11. Логика и эмпирическое познание. Сб. под ред. П. В. Таванца. М., «Наука», 1972.
12. *М. М. Бонгард, М. Н. Вайнцвайг, Ш. А. Губерман, М. Л. Извекова, М. С. Смирнов.* Использование обучающейся программы для выявления нефтеносных пластов.— Геология и геофизика, 1966, № 6.
13. *И. М. Гельфанд, Ш. А. Губерман, М. П. Житков, М. С. Калецкая, В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн.* Прогноз места возникновения сильных землетрясений как задача распознавания.— Наст. сб.
14. *М. М. Бонгард, И. С. Лосев, М. С. Смирнов.* Проект модели организации поведения — «Животное».— Наст. сб.

М. Н. Вайнцвайг, М. П. Полякова

ОБ ОДНОМ ПОДХОДЕ К ПРОБЛЕМЕ СОЗДАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

1. Что такое интеллект. Приступая к созданию систем искусственного интеллекта, необходимо прежде всего решить вопрос о том, какое из свойств естественного интеллекта следует считать определяющим. Чаще всего в качестве такого свойства рассматривают *способность решать широкий круг достаточно сложных задач*, поскольку сложностью и разнообразием решаемых задач обычно оценивается интеллектуальность человека. Эта точка зрения на интеллект согласуется и с практическими целями, а именно с желанием иметь устройства, способные выполнять функции специалистов в той или иной области. Распространение такой точки зрения привело к тому, что подавляющее большинство исследователей в области искусственного интеллекта пошло по пути построения систем, решающих некоторые классы достаточно сложных для человека задач, надеясь, по-видимому, что рано или поздно в принципах построения этих систем обнаружится нечто общее и это общее даст возможность построить модель интеллекта в целом. Так возникли программы для доказательства теорем, игры в шахматы, перевода с одного языка на другой, распознавания образов, а также различные системы типа «вопрос — ответ», «интеллектуальные» роботы и пр. Многие из этих работ быстро нашли практическое применение, что послужило стимулом для развития этого направления. Однако, если конечной целью таких исследований считать построение целостной модели интеллекта, то выбранный путь нам представляется неоправданно сложным и малоэффективным.

Дело прежде всего в том, что к моменту, когда человек способен решать сколько-нибудь сложные задачи, его память уже хранит чрезвычайно большое количество различных сведений, а

поскольку он часто находит аналогии между очень, казалось бы, далекими областями знаний, то любое из этих сведений может быть им использовано в процессе решения. Если кроме того учесть, что далеко не все информационные процессы протекают через сознание, то окажется, что выявить и формализовать реальные для человека пути решения сложных задач практически невозможно. Вспомним при этом, что само умение решать такие задачи человек приобретает лишь благодаря своей способности обучаться; без обучения круг доступных ему задач был бы всегда ограничен. В то же время, занимаясь исследованием путей решения сложных задач в привычной для специалиста области (а именно с такими задачами обычно имеют дело при моделировании), понять общие принципы обучения оказывается чрезвычайно трудно, так как в этом случае человек пользуется обучением лишь в очень специфическом и завуалированном виде. В результате при построении моделей, решающих такие задачи, обучение либо вообще приходится исключить из рассмотрения, как это обычно делается в большинстве игровых программ или программ, доказывающих теоремы, либо на основе самых общих соображений заменять некоторым суррогатом, как это обычно бывает с программами распознавания. Таким образом, принципы работы построенных программ оказываются весьма далекими от принципов работы человеческого интеллекта, а сами программы — приспособленными для решения лишь заранее определенного и очень узкого класса задач. Единственный выход из создавшегося положения состоит, с нашей точки зрения, в том, чтобы, отказавшись от непосредственного построения систем, решающих сложные задачи, переключить внимание на выяснение принципов обучения.

Способность обучаться мы будем рассматривать в качестве основного свойства интеллекта, считая, что умение решать сложные задачи является производным, поскольку оно целиком зависит от того, как протекал процесс обучения. Так как способностью обучаться человек наделен от рождения, то принципы его обучения можно пытаться исследовать уже на самых простых задачах. По существу, мы встаем на точку зрения Тьюринга, который писал: «Почему бы нам вместо того, чтобы пытаться создать программу, имитирующую ум взрослого, не попытаться создать программу, которая бы имитировала ум ребенка? Ведь если ум ребенка получает соответствующее воспитание, он становится умом взрослого человека. Как можно предположить, мозг ребенка в некотором отношении подобен блокноту, который мы покупаем в киоске: совсем небольшой механизм и очень много чистой бумаги. Наш расчет состоит в том, что механизм в мозгу ребенка настолько несложен, что устройство, ему подобное, может быть легко спрограммировано».

В связи с этим будем полагать, что интеллект состоит из двух компонент: 1) *механизма интеллекта*, неизменного на протяжении всей человеческой жизни и включающего в себя способы за-

полнения и использования памяти, которыми определяется, в частности, способность интеллекта к обучению, и 2) *содержимого памяти*, которое формируется в процессе обучения и которым, по существу, определяется то, что обычно называют степенью интеллектуальности человека.

Чтобы построить модель интеллекта, достаточно разобраться в устройстве механизма интеллекта, поскольку содержимое памяти представляет собой лишь отражение его деятельности. Для этого прежде всего попытаемся понять, какими дополнительными свойствами он обладает.

Основной функцией этого механизма является управление поведением человека. Под этим мы понимаем следующее. Человек живет во внешнем мире, информацию о состоянии которого он получает с помощью своих органов чувств. У человека есть эффекторы, посредством которых он может совершать какие-то действия и тем самым изменять состояние внешнего мира. В свою очередь поступающая информация приводит к тем или иным изменениям в состоянии самого человека. В каждом своем состоянии человек получает большее или меньшее удовольствие. Целью механизма интеллекта является такая организация поведения, чтобы возможно чаще приходить в состояния, приносящие возможно большее удовольствие. При этом мы исходим из того, что нет интеллекта, специально приспособленного для решения только зрительных или только слуховых задач, задач ведения разговора или движения — все они, как нам кажется, решаются одним и тем же механизмом. Переводя на единый внутренний язык и сопоставляя данные, поступающие от различных органов чувств, он находит в них общие закономерности и использует их для планирования тех или иных действий. Принимая во внимание такую неспециализированность механизма интеллекта, мы можем упростить процесс его исследования, позволив себе не рассматривать проблемы, связанные со спецификой различных органов чувств.

2. Игра. Чтобы построить модель механизма интеллекта, было бы, очевидно, желательно пронаблюдать его действия на ранних этапах обучения, когда можно проследить за первыми шагами заполнения и использования памяти. Казалось бы, самое естественное — это обратиться к исследованию поведения ребенка. Однако на ранних стадиях обучения контакт с ребенком настолько слаб, что трудно понять, какую из задач и с какой целью он решает. Попытаемся пойти несколько иным путем. Попробуем исследовать механизм интеллекта на взрослом человеке (испытуемом), поставив его в экспериментальные условия, когда, не имея возможности пользоваться каким-либо своим прошлым опытом, он должен будет решать некоторую последовательность усложняющихся задач. Принимая во внимание изложенные выше представления о свойствах интеллекта, можно сформулировать следующие требования к модельному миру, в котором должен учиться испытуемый. Во-первых, чтобы исключить возможность поль-

зования прошлым опытом, мир этот должен быть для него непривычен. Во-вторых, чтобы упростить процесс исследования поведения испытуемого, задачи этого мира должны быть минимально специализированы для его органов чувств. И, наконец, в-третьих, этот мир должен быть в определенном смысле адекватен по своей структуре реальному миру, чтобы имелась возможность постановки любых встречающихся человеку задач.

