описания работы мозга в строгих логических терминах, то это превосходит наши сегодняшние возможности. По мнению фон Неймана, может оказаться, что простейшим описанием конкретной деятельности мозга окажется полная схема всех нервных связей. Возможно, мозг удастся описать и в терминах математической логики, но, бесспорно, эта задача на несколько порядков более сложная, чем любые созданные до настоящего времени математические построения.

А не существуют ли такие свойства живых организмов, которые легче подвергнуть логическому анализу? Одно из таких свойств — самовоспроизведение. Вероятно, это — наименее «интеллектуальное» проявление жизни. Большинство организмов не «думает», а многие и вообще не имеют нервной системы, но все организмы воспроизводят себя. Есть надежда, следовательно, что самовоспроизведение окажется более простым, с точки зрения логики, чем мышление. Для того чтобы построить логическую модель самовоспроизведения, надо быть в состоянии по крайней мере сформулировать основные проблемы, касающиеся изучения характеристик этого процесса, и, возможно, даже наметить некоторые приемы, которыми эти проблемы могут быть решены; это может в свою очередь подсказать биологам пути исследования и, быть может, прольет свет на некоторые в настоящее время усиленно изучаемые биологические процессы.

Фон Нейман первый занялся детальным изучением того, как в машине воплотить процесс самовоспроизведения. Эта задача заключалась, по его собственным словам, в следующем: «Можно ли из простых элементов создать такую машину, которая, будучи помещенной в резервуар, где в достаточном количестве находятся подобные разрозненные элементы, стала бы создавать другие машины, являющиеся точной копией оригинала?» Далее показывается реальность подобной модели: определенным образом обученная машина, помещенная в «питательную» среду, будет блуждать в ней, собирая части, нужные для создания себе подобных, и со временем будет две, затем четыре, восемь, шестнадцать таких машин — и так далее до тех пор, пока хватает элементов и свободного пространства.

На первый взгляд это звучит совершенно неправдоподобно и даже несерьезно. Так ли уж все это просто? Если кристаллический «зародыш», состоящий всего из нескольких молекул вещества, поместить в среду, насыщенную такими же молекулами при соответствующей температуре и давлении, то свободные молекулы будут оседать на кристаллик, в точности повторяя его структуру. С этой точки зрения рост кристалла есть самовоспроизведение. Еще пример — застежка «молния»: здесь происходит сцепление пары зубчиков, находящихся в среде, состоящей из соединителя и множества зубчиков, расположенных в ряд. Все остальные зубчики тоже поочередно вступают в зацепление, образуя последовательность двухзубчиковых «машин» — что-то вроде одномерного кристалла.

Конечно, это тривиальные примеры, построенные на использовании простейшего изменения состояния: от аморфного к кристаллическому, от открытого к закрытому. А вот пример с перфокартами. Каждая из них — «машина», и она воспроизводит себя, конечно, с помощью сложного устройства дублирующего перфоратора, который, собственно, и вносит в систему необходимую упорядоченность. Попутно отметим, что некоторые одноклеточные организмы могут воспроизводиться в простой питательной среде, в то время как более высокоорганизованные могут потребовать более совершенной среды, содержащей, например, такие сложные вещества как витамины. Таким образом, представляет интерес процесс образования сложной машины из простых элементов — возможно, из большого их числа. Но количество разных типов элементов (или их состояний) должно быть невелико. Как раз этого и достиг фон Нейман в своей модели самовоспроизведения.

Тем не менее может возникнуть вопрос, насколько эта модель отражает специфику соответствующего биологического процесса. Одна из первых же трудностей, с которыми пришлось встретиться, и способ, каким она была преодолена, представляют собой хороший пример связи с биологией. Фон Нейман вскоре пришел к тому, что программа построения машиной своей копии не может быть полной. Для получения полной программы необходимо было бы дать описание не только автомата, но и программы; потребовалось бы описание описания — и так до бесконечности. Обойти эту трудность можно, создав две машины, которые рассматривали бы описание в двух аспектах. Одна из машин — копировальное устройство (C), другая — исполнительное устройство (O). Они работают в комплексе с устройством синхронизации (s), которое управляет работой каждого из них. Программа должна описывать все три устройства (B_{C+O+s}). Всю машину можно описать выражением: $C+O+s+$ $+B_{C+O+s}$. Если такой машине дано ее полное описание, то C копирует его, а O производит все действия, необходимые для построения C, O и s.

Результаты, недавно полученные в молекулярной генетике, обнаруживают поразительные аналогии между элементами фон Неймана и процессами в живой клетке. B — это набор генов, состоящих из дезоксирибонуклеиновой кислоты (ДНК), в которых закодированы наследственные признаки. C — фермент ДНК — полимер, который катализирует редупликацию ДНК, в результате чего нити ДНК копируют друг друга, удваивая таким образом набор генов. O — система, состоящая из рибонуклеиновой кислоты (РНК), ферментов и рибосом, которая соединяет аминокислоты в соответствии с инструкциями ДНК. Полученные таким образом ферменты и другие белки участвуют в построении новой клетки.

Первые модели фон Неймана были «механическими». Но эти модели не были физически построены. Однако несколько менее общие модели были продемонстрированы. В одной из них, построенной Г. Джекобсоном из Бруклинского колледжа, самовоспроизводящаяся машина представляла собой поезд из игрушечных вагонов*). Вагоны представляли собой автономные элементы системы, управление осуществлялось устройствами, которые устанавливались в каждом вагоне. Когда поезд, составленный из вагонов разного типа (в определенной последовательности), помещался на пути, он начинал двигаться, подталкивал неприцепленные вагоны, при необходимости менял их взаимное расположение, используя запасные пути, и составлял из них поезда «по своему образу и подобию». Джекобсон разработал подобную модель, которая допускала поезда произвольной длины и произвольную последовательность вагонов. Им была построена сравнительно простая демонстрационная модель, в которой воспроизводился поезд из двух вагонов.