Всем этим требованиям удовлетворяет мир реальных событий, описываемых, однако, на незнакомом испытуемому языке. Эксперимент будет сводиться к попытке обучить испытуемого адекватным этому миру реакциям, ведя с ним разговор (описывая ситуации, задавая вопросы, ставя задачи и пр.) на незнакомом ему языке. В этом случае испытуемый не будет знать, о чем идет речь и поэтому не сможет пользоваться своими прежними знаниями о мире для выбора правильных реакций на те или иные сообщения учителя. Он как бы заново будет помещен в этот мир и заново займется его исследованием. Конечно, по мере изучения этого мира у него будут возникать ассоциации с его представлениями о реальном мире, и дальнейшее его поведение будет определяться этими представлениями. Однако, как показывает опыт, такие ассоциации возникают относительно нескоро, и мы надеемся, что всю интересующую нас информацию удастся выявить до того, как эти ассоциации возникнут.

Эксперимент будет вестись в рамках следующей игры. Экспериментатор (учитель) будет произносить какие-то слова или фразы и, в зависимости от ответов испытуемого, ставить те или иные оценки — других способов общения испытуемого с экспериментатором и внешним миром не будет. Цель испытуемого в этой игре — научиться получать возможно более высокие оценки. Цель учителя — научить испытуемого языку и, в определенном смысле, правильным реакциям на поступающие сообщения. При этом оценка реакции может, вообще говоря, зависеть не только от непосредственно полученного, но и от предшествующих сообщений, т. е. от предыстории.

Несмотря на малую пропускную способность входного канала испытуемого в этой игре, при достаточном времени можно надеяться на успешное его обучение. Основания для такой надежды дают известные примеры Елены Келлер и Ольги Скороходовой. Отсутствие зрения и слуха не помешало каждой из них получить высшее образование и написать ряд книг. То обстоятельство, что получение высоких оценок учителя является единственной потребностью испытуемого, значительно упрощает процесс его направленного обучения (по крайней мере, на его начальных стадиях). В то же время это не лишает его и возможностей самообучения. Так, например, стремясь получать высокие оценки в будущем, он может начать исследовать поведение учителя, а научившись понимать язык учителя, он получит возможность задавать ему вопросы, например, относительно обстановки во внешней среде.

Таким образом, он сможет установить влияние этой обстановки на поведение учителя, начнет понимать, что учителю нравится, а что — нет, и на поздних этапах обучения сможет давать учителю полезные советы относительно выполнения тех или иных действий во внешней среде, постановки в ней тех или иных экспериментов и пр. В итоге интеллект испытуемого, работая на потребности учителя и используя его средства общения с внешним миром, будет развиваться практически независимо от интеллекта учителя и в своем развитии может, вообще говоря, превзойти его.

Для удобства и ускорения процесса обучения разговор испытуемого с учителем заменим их письменным общением, представляя сообщения одного и другого в виде слов (последовательностей букв). Письменное общение посредством слов оказывается удобным еще и потому, что для описания алгоритмов преобразования слов существуют достаточно хорошо разработанные формальные языки, которыми можно будет воспользоваться при построении модели интеллекта испытуемого. Для обучения подойдет любой язык, слова которого не будут вызывать у испытуемого каких-либо смысловых ассоциаций. Такой язык легко построить, например, перекодировав соответствующим образом слова русского языка. Ниже приводится пример реальной игры с испытуемым, которая велась именно на таком языке, а также словарь русских значений используемых в игре терминов.

Словарь русских значений используемых в игре терминов

a — скажи, *p* — мама, *c* — Света, *j* — чтобы, *el* — иди,
e — шла, *s* — скорей, *r* — домой, *k* — карандаш, *b* — Боря,
i — имеет, *h* — что, *n* — не знаю, *r* — Федя, *g* — Женя.
d — дает, *m* — мяч,

Например, *bdkg* — Боря дает карандаш Жене, *hif* — что имеет Федя? *kkkm* — три карандаша и один мяч. Оценка учителя в данном варианте игры могла принимать лишь два значения (+ и —), а цель испытуемого сводилась, соответственно, к получению как можно большего числа плюсов.

Очевидно, эту игру можно было бы продолжать сколь угодно далеко, вводя в нее новые термины с их естественными смысловыми отношениями и тем самым как бы расширяя мир испытуемого.

Интересно отметить, что даже в конце приведенного примера игры, когда испытуемый нашел алгоритмы, позволяющие ему правильно реагировать на относительно широкий класс сообщений учителя, у него не возникло ни малейших ассоциаций относительно подразумеваемого смысла используемых терминов. Он лишь установил некоторые формальные отношения между ними.

На приведенном примере можно проследить, насколько естественно велось обучение испытуемого. В частности, видно, как по мере обучения постепенно расширяются возможности учителя. Так, если в начале игры учитель мог рассчитывать лишь на слу-

№	Вопрос	Ответ	Оценка	№	Вопрос	Ответ	Оценка	№	Вопрос	Ответ	Оценка
1	ap	ap	—	39	hic	n	—	77	hig	n	+
2	p	p	+	40	ak	k	+	78	fdkb	∅	+
3	ap	p	+	41	hic	k	+	79	hib	k	+
4	ac	c	+	42	hif	∅	—	80	hif	n	+
5	acel	l	—	43	hif	k	—	81	hic	k	+
6.	acel	cel	+	44	hif	n	+	82	cdkb	∅	+
7	acelr	celr	+	45	hig	n	—	83	hib	n	—
8	ajce	jce	—	46	cdkg	∅	+	84	akk	kk	+
9	acel	cel	+	47	hig	k	+	85	hib	kk	+
10	ajce	ce	—	48	hif	k	—	86	hic	k	—
11	acel	cel	+	49	hif	n	+	87	hic	n	+
12	ajce	e	—	50	hic	k	—	88	hif	n	+
13	apel	pel	+	51	hic	n	+	89	bdkg	∅	+
14	ajpe	p	—	52	hib	k	+	90	hib	k	+
15	apel	pel	+	53	bdkf	∅	+	91	hig	k	+
16	ajpe	j	—	54	hif	k	+	92	hif	n	+
17	apel	pel	+	55	hig	n	+	93	fik	∅	+
18	ajpe	jpe	—	56	hib	k	+	94	hif	k	+
19	ajce	ajce	—	57	hic	n	+	95	fdkg	∅	+
20	acel	cel	+	58	bdkc	∅	+	96	hig	kk	+
21	ajce	cel	+	59	hic	k	+	97	bdkg	∅	+
22	ajpe	pel	+	60	hib	n	+	98	hig	kkk	+
23	ajcer	cerl	—	61	hif	k	+	99	fim	∅	+
24	acelr	celr	+	62	cdkf	∅	+	100	hif	m	+
25	ajcer	celr	+	63	hib	n	+	101	fdmb	∅	+
26	ajpe	pel	+	64	hic	n	+	102	hib	k	—
27	ajpes	pels	+	65	hig	n	+	103	am	m	+
28	ajcers	celrs	+	66	hif	n	—	104	hib	m	+
29	bik	bik	—	67	hif	k	—	105	bdmc	∅	+
30	a	∅	+	68	akk	kk	+	106	hic	m	+
31	bik	∅	+	69	hif	kk	+	107	hib	n	+
32	hib	hi	—	70	fdkc	∅	+	108	gdkc	∅	+
33	ak	k	+	71	hic	n	—	109	hic	mm	—
34	hib	k	+	72	ak	k	+	110	hic	mk	+
35	hic	k	—	73	hic	k	+	111	cdmg	∅	+
36	an	n	+	74	hif	n	—	112	hig	kkm	+
37	hic	n	+	75	hif	k	+				
38	cik	∅	+	76	hib	n	+				

чайную выдачу правильных ответов, то потом у него появляется возможность подсказывать правильные ответы и задавать «наводящие» вопросы. Легко себе представить, что, когда испытуемый в достаточной мере усвоит используемый учителем язык, то у последнего появится возможность вводить в явном виде определе-

ния новых терминов, сообщать испытуемому алгоритмы, по которым должны даваться правильные ответы, и пр. Все это, очевидно, вполне согласуется с нашими обычными представлениями о поведении ребенка при обучении в реальном мире. Поэтому можно надеяться, что, построив систему, которая бы имитировала поведение испытуемого, мы могли бы учить ее не менее успешно, чем учим ребенка. Построение такой системы и будет нашей дальнейшей целью.

3. О формализации задачи. Итак, задачу построения искусственного интеллекта мы свели к более частной и, на первый взгляд, более простой задаче построения системы, имитирующей поведение испытуемого в играх описанного выше типа. Заметим, что задача эта является не столько математической, сколько естественнонаучной, а потому далеко не формальной. Бессмысленно было бы сейчас, пока не принимаются во внимание никакие конкретные свойства человеческого поведения, заниматься сколько-нибудь детальной ее формализацией. Система, удовлетворяющая условиям такой формальной задачи, все равно оказалась бы не адекватной действительности и не удовлетворяла бы нашим желаниям.

Многие приемы, используемые обычно при формализации задачи распознавания, которая, казалось бы, очень близка к нашей задаче, здесь оказываются непригодными. Так, мы не можем, например, считать, как это обычно принято в задачах распознавания, что распределение вероятностей на множестве ситуаций, в которые может попадать система, не меняется с ее обучением, поскольку человек выбирает область, в которой будет работать, обычно в зависимости от того, что он умеет. Мы не хотим также с самого начала формально ограничить класс решающих правил, которые будет строить система, поскольку считаем, что для алгоритмов, которые способен находить человек, вряд ли существуют формальные ограничения. Поэтому мы будем уточнять свою задачу постепенно в самом процессе построения системы.

4. Цели системы. Связь системы с учителем будет осуществляться с помощью двух входов X и Z и одного выхода Y, а ее работа внешне будет выражаться в циклическом повторении следующих событий: 1) получение на входе X сообщения (вопроса) x_0, 2) выдача на выходе Y ответа y_0 и 3) получение на входе Z оценки z_0. Таким образом, историю внешней работы системы можно описать последовательностью слов (вопросов, ответов и оценок)

$$x_{-k}, y_{-k}, z_{-k}; \ldots; x_{-1}, y_{-1}, z_{-1}; x_0, y_0, z_0; x_1, y_1, z_1, \ldots,$$

где нулевой индекс (начало отсчета) приписан циклу (вопрос, ответ, оценка), принимаемому за *настоящее*, а отрицательные и положительные индексы, соответственно, циклам *прошлого* и *будущего*.

Начальный кусок этой последовательности, заканчивающийся вопросом x_0, будем называть описанием *внешней ситуации* S,

в которой дается ответ y_0. Оценка z_0 этого ответа может в общем случае зависеть (см. таблицу) не только от вопроса x_0, но и от любых предшествующих ему вопросов, ответов и оценок, т. е. определяется ситуацией S. В силу той же зависимости даваемый ответ y_0 может, вообще говоря, влиять не только на настоящую оценку z_0, но и на оценки z_i, получаемые в будущем. Поэтому естественно исходить из того, что система при выборе ответа должна стремиться к увеличению не только настоящей, но и этих будущих оценок, подобно тому, как человек планирует свои действия в расчете на возможность удовлетворения своих потребностей не только в данный момент, но и в будущем. Для простоты можно, например, считать, что *цель системы* при выборе ответа y_0 состоит в максимизации функции $\rho_n(y_0) = \sum_{i=0}^{n-1} z_i$, где n — число циклов, т. е. суммы прогнозируемых ею значений оценок на некотором отрезке будущего.

Вообще говоря, не в любой ситуации у системы будет возможность давать ответы с одинаково высокими оценками z_i. Поэтому, стремясь к своей цели, система должна, во-первых, уметь предсказывать оценку даваемого ею ответа, во-вторых, моделировать дальнейшее поведение учителя, чтобы знать, сможет ли она и на последующие его вопросы давать достаточно высоко оцениваемые ответы. Для того и для другого необходимы знания, которые в систему изначально не закладываются. Они постепенно будут приобретаться ею в процессе обучения. На ранних этапах обучения надежность этих знаний очень мала. Поэтому экстраполяция оценок, которые будет получать система на сколько-нибудь длительном отрезке будущего, становится бессмысленной, так как надежность такой экстраполяции с ростом длины отрезка будущего падает, грубо говоря, экспоненциально. Другими словами, необученная система оказывается не в состоянии преследовать далекие цели. Поэтому мы будем считать, что цель системы зависит от степени ее обученности, и будем рассматривать целевую функцию $\rho = \rho_n$, где значение параметра n (*глубины планирования* поведения системы) не фиксировано заранее и может быть сколь угодно большим. Система всякий раз будет планировать свое поведение на наибольшую глубину, при которой экстраполяция получаемых ею оценок будет еще достаточно надежной. Наличие такого свойства у человека можно проиллюстрировать на примере игры в шахматы, когда начинающий игрок не способен планировать свою игру более, чем на один ход, а опытный шахматист для получения в будущем позиционного или материального преимущества может преднамеренно идти на потерю какой-либо из своих фигур.

Внутреннюю работу системы можно, грубо говоря, разбить на два основных процесса: 1) обучение, т. е. приобретение знаний, и 2) использование этих знаний, направленное на максимизацию

целевой функции ρ. Естественно считать, что *цель обучения* состоит в построении алгоритма $F(S)$, формирующего внешнее поведение системы, т. е. определенного на возможно более широком классе ситуаций S и позволяющего системе строить ответ y_0, оптимальный с точки зрения оценки ρ, с возможно большей глубиной планирования. Таким образом, цель обучения определяется общей целью системы.

Естественно, что в начале обучения прежде всего формируется алгоритм $F(S)$ с глубиной планирования $n = 1$, который в зависимости от ситуации S строит ответ y_0 с наиболее высокой оценкой z_0. Для этого, очевидно, не требуется знаний того, какими будут дальнейшие вопросы учителя, и поэтому задача построения такого алгоритма оказывается существенно проще общей задачи обучения. Лишь накопив достаточный опыт и установив некоторые закономерности в поведении учителя, система сможет перейти к большей глубине планирования. Поскольку при фиксированной глубине планирования надежность экстраполяции системой значений будущих оценок оказывается, вообще говоря, разной, то естественно, что глубина планирования будет зависеть от ситуации.

Все это вполне согласуется с нашими представлениями о поведении человека. Так, например, на всем протяжении приведенного нами примера игры испытуемый так и не пытался перейти к глубине планирования $n > 1$, поскольку не мог найти никаких четких закономерностей в поведении учителя. В то же время легко привести пример игры, где поведение учителя просто формализуется, и испытуемый поэтому достаточно быстро начинает планировать на глубину $n > 1$.

Заметим, что в приведенной нами схеме работы системы нигде в явном виде не фигурировал параметр времени — рассматривалась лишь упорядоченность событий во времени. В этом некоторая неполнота нашей схемы, так как, во-первых, такая система в отличие от человека не сможет рассматривать временны́е закономерности, связывающие ее поведение с внешним миром, т. е. учителем, во-вторых (как следствие первого), она, вообще говоря, вправе не принимать во внимание скорость своей работы.

Первый недостаток нам не кажется существенным. Его можно будет потом исправить, введя в уже построенную систему некоторые дополнительные элементы. Мы не делаем этого сейчас, так как не хотим пока загромождать систему непринципиальными, с нашей точки зрения, деталями. Куда более важным нам представляется второй недостаток, поскольку система, у которой каждый акт выдачи ответа будет требовать практически бесконечного времени, нас, конечно же, не устраивает. Поэтому мы еще несколько видоизменим целевую функцию системы. Будем считать, что работа системы протекает в реальном времени, но состояния ее входов и выхода могут рассматриваться лишь в дискретные моменты времени через равные и достаточно малые интервалы. Будем, кроме того, считать, что учитель не тратит времени на то,

чтобы поставить системе оценку и задать следующий вопрос — это происходит непосредственно в момент выдачи ответа, и лишь в эти моменты времени значение на оценочном входе Z может быть отлично от ε — отрицательной величины, малой по сравнению с оценкой ответа.