---

*) H. Jacobson, On models of reproduction, Amer. Sci. **46**, № 3, 255—268. [Русский перевод см. в «Кибернетическом сборнике» № 7, 1963.]

Английскому генетику Л. С. Пенроузу принадлежит другая модель*), построенная частично на основе его предположений о принципах самовоспроизведения генетического материала. В простейшем случае в машину Пенроуза входят элементы двух видов A и B, вырезанные таким образом, что они могут соединяться друг с другом одним из двух способов. Если в поднос поместить машину AB и несвязанные элементы A и B, а затем начать встряхивать поднос (первая машина служит зародышом), то и возникают новые соединения AB (рис. 3). Если в качестве зародыша поместить машину BA, то воспроизводимыми парами будут BA.

К числу достоинств механических моделей следует отнести их наглядность и убедительность, но составить их математическое описание чрезвычайно трудно. В своих более поздних работах фон Нейман обратился к абстрактным моделям; он отказался от физических моделей, избежав этим всех трудностей, связанных с их изготовлением, движением и управлением. Сформулированная им задача относилась скорее к логике и математике, чем к механике и электротехнике. Вместо «питательной среды» появилась математическая двумерная решетка или плоскость, разбитая на квадраты. В каждый квадрат помещается один элемент — машина с конечным числом состояний (конечный автомат). Машины фон Неймана не имеют ни входов, ни выходов, они имеют лишь некоторое количество допустимых состояний. Таблица этих состояний и правила, определяющие переходы из одного состояния в другие, являются общими для всех элементов; при этом в любой момент времени разные элементы могут находиться в разных состояниях. Каждая машина является детерминированным синхронным автоматом: в каждый дискретный момент времени T (за исключением начального состояния при $T=0$) состояние каждого элемента зависит только от своего собственного состояния и состояния ближайших

*) L. Penrose, Automatic mechanical self-reproduction, New Biology, № 28, England, p. 92 (1959). (*Прим. ред.*)

соседей в момент времени $T-1$. Существует особое состояние, называемое состоянием «покоя». В этом состоянии находятся все элементы, за исключением

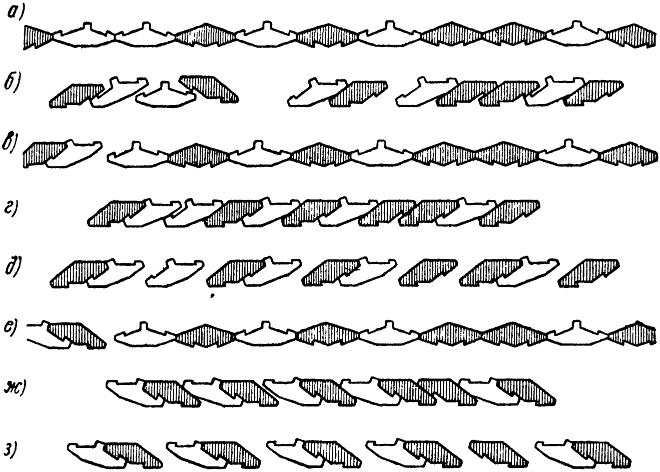


Рис. 3. Самовоспроизведение является одним из биологических процессов, подвергнутых математическому анализу. Маленькие заштрихованные и незаштрихованные «существа» — это два типа составных частей простейшей самовоспроизводящейся машины, сконструированной генетиком Л. С. Пенроузом. Если эти части поместить в поднос (*а*) и начать покачивать его взад-вперед, то части будут сталкиваться друг с другом, не вступая в зацепление (*б*). Но если в поднос поместить черно-белый зародыш или «машину» (*в*) и снова покачать его, то зародыш будет способствовать приданию определенного наклона другим частям, которые, вступая в зацепление, снова создают черно-белые комплексы (*г*). Для наглядности эти машины показаны отдельно (*д*). Если теперь за исходный зародыш взять бело-черную конфигурацию, то в результате покачивания будут возникать только бело-черные машины (*ж*, *з*): машины дают «чистокровный приплод». У Пенроуза «существа» были сделаны из фанеры, а те, что показаны на рисунке — из алюминия в Телефонных лабораториях Белла.

некоторого конечного их числа. Если некоторый элемент и все его ближайшие соседи находятся в покое в момент времени $T-1$, то этот элемент будет находиться в покое и в момент времени T. Вся система в целом: плоскость, разбитая на квадраты, элементарные машины, допустимые состояния и правила перехода — названа «мозаичной структурой» (рис. 4). Конечная совокупность элементов называется «конфигурацией», если определены состояния всех входящих в нее элементов.