Теперь целевую функцию системы будем представлять в виде суммы значений на входе Z по всем моментам интервала времени, соответствующего глубине планирования в функции ρ. Поскольку цель системы при выборе ответа y_0 будет состоять в максимизации зависящей от времени целевой функции, то для успешной работы ей будет необходимо отвечать не только правильно, но и возможно быстрее. Системе придется выбирать свои внутренние действия (на какое понятие в первую очередь обратить внимание, какое произвести преобразование и т. п.) с таким расчетом, чтобы исключить излишние операции и следовать по пути, быстрее всего приводящему к решению поставленной задачи. Таким образом, выбираемое внутреннее действие должно быть оптимально и относительно затрачиваемого на его поиск времени.

Целесообразность тех или иных внутренних действий, так же как и внешних, очевидно, зависит от того, с каким классом задач системе приходится иметь дело, т. е. определяется устройством внешнего мира, которое изначально системе не известно, но может ею постепенно познаваться в процессе обучения. Поэтому для достижения своей цели *система должна быть обучаемой на всех этапах организации своего поведения, где имеется возможность альтернативного выбора действия.*

5. Построение ответа в незнакомой ситуации. Вернемся теперь к рассмотрению приведенного выше примера игры и попытаемся проследить за поведением испытуемого (ученика), с тем чтобы построить какую-то, пока чисто описательную, модель этого поведения.

На первый взгляд, казалось бы естественным предположить, что, пока у испытуемого не накопится определенный опыт, и он не найдет закономерностей, позволяющих устанавливать, какие ответы верны (с оценкой $+$), а какие — нет (с оценкой $-$), у него не должно быть никаких оснований предпочитать одни ответы другим, т. е. все мыслимые ответы должны быть для него априори равновероятны. Однако, как показывает эксперимент, дело обстоит далеко не так. Никто из испытуемых не пытается давать в начале игры сколько-нибудь «сложных» ответов. В частности, практически все испытуемые с самого или почти с самого начала игры пытаются использовать в качестве ответа (см. п.п. 1, 2 таблицы) сам вопрос, и лишь в случае неудачи меняют стратегию. Такое поведение испытуемых можно объяснить наличием у человека некоторых априорных методов поведения (эвристик), позволяющих с самого начала в зависимости от ситуации отдавать предпочтение тем или иным ответам.

Именно наличием эвристик определяется приспособленность

человека к тому миру, в котором он живет. Они позволяют ему в неисследованных ситуациях совершать действия, которые с достаточно большой вероятностью оказываются «правильными» (полезными). Так накапливается опыт, обобщив который, человек получает возможность действовать целенаправленно. Отсутствие такого рода эвристик привело бы, очевидно, к тому, что этот опыт накапливался чрезвычайно медленно, и обучение было бы практически неосуществимо.

Сформулируем некоторые эвристики, применением которых можно было бы объяснить поведение испытуемых на ранних этапах нашей игры.

Эвристика 0.1. С большой вероятностью, правильный ответ можно построить из тех букв, которые уже встречались в вопросах.

Эвристика 0.2. С большой вероятностью правильным окажется ответ, тождественный вопросу. (Эта эвристика очень похожа по своим функциям на инстинкт подражания, часто наблюдаемый у детей.)

Эвристика 0.3. С большой вероятностью правильным окажется ответ, тождественный одному из предшествующих правильных ответов (чем позже был дан этот предшествующий ответ, тем вероятнее успех при его повторении).

Эвристика 0.4. С большой вероятностью ответ B, даваемый на вопрос A, однажды оказавшись правильным (неправильным), будет таким же и в следующий раз. (Эта эвристика может, очевидно, не только рекомендовать ответы, но и запрещать их.)

Используемый в формулировках эвристик термин «вероятность» можно пока интерпретировать как априорную вероятность истинности соответствующих высказываний об ответе. Употребление именно этого термина, вообще говоря, не является необходимым, поскольку нас здесь интересует лишь определяемый этими вероятностями порядок рассмотрения различных гипотез об ответах. Вводя вероятности, мы всего лишь конкретизируем механизм этого упорядочения. Однако, если учесть, что эвристики могут использоваться совместно, причем один и тот же ответ может «рекомендоваться» одними эвристиками и «запрещаться» другими, то введение этого термина становится вполне оправданным, поскольку позволяет воспользоваться обычными методами вычисления вероятностей, развиваемыми в индуктивной логике (см. также раздел 9).

Следует подчеркнуть, что обучение в рамках нашей игры может протекать успешно лишь в случае, если учитель сознательно или бессознательно будет строить свое обучение в расчете на наличие у испытуемого тех или иных эвристик. В нашем случае это особенно относится к эвристикам 0.2—0.4. В частности, рассчитывая сначала на использование эвристики 0.2, а затем на совместное использование эвристик 0.3 и 0.4 (последней 0.4 — как запрещающей), учитель относительно быстро добивается нужного

ему ответа на тот или иной вопрос (см. пп. 2, 3 таблицы). Это позволяет испытуемому быстро накопить опыт, обобщение которого даст ему возможность не только правильно отвечать на вопросы данного типа (см. пп. 6, 7), но и сформировать правило, позволяющее ему в дальнейшем воспринимать «подсказку» учителя (см. пп. 21, 24). Это правило после соответствующего перевода можно сформулировать в следующем виде: «если на вопрос *A* был дан неправильный ответ, а следующий вопрос (сообщение) имел вид «скажи *B*», то на *A* нужно давать ответ *B*» (слово «скажи» в данном случае правильнее интерпретировать как «нужно говорить»).

Нас будут, конечно, интересовать следующие вопросы. Много ли у человека таких эвристик? В каком виде они заложены и как реализуются? Являются ли уже найденные нами эвристики исходными, или они всего лишь некоторые следствия каких-то других более общих эвристик? Чтобы разобраться в этих вопросах, нам придется проанализировать дальнейшие действия испытуемого и проследить за тем, как он находит закономерности.

6. Обучение (поиск закономерностей). Построив некоторое количество правильных и неправильных ответов, испытуемый пытается обобщить накопленный опыт. Как уже отмечалось, на ранних этапах обучения целью этого обобщения является построение алгоритма $F(S)$ с единичной глубиной планирования, который позволит с достаточно высокой надежностью экстраполировать правильные ответы в возможно более широком классе ситуаций S. Казалось бы самым простым способом поиска такого алгоритма является обычный метод проб и ошибок. Этот метод в грубом виде сводится к синтезу всевозможных алгоритмов (например, с помощью некоторого порождающего процесса) и последующему отбору среди них алгоритмов, удовлетворяющих критерию правильности, т. е. таких, которые во всех встречавшихся ранее ситуациях строят все правильные ответы и не строят ни одного неправильного. Здесь, однако, приходится столкнуться с той же трудностью, что и в случае выдачи ответа. При произвольном порядке рассмотрения алгоритмов неизвестно, как долго придется вести такой поиск, прежде чем удастся найти «подходящий» алгоритм. Но самое опасное даже не в этом. В подавляющем большинстве случаев найденные таким образом алгоритмы не будут обладать способностью к правильной экстраполяции — в ситуациях, отличных от уже встречавшихся, эти алгоритмы почти всегда будут давать неправильные ответы, т. е. обучение не будет достигать своей цели.

У человека процесс формирования алгоритма выдачи ответов протекает иначе. Прежде всего оказывается, что он совсем не сводится к рассмотрению всевозможных алгоритмов и проверке их по критерию правильности. Это можно продемонстрировать на следующем простом примере игры. В этом примере мы для краткости оставили лишь те строки, которым соответствует по-

ложительная оценка ответа и для которых должен быть найден единый алгоритм преобразования вопроса в ответ.