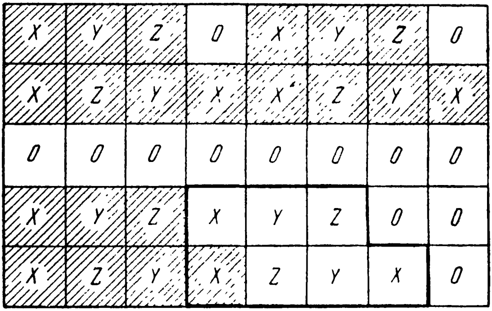


Рис. 4. Мозаичная структура — это часть плоскости, разделенная на квадратные ячейки. Рассматриваемая область представляет собой конфигурацию из 40 ячеек, состояния которых обозначены X, Y, Z и 0 (невозбужденные). Вся конфигурация состоит из трех экземпляров семиклеточных конфигураций (заштрихованы). Отдельные экземпляры должны быть непересекающимися подмножествами. Четвертое множество (заключенное в черный контур), хотя и повторяет конфигурацию остальных, не является самостоятельной копией из-за пересечения с остальными.

Какая же связь между всеми этими определениями, машинами и самовоспроизведением? Будем рассматривать мозаичную структуру как некую абстрактную среду, в которой квантованы пространство и время, а движение и прочие плавные изменения заменены последовательностью дискретных состояний; переходы от одних состояний к другим описываются некоторыми правилами. В этой среде существуют конфигурации, составленные из элементов, — это наши «машины». Эти машины и способны к самовоспроизведению. Отдельные элементы — это простейшие частицы, скажем молекулы. Изменение их состояний можно представлять себе в виде энергетических переходов, изменения уровня химической активности, изменения геометрического положения. А правила перехода из состояния в состояние — это физические и химические законы среды, определяющие изменения, происходящие в самих элементах, и их связь с другими элементами. Элементы, находящиеся в покое, — это неиспользованное сырье, а правила, определяющие состояние покоя, фактически гласят, что ни один элемент, не входящий в какую-либо конфигурацию, не может внезапно стать активным; машина «подбирается» к окружающему сырью лишь путем локальных действий.

Таким образом, необходимо построить мозаичную структуру, состоящую из элементов с небольшим числом состояний (другими словами, весьма простых элементов), сформулировать правила переходов, а затем создать такую конфигурацию, которая бы обладала способностью к самовоспроизведению. Это до некоторой степени напоминает процесс написания программы для вычислительной машины. Фон Нейман выдвинул дальнейшее требование: каждая конфигурация должна содержать в себе машину Тьюринга. Затем он довольно подробно описал самовоспроизводящуюся конфигурацию, включающую около 200 000 элементов с 29 внутренними состояниями. После его смерти в 1957 г. мозаичными структурами продолжали заниматься; исследовались их структурные особенности, процесс воспроизведения пытались записать в строгой логической форме, результаты облекались в форму доказываемых теорем.

Мы уже интересовались вопросом: насколько быстро может расти популяция самовоспроизводящихся конфигураций? В действительности она не может возрастать экспоненциально, например удваиваться с каждым поколением. Численность популяции двумерной мозаичной структуры в любой момент времени не может превышать квадрата времени ее воспроизведения.

Сформулируем наше утверждение в качестве теоремы: *если самовоспроизводящаяся конфигурация* *создает $f(T)$ себе подобных конфигураций за время T, то существует положительная константа k, такая что*

$$f(T) \leqslant kT^2.$$

Доказательство состоит в следующем. Пусть C — самовоспроизводящаяся конфигурация, умещающаяся на квадрате размером $D \times D$. Тогда в каждый момент времени общее количество возбужденных элементов не может превысить величины $(2T+D)^2$, поскольку совокупность возбужденных элементов, имеющая форму квадрата, может увеличивать свои размеры лишь на один элемент с каждой стороны за единицу времени. Если r — количество элементов в C, то

$$f(T) \leqslant \frac{(2T+D)^2}{r}.$$

Путем несложных упрощений из этого неравенства можно прийти к доказательству теоремы. Эта ситуация напоминает закон Мальтуса: имеющееся пространство ограничивает рост населения, поскольку скорость распространения границы области возбуждения не может превысить некоторой постоянной величины.

Еще один важный вопрос: любая ли конфигурация способна к самовоспроизведению? Оказывается, существуют конфигурации, которые не могут возникнуть, если только они не существовали в самом начале при $T=0$. Эти конфигурации неконструируемы в том смысле, что не существует такой предшествующей конфигурации, из которой могла бы быть получена данная при помощи установленных правил перехода. Джон Тьюки (Принстонский университет) предложил называть конфигурации, не имеющие предшественников «садами Эдема». Поскольку такие конфигурации не могут быть получены ни из каких других, они не способны к самовоспроизведению. Поэтому исследование условий, при которых они могут возникать, позволяет установить основные ограничения на способность машин к самовоспроизведению.

Эти условия включают в себя нечто, что можно было бы назвать возможностью «стирания». Когда с классной доски исчезают следы мела, невозможно сказать, что там было написано до этого. В области вычислительной техники термин «стирание» относится к операции над элементами памяти, в результате которой они устанавливаются в исходное состояние независимо от хранимой перед этим информации.

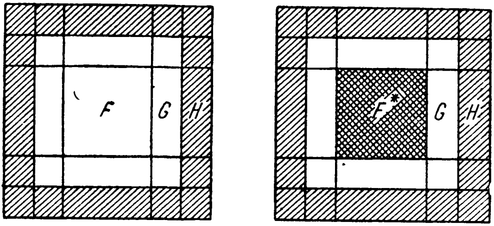


Рис. 5. Конфигурации, соответствующие двум областям одинакового размера, представлены в момент T. Конфигурации внутренних областей отличаются друг от друга (на рисунке внутренние области имеют определенные размеры, но, вообще говоря, они могут быть произвольными). Промежуточные конфигурации G и внешние конфигурации H идентичны друг другу.