Вопрос — $x(S)$	*Ответ — $y(S)$*
аваавсвв	ава
ввввваасва	вввв
ававасвва	ав
вавасв	вав
ваавасавва	в
авввавса	аввва

Следует напомнить, что для испытуемого эти строки появляются на самом деле постепенно. Между ним могут быть и ошибки, и, чтобы от них быстрее избавиться, ему необходимо искать закономерности с появлением первой же такой строки.

Почти сразу испытуемый обнаруживает, что правильный *ответ является некоторым* (пока еще не известно какой длины) *началом вопроса.* Следует подчеркнуть, что это утверждение формулирует лишь некоторое частное свойство задачи (хотя конечная цель обучения состоит в формировании алгоритма выдачи ответа). Оно не определяет ответ однозначно и поэтому не может быть найдено в процессе рассмотрения различных алгоритмов преобразования вопроса. В то же время это утверждение описывает некоторую закономерность, связывающую вопрос с правильным ответом. Отыскав эту закономерность на малом числе строк, человек с ее помощью как бы частично решает задачу — он ограничивает область, в которой в дальнейшем ищет правильный ответ. Лишь построив таким образом достаточное количество правильных ответов, человек использует то же самое утверждение уже в новом качестве, а именно, как условие целесообразности отделения от вопроса его начального куска, тождественного ответу. Такая операция также не могла бы быть проделана при рассмотрении лишь различных преобразований вопроса, так как в качестве переменного параметра такого преобразования здесь используется ответ. Заметив далее, что остаток вопроса содержит единственную букву *с* и, используя это утверждение в качестве условия целесообразности разбиения этого остатка по букве *с* на две части, человек обнаруживает, что *эти части имеют равное число букв.* Это утверждение также не является описанием алгоритма, однако в совокупности с первоначально найденным утверждением оно определяет ответ уже однозначно. Поэтому, найдя его, человек быстро строит алгоритм выдачи ответа. Для этого он, используя логический вывод, разрешает выделенные курсивом утверждения относительно ответа как некоторой неизвестной величины. (Подробнее об этом см. раздел 11.)

Итак, оказывается, что непосредственный поиск алгоритма выдачи ответа у человека заменен поиском закономерностей в ви-

де утверждений (высказываний), истинных для всех встречавшихся ранее ситуаций, с последующим разрешением набора этих утверждений относительно неизвестной величины — ответа — подобно тому, как многим физическим законам при поиске их окончательной формы придают с самого начала вид уравнений (которые являются частным случаем утверждений), а процесс решения конкретных задач сводится к разрешению начальных данных этих задач и относящихся к этим задачам систем уравнений (законов) относительно соответствующих неизвестных.

Естественно возникает вопрос: почему же закономерности отыскиваются человеком в виде утверждений, а не непосредственно в виде алгоритмов. Ведь в этом случае для построения ответа становится необходим вывод — процесс казалось бы существенно более сложный, чем применение алгоритма. Причины этому следующие.

1. Прежде всего, утверждения описывают закономерности более общего вида, чем алгоритмы. Если алгоритм при заданных значениях своих входов определяет значение выхода всегда однозначно, то утверждение, связывающее те же самые параметры, определяет, вообще говоря, некоторую область значений выхода. Поэтому, как правило, оказывается, что преобразование, которому соответствует достаточно сложный алгоритм, может быть описано набором относительно простых утверждений, и легче бывает найти набор утверждений, которые в своей совокупности однозначно определяют преобразование, чем устраивать непосредственный поиск сложного алгоритма.

2. Связанные утверждением понятия, будь то слова, отношения или преобразования, в определенном смысле равноправны, т. е. каждое из них может в принципе рассматриваться как неизвестное, относительно которого это утверждение должно быть разрешено. Это обстоятельство существенным образом расширяет круг промежуточных задач, в которых могут быть использованы находимые в таком виде закономерности.

3. В процессе работы практически всегда рано или поздно оказывается, что найденная в свое время закономерность опровергается какими-то противоречащими примерами. В таком случае человек пользуется следующей эвристикой. Прежде чем окончательно расстаться с этой закономерностью, он пытается найти зависящее от ситуации условие ее применимости, т. е. условие, при выполнении которого эту закономерность можно использовать для экстраполяции. Это в равной мере может относиться и к алгоритмам и к утверждениям. Таким образом, оказывается, что в памяти человека в основном хранятся именно условные закономерности. Надежность экстраполяции с помощью каждой такой условной закономерности, в общем случае тем больше, чем больше было таких ситуаций, в которых ее удалось проверить и она оказалась истинной, т. е. ситуаций, в которых выполнялось условие ее применимости [2]. Поскольку утверждения описывают

более общий вид закономерностей, чем алгоритмы, то и выполняются они, как правило, на большем числе ситуаций, а значит, для них обычно удается найти и более общие, т. е. выполнимые на большем числе ситуаций, условия применимости. Поэтому надежность условных утверждений в общем случае оказывается больше надежности условных алгоритмов.

7. Язык описания закономерностей. Наша система, подобно человеку будет находить закономерности в виде утверждений и разрешать их потом относительно неизвестных. Для описания этих закономерностей системе необходим некоторый язык. Мы приведем здесь начальный вариант такого языка. При его выборе мы исходили из следующих соображений. Во-первых, этот язык должен быть универсальным, поскольку, как уже говорилось ранее, мы не хотим ограничивать класс закономерностей, построение которых будет доступно системе. Во-вторых, поскольку мы хотим, чтобы поиск закономерностей велся достаточно естественным для человека образом, а именно, чтобы простые для человека закономерности обнаруживались системой в первую очередь, то естественно стремиться к тому, чтобы запись этих закономерностей на языке системы была возможно короче. И, наконец, в-третьих, поскольку найденные закономерности придется потом использовать для синтеза ответа, т. е. разрешать соответствующие им утверждения относительно неизвестных, то, конечно, желательно, чтобы процедура такого разрешения была бы достаточно удобной и быстрой. (Обычно используемый в таких случаях язык исчисления предикатов [3, 4] нам кажется плохо приспособленным для этих целей.)

Мы решили отойти от обычной схемы языков исчисления предикатов с кванторами и остановиться на языке, использующем некоторые идеи исчисления равенств Гудстейна [5]. Все утверждения и преобразования, описываемые на нашем языке, строятся в виде программ, команды которых являются либо операциями над словами, либо командами условного перехода. Последние могут одновременно выполнять роль элементарных утверждений. Команды в программе занумерованы. Каждая программа может рассматриваться либо как алгоритм преобразования слов, либо как запись некоторого утверждения (или системы утверждений) над словами. В последнем случае она может быть использована для вывода других утверждений или алгоритмов.

Для выполнения преобразований по данной программе в нашей системе должен быть соответствующий интерпретатор. В этом режиме команды программы выполняются последовательно в порядке возрастания их номеров, если другое не предписывается командами условного перехода. Преобразуемые слова лежат в ячейках, адреса которых указываются в записях соответствующих команд. Таким образом, адрес ячейки является как бы названием переменной, а слово, лежащее в этой ячейке,— ее значением. Длина слов не фиксирована.

В приведенном здесь варианте языка имеются следующие пять операций над словами: 1) $xy = z$ — соединение (запись подряд) слов из ячеек x и y с занесением результата в ячейку z, 2) $x\vec{y} = z, z'$, 3) $x\overleftarrow{y} = z, z'$ — разделение на части слова из ячейки x по первому слева (для 2) или справа (для 3) вхождению в него слова из ячейки y с занесением левой части в ячейку z, а правой в z', 4) $l(y)x = z, z'$, 5) $xl(y) = z, z'$ — отделение от слова из ячейки x слева (для 4) или справа (для 5) слова, равного по длине слову из ячейки y с занесением оторванной части в ячейку z (для 4) или z' (для 5), а остатка, соответственно, в ячейки z или z'.