В общем стирание — необратимый процесс, т. к. после его выполнения невозможно установить предыдущее состояние, из которого был осуществлен переход в данное. Тогда должно существовать по крайней мере два предшествующих состояния, которые определенными переходами переводились бы в одно. В случае мозаичной структуры понятие стирания должно быть определено несколько более строго, чтобы действительно гарантировать стирание, а не просто переход к новому состоянию.

Чтобы сформулировать такое определение, рассмотрим две конфигурации на рис. 5. Здесь конфигурации внутренних областей, состоящих из девяти элементов, в данный момент времени различны; обозначим их F и F^*. В то же время конфигурация промежуточных областей — окаймлений квадратов в обоих случаях одна и та же — G. Другими словами, эти области являются точными копиями одна другой. Это же относится и к конфигурациям внешних областей: обе обозначены, как H. Теперь обратите внимание на конфигурации в следующий момент (рис. 6).

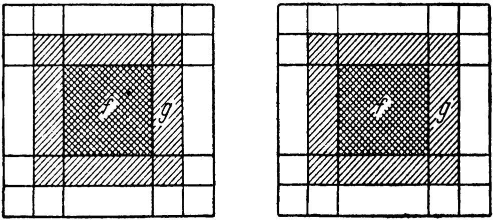


Рис. 6. Новые конфигурации областей, рассмотренных на рис. 5, в момент $T+1$. Внутренние области преобразовались в новые конфигурации, которые оказались совпадающими. Совпадают и промежуточные области. Состояние внешних областей не определено. Если две области, первоначально отличающиеся друг от друга, как, например, на рис. 5, переходят в идентичные конфигурации, то говорят, что они «взаимно стираемы».

Если при этом конфигурации f и g, которыми обозначены соответственно внутренние и промежуточные области, повторяют друг друга, то говорят, что две исходные конфигурации «взаимно стираемы». Будучи разными, они превратились в одинаковые. Обратите внимание, что в этом случае невозможно определить конечную конфигурацию внешней области, потому что на состояния ее элементов оказывают влияние элементы, внешние по отношению к рассматриваемой области, а относительно последних нам ничего не известно. Функция внешней области состояла лишь в защите промежуточной области. Таким образом, будем называть конфигурацию *стираемой*, если существует другая такая конфигурация, что обе они являются взаимно стираемыми. Наконец, из этого определения следует, что если стираемая конфигурация любой формы заключена внутри прямоугольной области, то эта прямоугольная область также стираема. Поэтому приведенные соображения можно рассматривать на областях квадратной формы, с которыми удобней работать, чем с областями произвольной формы.

Теперь можно сформулировать следующую теорему: *если в мозаичной структуре существуют стираемые конфигурации, в ней существуют конфигурации «сад Эдема»*. (Можно, впрочем, построить мозаичную структуру, в которой бы отсутствовали стираемые конфигурации; такая структура была бы весьма тривиальной.)

Мы не станем рассматривать подробного доказательства этой теоремы; я приведу лишь некоторые общие соображения. Пусть n — такое целое число, что в области размером $n \times n$ имеется стираемая конфигурация. Рассмотрим бо́льшую область $kn \times kn$ (рис. 7). Каждая из k^2 областей размером $n \times n$ имеет достаточные размеры, чтобы в ней могла расположиться стираемая конфигурация; k выбирается достаточно большим, так что в большой области возможно наличие многих стираемых конфигураций. Если каждый элемент может находиться в одном из A состояний, то во всей области возможны $A^{(kn)^2}$ конфигураций. Рассмотрим, в какое состояние перейдет вся наша область в следующий момент. Напомним, что в момент $T+1$ мы не сможем точно указать состояния пограничных элементов нашей области. Поэтому в момент $T+1$ мы вправе рассматривать область, имеющую лишь $A^{(kn-2)^2}$ состояний.

Далее, если в нашей исходной области $kn \times kn$ существовала область $n \times n$, содержащая стираемую конфигурацию, тогда два возможных состояния, соответствующие двум взаимно стираемым конфигурациям, перейдут в одно и то же состояние. Если существовало две стираемых конфигурации, тогда четыре возможных состояния перейдут в одно. В общем, если в момент T существует s стираемых конфигураций, то в момент $T+1$ 2^s состояний сольются в одно (рис. 8). Теперь необходимо лишь показать, что потери числа состояний из-за стираний должны превышать потери из-за сокращения внешней границы области, т. е. потери, выражаемые разностью между $A^{(kn)^2}$ и $A^{(kn-2)^2}$.

Предпочтительнее рассматривать не количество состояний, а их логарифмы. Логарифм отношения,

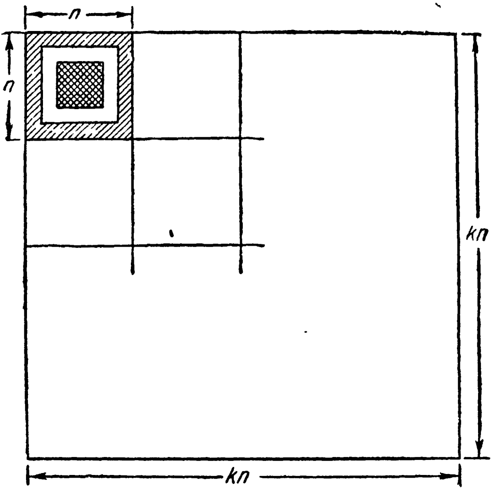


Рис. 7. Теорема о «садах Эдема» доказывается при помощи этого рисунка. Небольшая стираемая конфигурация имеет размеры $n \times n$. Большая область имеет размеры $kn \times kn$ (здесь натуральное число k равно 4, но на самом деле оно является достаточно большим). Эта область показана в момент T. В момент $T+1$ ее размеры уменьшатся до $(kn-2) \times (kn-2)$.