В случае неприменимости этих команд управление передается стандартной ячейке α интерпретатора, которая играет роль «аварийного останова» программы, записанной на этом языке (но не останова всей системы!). В языке имеются следующие команды условного перехода:

1) $x \equiv y, z$; 2) $x \supseteq y, z$; 3) $l(x) = l(y), z$; 4) $l(x) \geqslant l(y), z$;

1′) $x \not\equiv y, z$; 2′) $x \not\supseteq y, z$; 3′) $l(x) \neq l(y), z$; 4′) $l(x) \not\geqslant l(y), z$.

При выполнении записанного в команде условия управление передается следующей команде программы, в противном случае оно передается команде с номером z. В этом языке имеется также команда «конец», которая передает управление стандартной ячейке β интерпретатора. С помощью ячеек α и β осуществляется связь интерпретатора с другими блоками системы.

Легко показать, что на приведенном здесь языке выразим любой из нормальных алгоритмов Маркова [6]. Поскольку язык нормальных алгоритмов принято считать одним из эталонных универсальных языков, то наш язык можно, следовательно, также считать универсальным.

Как уже говорилось, программы этого языка могут быть записями не только алгоритмов, но и утверждений (или систем утверждений). Элементарные утверждения записываются при этом в виде команд условного перехода, в последнем адресе которых стоит символ α. Символ α играет здесь ту же роль, что и квантор общности. Так, например, выражение $x \equiv y, \alpha$ следует читать как «$x \equiv y$ верно всегда» (т. е. для любых ситуаций S). Заметим, что если для данных значений x и y высказывание $x \equiv y$ окажется ложным, то при выполнении этой команды интерпретатором управление будет передано ячейке α, что соответствует требованию дальнейшего анализа и перестройки программы другими блоками системы. Сложные утверждения имеют вид преобразований, результаты которых связаны некоторыми элементарными утверждениями. Можно показать, что на приведенном здесь языке могут быть выражены любые утверждения исчисления предикатов, не использующие неограниченных кванторов существования. В частности, система утверждений, найденная в примере игры, рассмотренном в предыдущем разделе, записывается на этом

языке в следующем виде:

1. $x\vec{y} = v_1, v_2$
2. $v_1 \equiv \phi, \alpha$
3. $v_2\overleftarrow{c} = t_1, t_2$
4. $l(t_1) = l(t_2), \alpha.$

Команды 1 и 2 соответствуют записи первого из утверждений, а команды 1, 3 и 4 — записи второго. Алгоритм построения ответа для этого случая может быть записан в виде следующей программы:

1. $xc = p,\ p'$ 2. $pl(p') = y,\ q.$

8. Эвристики поиска закономерностей. Попытаемся теперь выяснить, каким именно образом протекает у человека процесс поиска закономерностей, которые, как было показано, представимы в виде некоторых утверждений, истинных для всех встречавшихся ранее ситуаций. В общем случае эти утверждения должны, очевидно, устанавливать зависимость между 1) даваемым ответом, 2) ситуацией, в которой этот ответ дается, и 3) целевой функцией ρ.

Следует напомнить, что описание ситуации, в которой дается ответ, включает в себя описание всей истории внешнего поведения человека, т. е. представляет собой последовательность всех предшествовавших вопросов, ответов и оценок. С течением времени это описание удлиняется и становится все менее обозримым. Для того чтобы поиск закономерностей при этом не усложнялся, человеку необходимы специальные механизмы, позволяющие обращать внимание лишь на достаточно важные для настоящего момента элементы описания ситуации. Эти механизмы и само понятие *важности* оказываются в общем случае применимыми не только к элементам описания ситуации, но и к любым другим используемым понятиям (операторам, высказываниям, задачам и пр.). *Важность* понятия мы рассматриваем как приписанную этому понятию величину, от которой зависит вероятность обращения на это понятие *внимания* (подробнее об этом см. раздел 9). При этом естественно считать, что вероятность применения оператора P (например, предиката, преобразования, правила вывода) к понятиям $x_1, x_2, \ldots, x_k$, для которых этот оператор определен, тем больше, чем больше важности P и x_i $(i = 1, 2, \ldots, k)$. *Важность* понятия можно также интерпретировать как величину, определяющую априорную (субъективную) вероятность того, что это понятие удастся существенным образом использовать при решении поставленной задачи.

Пользуясь *важностями*, человек может вести поиск решения задачи, начиная с рассмотрения самых важных понятий и постепенно переходя к все менее и менее важным, пока либо не решит задачу, либо откажется от ее решения, так как *важность* остав-

шихся понятий столь мала, что никакие их преобразования не дают ему надежды на успех.

Исходным понятиям, таким, как ответ, ситуация, целевая функция, элементарные операторы, изначально приписаны достаточно высокие *важности*. Другие понятия строятся из исходных с помощью эвристик, представляющих собою механизмы, которые, в зависимости от *важностей* рассматриваемых понятий, преобразуют (или не преобразуют) эти понятия в другие понятия и приписывают последним определенные значения *важности*.

Элементами описания ситуаций могут быть либо слова-константы, либо переменные величины, для которых определен способ получения их значений в произвольной ситуации. Такие переменные мы будем называть ситуационными в отличие от свободных переменных, обозначающих места аргументов в функциях и предикатах. Простейший способ получения ситуационных переменных дает уже само представление ситуации в виде последовательности: настоящий вопрос, предшествующая ему оценка, предыдущий ответ, предыдущий вопрос и т. д. Для таких переменных имеет место следующая эвристика.

Эвристика 1.1. Чем дальше в прошлом находится некоторая ситуационная переменная (вопрос, ответ или оценка), тем меньше ее *важность* (т. е. тем меньше субъективная вероятность того, что правильность даваемого ответа должна зависеть от значения этой переменной) [1].

Заметим, что эвристику 0.4 можно, вообще говоря, рассматривать как некоторое следствие эвристики 1.1, так как эвристика 0.4 может быть переформулирована в следующем виде: «с большой вероятностью правильность ответа зависит только от вопроса».

Приведем теперь примеры эвристик, с помощью которых строятся высказывания, предикаты и преобразования с приписыванием этим понятиям значений *важности*. Следует заметить, что приводимые здесь примеры представляют собой лишь частные случаи некоторых других более общих эвристик, которыми, как нам кажется, пользуется человек и которые предполагается заложить в нашу систему. Мы описываем их здесь в таком частном виде лишь для того, чтобы, не вводя громоздких определений, проиллюстрировать с их помощью процесс поиска закономерностей для примера игры, описанного на стр. 214.

Эвристика 1.2. Если $x\,(S)$ — ситуационная переменная, принимающая в данной ситуации значение a, то чем больше *важность* x, тем больше *важность* высказывания $\forall\, S\ (x\,(S) \equiv a)$.

Эвристика 1.3. Если $x_1(S)$, $x_2(S)$, . . ., $x_k(S)$ — ситуационные переменные и $P\,(u_1, u_2, \ldots, u_k)$ — некоторый предикат со сво-

---

[1] Поскольку точные формулировки эвристик должны основываться на конкретном способе вычисления *важностей* и требуют достаточно громоздких определений, что не входит в задачи настоящей статьи, то все формулировки даются здесь лишь в интуитивно понятном виде.

бодными переменными, то чем больше *важности* P и $x_i (i = 1, 2, \ldots, k)$, тем больше *важность* высказывания $\forall\, SP\,(x_1(S), x_2(S), \ldots, x_k(S))$.

Замечание. Эвристики 0.2 и 0.3 начальной выдачи ответа можно рассматривать как следствия этой эвристики.

Эвристика 1.4. Если $x\,(S)$ — ситуационная переменная, то чем больше ее *важность*, тем больше 1) *важность* оператора $\Pi K(x\,(S))$ — поиска контрастной (например, отличающейся по алфавиту от остальных частей) части значения x, соответствующего данной ситуации, и 2) *важность* высказывания $\forall\, S\,(x\,(S) \geqslant b)$, если такая часть b найдена.

Замечание. Несмотря на то, что оператора ПК (u) нет в приведенном нами варианте языка, мы рассматриваем его как исходный, которому изначально приписана достаточно высокая *важность*. Именно наличием этого оператора можно объяснить то, что новорожденный ребенок следит глазами за яркими предметами, обращает внимание на резкие звуки и т. п. Благодаря этому оператору человек обращает внимание на всевозможные неоднородности в рассматриваемых им описаниях событий и, выделяя с помощью этих неоднородностей отдельные элементы этих описаний, строит более короткие и обобщенные описания, в терминах которых ищет потом закономерности.

Эвристика 1.5. Чем больше *важность* высказывания $\forall S\,(x_1\,(S) \supseteq x_2(S))$, тем больше 1) *важности* операций $x_1\vec{x}_2 = r_1 r_2$ и $x_1\overleftarrow{x}_2 = r_1$, r_2 разделения x_1 по вхождению x_2 и 2) *важности* переменных $r_1\,(S)$ и $r_2\,(S)$, являющихся результатами такого разделения.

Эвристика 1.6. *Важность* высказывания $\forall\, SP\,(x_1\,(S), x_2\,(S), \ldots, x_k\,(S))$ тем больше, чем больше ситуаций S, в которых оно было проверено с помощью предиката $P\,(u_1, u_2, \ldots, u_k)$ и оказалось истинным. Если оно не было ложным, то это — закономерность.

Эвристика 1.7. Чем больше *важность* высказывания, тем больше *важность* любого из понятий, используемых в этом высказывании.

Замечание. Эвристики 1.6 и 1.7 позволяют увеличивать *важности* предикатов, операций и других используемых в высказывании понятий при удачной проверке этого высказывания.

Попытаемся теперь с помощью приведенных здесь эвристик описать процесс поиска закономерностей для примера игры, представленного на стр. 214. Для описания этих закономерностей мы будем пользоваться языком, введенным в предыдущем разделе. Поэтому употребляемый для упрощения формулировок эвристик квантор общности будет теперь заменен символом α. В разбираемом примере мы для краткости нигде в явном виде не будем рассматривать использование эвристики 1.1, хотя уже само выделение представленной части описания ситуации предполагает, в частности, использование этой эвристики. Той ветви дерева поиска

закономерностей, которая приводит к решению, можно поставить в соответствие следующую последовательность процедур.

1. Поскольку предикат $u_1 \geqslant u_2$, вопрос x и ответ u исходно являются достаточно важными понятиями, то в соответствии с эвристикой 1.3 высказывание $x \geqslant y$, α получает достаточно высокую *важность*.

2. В силу *важности* это высказывание проверяется на истинность в каждой ситуации и в соответствии с эвристикой 1.6 с каждой такой проверкой важность этого высказывания растет. В результате высказывание $x \geqslant y$, α оказывается закономерностью.

3. Согласно эвристике 1.5 осуществляется действие $x\vec{y} = v_1$, v_2 и переменным v_1 и v_2 приписывается достаточно высокая *важность*.

4. В силу *важности* переменной v_1 согласно эвристике 1.2 высказыванию $v_1 \equiv \phi$, α приписывается достаточно высокая *важность*.

5. Согласно эвристике 1.6 с каждой проверкой на истинность высказывания $v_1 \equiv \phi$, α его *важность* растет, и в результате это высказывание оказывается закономерностью.

6. В силу *важности* переменной v_2 согласно эвристике 1.4 на первой же строке примера находится контрастная часть c значения этой переменной, и высказыванию $v_2 \supset c$, α приписывается высокая *важность*.

7. Согласно эвристике 1.6 *важность* этого высказывания с каждой его проверкой на истинность увеличивается, и в результате высказывание $v_2 \geqslant c$, α оказывается закономерностью.

8. Согласно эвристике 1.5 осуществляется действие $v_2\vec{c} = t_1$, t_2 и переменным t_1 и t_2 приписывается достаточно высокая *важность*.

9. В силу *важности* элементарного предиката $l(u_1) = l(u_2)$ и переменных t_1 и t_2 в соответствии с эвристикой 1.3 высказывание $l(t_1) = l(t_2)$, α получает достаточно высокую *важность*.

10. Согласно эвристике 1.6 *важность* этого высказывания с каждой проверкой увеличивается, и в результате высказывание $l(t_1) = l(t_2)$, α оказывается закономерностью.

Итак, найдена система закономерностей, однозначно определяющая ответ (см. стр. 225).

Из приведенного здесь примера видно, что *эвристики, поставляя материал друг для друга, образуют единую рекурсивную схему анализа ситуаций и поиска закономерностей.*

9. Память и механизм обращения внимания. Обсудим теперь некоторые вопросы, касающиеся устройства памяти человека и его механизма вспоминания и обращения внимания. Легко провести аналогию между работой этого механизма и процессом выборки из памяти в обычной вычислительной машине. При этом сознанию человека можно сопоставить арифметическое устройство этой машины и считать, что понятия, на которые обращено внимание, обрабатываются механизмом сознания человека, подобно тому как содержимое ячеек памяти машины, будучи помещенным в регистры арифметического устройства, обрабатывает-

ся этим устройством. Отдельные эвристики, осуществляющие преобразования понятий и приписывание значений *важности*, можно рассматривать как конкретные операторы механизма сознания, которые начинают работать при выполнении соответствующих условий.

Понятия, на которые обращается внимание (попадающие в поле внимания), могут поступать либо из памяти человека, либо, может быть, непосредственно от его органов чувств. При этом память человека и его механизм вспоминания и обращения внимания устроены таким образом, что позволяют ему очень быстро находить в ней многие необходимые в данный момент сведения. В частности, человек, как правило, быстро вспоминает ситуацию, обладающую какими-то определенными свойствами, быстро находит законы, которыми можно воспользоваться при решении данной задачи, быстро вспоминает способ решения задачи, если задача решалась ранее. На этом основании иногда высказывается мнение (см., например, [7]), что память человека образует некоторую иерархическую структуру деревообразного или какого-либо другого типа, которая позволяет ему, производя последовательные проверки в узлах ветвления, не рассматривать бесперспективные пути и за счет этого относительно быстро находить нужные сведения. Нам, однако, более вероятной представляется другая гипотеза об устройстве памяти, которая использует идею ассоциативной памяти.

Поскольку закономерности человек находит в виде утверждений (высказываний), то естественно предположить, что высказывания являются одним из основных типов элементов памяти. При этом все понятия, содержащиеся в высказывании, можно считать естественным образом связанными с этим высказыванием отношением включения записи в запись, и можно предположить, что переключение внимания происходит следующим образом. Если внимание обращено на некоторое высказывание, то последнее с помощью операторов (эвристик) механизма сознания как-то перерабатывается в другие (например, составляющие его) понятия. По этим понятиям из памяти выбирается некоторое связанное с ними (содержащее их) высказывание, на которое и переключается внимание. Заметим, однако, что таких высказываний может быть очень много. Поэтому возникает вопрос — на какое же из них должно быть переключено внимание.