характеризующий потери на пограничном слое, находится примерно в линейной зависимости от k. Но логарифм количества состояний, теряемых из-за стирания, растет так же как число стираемых конфигураций, а следовательно, так же как площадь всей области, т. е. приблизительно пропорционально квадрату k. Таким образом, для больших k бо́льшее число состояний будет теряться из-за стирания, чем от сокращения пограничной области. В момент $T+1$ должно существовать некоторое состояние P, которому не может предшествовать никакое состояние в момент T. Это состояние P и есть конфигурация «сад Эдема», о которой говорится в теореме. Это состояние относится к числу возможных, но его нельзя достигнуть ни из

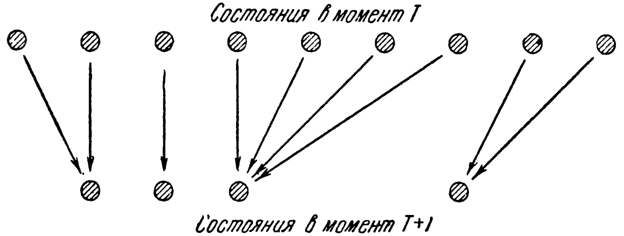


Рис. 8. Число допустимых состояний для области, изображенной на рис. 7, за промежуток времени от T до $T+1$ уменьшается. Это происходит из-за стирания и сокращения пограничной области. Можно показать, что из-за стирания происходят бо́льшие потери, чем от сокращения. Таким образом, в момент $T+1$ появляется лишнее состояние P, называемое конфигурацией «сад Эдема».

какого предыдущего состояния. Оно соответствует машине, которую можно описать как совокупность ряда элементов, но которую нельзя создать из этих элементов. Поскольку эта машина не может возникнуть путем воспроизведения, она не в состоянии воспроизводить себя.

Двух теорем, которые я привел выше, вероятно достаточно, чтобы понять, каким образом такой процесс как воспроизведение может быть описан абстрактно и сделан доступным математическому анализу. Я не хочу этим дать повод к утверждению, что обязательно существует тесная связь между мозаичной структурой и биологическим воспроизведением; бесспорно это еще нужно показать. Однако, как я утверждал выше, весьма вероятно, что рассмотрение подобных абстрактных моделей по крайней мере поможет прояснить некоторые аспекты, установить некоторые критерии биологических процессов.

Рассмотрим следующий пример. Жизнь на Земле, как полагают, возникла вследствие случайного соединения неорганических веществ. Насколько это правдоподобно? Из рассмотрения мозаичных структур можно сделать вывод о том, насколько сложной должна быть система элементов, которая была бы в состоянии проявлять свойство самовоспроизведения, а также способность претерпевать эволюцию, совершенствуя свой вид. Джекобсон, изучавший поведение машин на игрушечных поездах, сделал шаг в этом направлении, охарактеризовав понятие сложности в битах информации.

Даже если машинные модели самовоспроизведения в конце концов окажутся неприменимыми к биологии, они могут представить исключительный интерес сами по себе. Вернемся на минуту к модели самовоспроизведения фон Неймана. Что если построить подобную машину не для некой абстрактной среды, а для существования в реальном физическом мире? Подобная машина, как обычное живое растение, будет синтезировать свои элементы из окружающих веществ, создавая из них необходимые конструкции, воспроизводя себя. При этом она будет из окружающей среды извлекать необходимые вещества, очищать их и накапливать. Например, установки, предназначенные для самовоспроизведения в океанах, могут в качестве конструкционного материала использовать магний, присутствующий в воде, при этом человек получил бы возможность собирать «урожаи» магния. До сих пор еще никому не удавалось провести исследовательскую работу, необходимую для создания подобной машины, но я думаю, что однажды она будет построена.

Библиография

Более детальное введение в нейрофизиологию для неспециалистов дано в

1. Eccles J. C., The physiology of the imagination, Sci. Amer. **199**, № 3 (1958), 135—146.

Подробное описание можно найти в

2. Ranson S. W., Clark S. L., The anatomy of the nervous system, 10th ed., W. B. Saunders Company, Philadelphia and London, 1959.

Модель Маккаллока—Питтса впервые была описана в

3. Mc Culloch W. S., Pitts E. W., A logical Calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. **5** (1943), 115—133. [Русский перевод в сб. «Автоматы», М., ИЛ, 1956, стр. 362—384. Предупреждаем читателя, что часть 3 непонятна]

Превосходной статьей по конечным автоматам (выходы которых принимают значения «да» или «нет») является

4. Rabin M. O., Scott D., Finite automata and their decision problems, IBM J. Res. and Develop. **3**, № 2 (1959), 114—125. [Русский перевод в Кибернетическом сборнике, вып. 4, М., ИЛ, 1962, стр. 58—91.]

Понятие регулярного события впервые было введено в

5. Kleene S. C., Representation of events in nerve nets and finite automata, Automata Studies, Princeton Univ. Press, Princeton, N. Y. (1956) 3—41. [Русский перевод в сб. «Автоматы», М., ИЛ, 1956, стр. 15—67.]

Основной результат Клини в более доступном изложении содержится в

6. Copi I. M., Elgot C. C. and Wright J. B., Realization of events by logical nets, J. Assoc. Comput. Mach. **5**, № 2 (1958) 181—196. [Русский перевод в Кибернетическом сборнике, вып. 3, М., ИЛ, 1961, стр. 147—166.]

Это изложение легло в основу § 1.7. Некоторые факты в менее доступной форме были изложены в

7. Arbib M., Turing machines, finite automata and neural nets, J. Assoc. Comput Mach., **8**, № 4 (1961), 467—475.

Машины Тьюринга названы в честь А. Тьюринга, который впервые определил их в статье

8. Turing A. M., On computable numbers with an application to the Entscheidungsproblem, Proc, London. Math. Soc., ser. **2**, **42**;

№ 3—4 (1936), 230—265. См. также исправление в Proc. London. Math. Soc. **43**, № 7 (1937), 544—546.

Изложение теории машин Тьюринга и их использования в теории рекурсивных функций имеется в

9. Davis M., Computability and Unsolvability, Mc Graw-Hill Book Company, Inc., N. Y., 1958.

Ясные, неформальные доказательства теорем 1.6.2 и 1.6.3 содержатся в поистине замечательной статье (которую я рекомендую Вашему вниманию).

10. Post E. L., Recursively enumerable sets of positive integers and their decision problems, Bull. Amer. Math. Soc. **50**, № 5 (1944), 284—316.

Читателю, желающему познакомиться с моделированием нервной системы на уровне «проводов и транзисторов», мы рекомендуем

11. Harmon L. D., Studies with artificial neurons, Kybernetik **1**, № 3 (1961), 89—101.

Дополнительную литературу можно найти в недавно опубликованной книге

12. Wooldridge D. E., The machinery of the brain, Mc Graw-Hill Book Company, Inc., New York, 1963. [Русский перевод: Вулдридж Д., Механизмы мозга, М., Изд-во «Мир», 1965]

Алгебраическим аспектам теории конечных автоматов посвящены работы:

13. Kalman R. E., Falb P. D., Arbib M. A., Topics in modern system theory, Mc Graw-Hill Book Company, N. Y. 1966.

14. Hartmanis J. and Stearns R. E., Algebraic Structure theory of sequential machines, Prentice-Hall, Inv., Engleewood Clyft N. Y., 1966.

Проблемам самовоспроизведения посвящены работы

15. Burks A. W., Computation, behavior and structure in fixed and growing automata, Self-organizing systems (ed. M. Yovits and S. Cameron), N. Y., Pergamon Press, 1960. [Русский перевод в сб. «Самоорганизующиеся системы», Изд-во «Мир», 1964.]

16. Lee C. Y., A Turing machine which prints its own code script, Proc. Sumpos. Math. Theory of Automata, Polytechnic Press, Brooklyn (1963), 155—164.

17. Moore E. F., Machine models of self-reproduction, Mathematical Problems in the Biological Sciences, Proc. Sympos. Applied. Math. **14** (1962), 17—33.

18 Morrison P., A thermodynamic characterization of self-reproduction, Rev. Mod. Phys. **36**, № 2 (1964), 517—523.

19. Myhill J., The abstract theory of self-reproduction, Views on General Systems Theory, J. Wiley (1964), 106—118.

20. Shannon C. E., Von Neumann's contributions to automata theory, Bull. Amer. Math. Soc. **64**, № 2 (1958), 139—173. [Русский перевод в сб. Шеннон К. Э., «Работы по теории информации и кибернетике», ИЛ, 1963, стр. 233—239.]

21. Stahl W. R., Self-reproducing automata, Perspectives in Biology and Medicine **8**, № 3 (Spring 1965), 373—393.

22. Thatcher J. W., The construction of a self-describing Turing machine, Proc. Sympos. Math. Theory Automata, Polytechnic Press, Brooklyn (1963), 165—171.

23. Von Neumann J., The general and logical theory of automata, Celebral Mechanisms in Behavior, The Hixon Sympos. (L. A. Jeffress, ed.), New York — London (1951), 2070—2098. [Русский перевод в кн. Тьюринг А., Может ли машина мыслить, М., Физматгиз, 1960, стр. 59—101.]

24. Rabin M. O., Degree of difficulty of computing a function and a partial ordering of recursive sets, Hebrew University, Jerusalem, Israel, April 1960.

25. Ritchie R. W., Classes of predictably computable functions, Trans. Amer. Math. Soc. **106**, № 1 (1963), 139—173.

26. Arbib M. A., Blum M., Machine dependence of degrees of difficulty, Proc. Amer. Math. Soc. **16**, № 3 (1965), 442—447.

Литература, добавленная редактором перевода

К главе 1

1. Автоматы, Сб. статей, М., ИЛ, 1956.

2. Барздинь Я. М., Проблемы универсальности в теории растущих автоматов, ДАН СССР 157, № 3 (1964), 542—545.

3. Барздинь Я. М., Моделирование логических сетей на автоматах Неймана—Черча, «Проблемы кибернетики», вып. 17, Изд-во «Наука», 1966, 5—26.

4. Глушков В. М., Синтез цифровых автоматов, Физматгиз, 1962.

5. Гнеденко Б. В., Королюк В. С., Ющенко Е. Л., Элементы программирования, Физматгиз, 1961.

6. Гутенмахер Л. И., Электронные информационно-логические машины, Изд-во АН СССР, 1960.

7. Карр Дж., Лекции по программированию, ИЛ, 1963.

8. Карцев М. А., Арифметические устройства электронных цифровых машин, Физматгиз, 1958.

9. Китов А. И., Криницкий Н. А., Электронные цифровые машины и программирование, Физматгиз, 1959.

10. Кобринский Н. Е., Трахтенброт Б. А., Введение в теорию конечных автоматов, Физматгиз, 1962.

11. Колдуэлл С., Логический синтез релейных устройств, ИЛ, 1962.

12. Козмидиади В. А., О множествах, разрешимых и перечислимых автоматами, «Проблемы логики», Изд-во АН СССР, 1963, 102—115.

13. Кратко М. И., Алгоритмическая неразрешимость проблемы распознавания полноты для конечных автоматов, ДАН СССР 155, № 1 (1964), 35—37.

14. Лупанов О. Б., О сравнении двух типов конечных источников, «Проблемы кибернетики», вып. 9, Физматгиз, 1963, 321—326.

15. Лупанов О. Б., О синтезе некоторых классов управляющих систем, «Проблемы кибернетики», вып. 10, Физматгиз, 1963, 63—97.

16. Лупанов О. Б., Об одном подходе к синтезу управляющих систем — принципе локального кодирования, «Проблемы кибернетики», вып. 14, Изд-во «Наука», 1965, 31—110.

17. Мурский В. Л., О преобразованиях конечных автоматов. «Проблемы кибернетики», вып. 15, Изд-во «Наука», 1965, 101—116.

18. Мэрфи Дж., Как устроены и работают электронные цифровые машины, Изд-во «Мир», 1965.

19. Нечипорук Э. И., О сложности схем в некоторых базисах, содержащих нетривиальные элементы с нулевыми весами. «Проблемы кибернетики», вып. 8, Физматгиз, 1962, 123—160.

20. Нечипорук Э. И., О синтезе схем из пороговых элементов. «Проблемы кибернетики», вып. 11, Физматгиз, 1964, 49—62.

21. Поспелов Д. А., Игры и автоматы, Изд-во «Энергия», 1966.

22. Ричардс Р. К., Арифметические операции на цифровых вычислительных машинах, ИЛ, 1957.

23. Ричардс Р. К., Элементы и схемы цифровых вычислительных машин, ИЛ, 1961.

24. Саломаа А., Аксиоматизация алгебры событий, реализуемых логическими сетями. «Проблемы кибернетики», вып. 17, Изд-во «Наука», 1966, 237—246.

25. Трахтенброт Б. А., Конечные автоматы и логика одноместных предикатов, Сиб. матем. журнал 3, № 1 (1962), 103—131.

26. Фистер М., Логическое проектирование цифровых вычислительных машин, Изд-во «Техніка», Киев, 1964.

27. Цетлин М. Л., Конечные автоматы и моделирование простейших форм поведения, УМН **18**, вып. 4 (1963), 3—28.

28. Яблонский С. В., Функциональные построения в k-значной логике, Труды матем. института им. Стеклова, Изд-во АН СССР **51** (1958), 5—142.

29. Яблонский С. В., Основные понятия кибернетики, «Проблемы кибернетики», вып. 2, Физматгиз, 1959, 7—38.

30. Яблонский С. В., Гаврилов Г. М., Кудрявцев В. Б., Функции алгебры логики и классы Поста, Изд-во «Наука», 1965.

31. Янов Ю. И., О логических схемах алгоритмов. «Проблемы кибернетики», вып. 1, Физматгиз, 1958, 75—127.

К главе 2

1. Анохин П. К., Физиология и кибернетика. Философские проблемы кибернетики, Соцэкгиз, 1961.

2. Бехтерев В. М., Современные проблемы строения и функций мозга в норме и патологии, Л., Медгиз, 1959.

3. Бионика. Сб. статей, Изд-во «Наука», 1965.

4 Брайнес С. Н., Напалков А. В., Свечинский В. Б., Нейрокибернетика, Медгиз, 1962.

5. Глушков В. М., Введение в кибернетику, Изд-во АН УССР, Киев, 1964.

6. Математические проблемы в биологии, Сб. статей под ред. Р. Беллмана, М., 1966.

7. Моделирование в биологии, М., 1963.
8. Проблемы бионики, Сб. статей, Изд-во «Мир», 1965.
9. Розенблатт Ф., Принципы нейродинамики, Изд-во «Мир», 1965.
10. Электроника и кибернетика в биологии и медицине, ИЛ, 1963.
11. Эшби У. Р., Конструкция мозга. Происхождение адаптивного поведения, ИЛ, 1962.
12. Шредингер Э., Что такое жизнь с точки зрения физики, ИЛ, 1947.

К главе 3

1. Бриллюэн Л., Научная неопределенность и информация, Изд-во «Мир», 1966.
2. Бриллюэн Л., Наука и теория информации, Физматгиз, 1960.
3. Гаврилов М. А., Структурная избыточность и надежность релейных устройств, Труды I конгресса ИФАК, 1960.
4. Гиндикин С. Г., Мучник А. А., Решение проблемы полноты для систем функций алгебры логики с ненадежной реализацией. «Проблемы кибернетики», вып. 15, Изд-во «Наука», 1965, 65—84.
5. Кириенко Г. И., О самокорректирующихся схемах из функциональных элементов. «Проблемы кибернетики», вып. 12, Изд-во «Наука», 1964, 29—37.
6. Месси Дж., Пороговое декодирование, Изд-во «Мир», 1966.
7. Методы введения избыточности для вычислительных систем, Сб. статей, Изд-во «Сов. радио», 1966.
8. Нечипорук Э. И., О самокорректирующихся вентильных схемах, ДАН СССР **156**, № 5 (1964), 1045—1048.
9. Питерсон У., Коды, исправляющие ошибки, Изд во «Мир», 1964.
10. Седов Е., Репортаж с ничейной земли, Изд-во «Молодая гвардия», 1963.
11. Фано Р., Передача информации. Статистическая теория связи, Изд-во «Мир», 1965.
12. Феллер В., Введение в теорию вероятностей и ее приложения, т. 1, Изд-во «Мир», 1964.
13. Яглом А. М., Яглом И. М., Вероятность и информация, Физматгиз, 1960.

К главе 4

1. Айзерман М. А., Браверман Э. М., Глушков В. М., Ковалевский В. А. и Летичевский А. А., Теория опознавания образов и обучающихся систем, «Техническая кибернетика», Изв. АН СССР, № 5 (1963), 98—101.
2. Айзерман М. А., Теория автоматического регулирования, Изд-во «Наука», 1966.
3. Бжалава И. Т., Психология установки и кибернетика, Изд-во «Наука», 1966.

4. Винер Н., Кибернетика и общество, ИЛ, 1958.
5. Вычислительные машины и мышление, Сб. статей, Изд-во «Мир», 1967.
6 Гаазе-Рапопорт М. Г., Автоматы и живые организмы. Моделирование поведения живых организмов, Физматгиз, 1961.
7. Гельфанд И. М., Цетлин М. Л., О некоторых способах управления сложными системами, Успехи матем. наук, 17, вып. 1 (1962), 3—25.
8. Греневский Г., Кибернетика без математики, Изд-во «Сов. радио», 1964.
9. Гродинз Ф., Теория регулирования и биологические системы, Изд-во «Мир», 1966.
10. Гуд Г. X., Макол Р. Э., Системотехника. Введение в проектирование больших систем, Изд-во «Сов. радио», 1962.
11. Евграфов М. А., Задыхайло И. Б., Некоторые соображения о программировании шахматной игры, «Проблемы кибернетики», вып. 15, Изд-во «Наука», 1965, 135—156.
12. Зарипов Р. X., О моделировании мелодий заданного стиля на цифровых вычислительных машинах. «Проблемы кибернетики», вып. 15, Изд-во «Наука», 1965, 157—200.
13. Ивахненко А. Г., Техническая кибернетика. Системы автоматического управления с приспособлением характеристик, Гостехиздат УССР, Киев, 1962.
14. Кибернетику на службу коммунизму, Сб. статей (т. 1, Госэнергоиздат, 1961; т. 2, Изд-во «Энергия», 1964; т. 3, Изд-во «Энергия», 1966).
15. Кобринский А. Е., Использование биотоков для целей управления. Изв. АН СССР, Энергетика и автоматика, № 3 (1959), 152—154.
16. Ковалевский В. А., Корреляционный метод распознавания изображений. Журнал вычисл. матем. и матем. физики, вып. 4 (1962), 684—694.
17. Косса П., Кибернетика. От человеческого мозга к мозгу искусственному, ИЛ, 1958.
18. Лук А. Н., Память и кибернетика, Изд-во «Наука», 1966.
19. Ляпунов А. А., Яблонский С. В., Теоретические проблемы кибернетики, «Проблемы кибернетики», вып. 9, Физматгиз, 1963, 5—22.
20. Моль А., Теория информации и эстетическое восприятие, Изд-во «Мир», 1966.
21. Полетаев И. А., Сигнал, Изд-во «Сов. радио», 1958.
22. Принципы самоорганизации, Сб. докладов, Изд-во «Мир», 1966.
23. Ревзин И. И., Модели языка, Изд. АН СССР, 1962.
24. Самоорганизующиеся системы, Сб. докладов, Изд-во «Мир», 1964.
25. Система «Человек и автомат», Сб. статей, Изд-во «Наука», 1965.
26. Эшби У. Р., Введение в кибернетику, ИЛ, 1959.

К главе 5

1. Возможное и невозможное в кибернетике, Сб. статей, Изд-во АН СССР, 1963.

2. Гильберт Д., Аккерман В., Основы теоретической логики, ИЛ, 1947.

3. Гудстейн Р. Л., Математическая логика, ИЛ, 1961.

4. Джордж Ф., Мозг как вычислительная машина, ИЛ, 1963.

5. Клини С. К., Введение в метаматематику, ИЛ, 1957.

6. Логические исследования, Сб. статей, Изд. АН СССР, 1959.

7. Мальцев А. И., Алгоритмы и рекурсивные функции, Изд-во «Наука», 1965.

8. Марков А. А., Теория алгорифмов. Труды матем. института им. В. А. Стеклова, Изд-во АН СССР, 42 (1954).

9. Новиков П. С, Элементы математической логики, Физматгиз, 1959.

10. Робинсон А., Введение в теорию моделей и метаматематику алгебры, Изд-во «Наука», 1967.

11. Слупецкий Е., Борковский Л., Элементы математической логики и теория множеств, Изд-во «Прогресс», 1965.

12. Тарский А., Введение в логику и методологию дедуктивных наук, ИЛ, 1948.

13. Трахтенброт Б. А., Алгоритмы и машинное решение задач, Физматгиз, 1960.

14. Тьюринг А., Может ли машина мыслить? Физматгиз, 1960

15 Успенский В. А., Лекции о вычислимых функциях, Физматгиз, 1960.

Цена 55 коп.

Нервные сети,
конечные автоматы
и машина Тьюринга

•

Структура
и случайность

•

Исправление ошибок
при передаче
и вычислениях

•

Кибернетика

•

Теорема Гёделя
о неполноте

•

Математика
в биологии