Вспомним прежде всего, что различным понятиям могут быть приписаны различные значения *важности*. Тогда естественно считать, что решение вопроса о том, на какое из высказываний должно быть обращено внимание, определяется не только их связями с рассматриваемыми понятиями, но и *важностями* самих высказываний. Кроме того, заметим, что существует «настройка на контекст», благодаря которой выбор понятий, на которые переключается внимание, зависит не только от того, с чем работает сознание в настоящий момент, но и от того, о чем недавно шла

речь. В общем случае, очевидно, само понятие *контекст* может относиться не только к информации, возникающей в процессе разговора, но и к любым другим понятиям, на которые недавно обращалось внимание. Так, например, если нам задают вопрос: «Давно ли вы видели Петра?», то возникающий в нашем сознании образ Петра неизбежно оказывается связанным и с образом того человека, который задает нам вопрос, и с тем местом, где этот вопрос задается, и с тем, о чем мы думали незадолго до этого. Нам может быть известно несколько лиц с этим именем, однако прежде всего у нас обычно возникает образ Петра, знакомого не только нам, но и задающему вопрос. Следует заметить, что процесс поиска именно этого образа проходит, как правило, вне нашего сознания — в нем обычно не возникают образы других лиц с тем же именем. Это делает маловероятной возможность перебора образов всех таких лиц. Более правдоподобным нам представляется непосредственный выбор из памяти высказывания, устанавливающего сам факт знакомства Петра с лицом, задающим вопрос, т. е. одновременно связанного и с понятием (образом) «Петр» и с понятием «лицо, задающее нам вопрос». Отыскав это высказывание и разбив его в своем сознании на отдельные понятия, мы, очевидно, можем найти среди них образ именно того Петра, который одновременно знаком и нам, и тому, кто задает вопрос. Если это условие удовлетворяется для нескольких образов, то в нашем сознании возникает тот из них, который больше соответствует данной обстановке, например, образ того Петра, с которым мы вместе с лицом, задающим вопрос, встречались на том месте, где этот вопрос задается, и т. д. Поэтому можно предположить, что из памяти выбирается высказывание, связанное одновременно и с образом Петра, и с лицом, задающим вопрос, и с местом, где этот вопрос задается.

Таким образом, «лицо, задающее вопрос» и «место, где задается вопрос», служат для нас понятиями, как бы образующими контекст. Следует подчеркнуть, что на эти понятия мы можем в момент получения вопроса уже не обращать своего внимания. В то же время, они совсем еще недавно были в нашем сознании и сейчас находятся где-то «рядом» (имеют большую важность), так что внимание может быть на них легко обращено. Заметим, что эти понятия существенно отличаются по своим функциям от высказываний, которые мы вспоминаем, так как используются для установления связей с этими высказываниями. Поэтому естественно считать, что такие понятия лежат в специальной памяти, которую мы будем в дальнейшем называть *контекстной памятью* (или просто *контекстом*) в отличие от *основной памяти*, где лежат вспоминаемые высказывания.

Очевидно, что не все понятия, лежащие в контекстной памяти, равноправны. Есть более важные и менее важные. Так, например, «лицо, задающее вопрос» является, как правило, более важным понятием, чем «место, где задается вопрос» (на первое легче об-

ратить внимание), хотя возможно, вообще говоря, и обратное. Эта *важность* сказывается, по-видимому, на «весе» связей этих понятий с высказываниями из основной памяти (чем больше *важность* понятия, тем больше «вес» его связи). Таким образом, можно считать, что вероятность выбора высказывания из основной памяти определяется *важностью* этого высказывания и общим весом его связей с понятиями из контекстной памяти.

Смысл термина «связь» следует, по-видимому, несколько расширить, считая, что связанными могут быть не только высказывания и используемые в них понятия, но и любые другие пары понятий, такие, что запись одного содержит в себе запись другого (слова и куски этих слов; картинки и их элементы и пр.). Тогда наше представление об устройстве памяти можно резюмировать следующим образом.

1. Имеется два типа памяти — основная и контекстная. В основную (долговременную) память поступают понятия только из контекстной памяти. В контекстную (кратковременную) память поступает вся информация от органов чувств, а также все результаты работы механизма сознания (в частности, высказывания о том, каким образом и в каком порядке понятия, на которые обращалось внимание, обрабатывались в соответствующей ситуации). В поле внимания (входной регистр механизма сознания) могут поступать как понятия из контекстной памяти, так и из основной памяти.

2. Всем понятиям, находящимся в основной и контекстной памяти, всегда приписаны некоторые значения *важности* (определяющие вероятности обращения внимания на эти понятия без учета связей с контекстом). Информации, поступающей от органов чувств, *важность* приписывается с помощью специальных выходных операторов. Другим понятиям контекстной памяти *важности* приписываются с помощью операторов механизма сознания. Понятия, поступающие в контекстную память, сразу же поступают и в основную память, но с очень малой *важностью*. По мере того, как со временем *важности* этих высказываний в контекстной памяти падают (см. эвристику 1.1), автоматически растут их *важности* в основной памяти и в итоге принимают значения, пропорциональные (с коэффициентом, меньшим единицы) начальным *важностям* в контекстной памяти.

3. Вероятность обращения внимания на понятие из контекстной памяти определяется лишь *важностью* этого понятия в этой памяти. Вероятность обращения внимания на понятие из основной памяти определяется 1) *важностью* этого понятия в этой памяти и 2) общей *важностью* связанных с ним понятий из контекстной памяти. Первая устанавливает вероятность выбора без учета связей с контекстом. Вторая увеличивает эту вероятность при наличии таких связей.

Таким образом, вероятность выбора из основной памяти понятия, связанного с понятиями контекста, как правило, оказывает

ся существенно больше вероятности выбора любого из понятий основной памяти, не имеющего этих связей. Вследствие этого с большой вероятностью внимание переключается на «нужное» понятие. В рамках такой модели можно объяснить и многие случаи «срыва» внимания, когда внимание обращается на понятие, не связанное (или слабо связанное) с понятиями контекста.

10. Возможный механизм реализации памяти. При реализации на обычной вычислительной машине предложенной здесь гипотезы об устройстве памяти выбор нужного элемента памяти будет требовать полного ее просмотра. Поиск по иерархической структуре был бы существенно более эффективным. Отсюда, однако, не следует, что наша гипотеза не верна, поскольку у человека устройство памяти может быть реализовано совсем не так, как на обычной вычислительной машине. В частности, более приспособленной для реализации этого устройства, с нашей точки зрения, была бы молекулярная вычислительная машина [8, 9]. Элементами памяти такой машины являются слова-молекулы, например ДНК, РНК или белков. В нашем случае эти молекулы могут служить записями понятий, в частности высказываний. Поиск элементов памяти ведется за счет теплового движения молекул, а выбор «нужного» элемента памяти происходит за счет комплементарности участков поверхностей одних молекул поверхностями других. В результате к молекулам, соответствующим полю внимания или понятиям контекстной памяти, прилипают лишь те из молекул, соответствующих понятиям основной памяти, участки поверхностей которых (например, отдельные понятия, входящие в высказывание) комплементарны поверхностям молекул, составляющих поле внимания или контекст. Таким образом, связь между понятиями устанавливается с помощью комплементарности поверхностей молекул, соответствующих этим понятиям.

Заметим, что в рамках такой модели важность понятий естественным образом задается в виде концентрации соответствующих этим понятиям типов молекул. Чем выше концентрация молекул данного типа, тем выше вероятность столкновения с одной из молекул этого типа. Это в равной мере относится к молекулам как основной, так и контекстной памяти, с той лишь разницей, что молекулы контекстной памяти естественно считать привязанными к определенному месту, с которого они легко могут передать «захваченные» ими молекулы из основной памяти в поле внимания. Оттуда понятия поступают в механизм сознания, который, как можно предположить, реализуется не на молекулярном, а на нейронном уровне.

11. Разрешение утверждений. Как уже говорилось, при замене поиска алгоритма, позволяющего строить правильные ответы, поиском системы утверждений (см. раздел 6) появляется необходимость в некотором механизме, способном разрешать эту систему, т. е. строить такие значения неизвестных, которые бы удовлетворяли каждому из входящих в нее утверждений. Простейший

и, казалось бы, универсальный способ такого разрешения состоит в следующем. Для набора неизвестных, входящих в данную систему утверждений, будем последовательно синтезировать всевозможные наборы слов в алфавите, представляющем собой объединение алфавитов, соответствующих всем известным словам из этой системы, и для каждого из синтезированных наборов будем проверять, удовлетворяет ли он этой системе. Если система имеет хотя бы одно решение, то оно будет, очевидно, таким способом рано или поздно найдено. Тем не менее, в чистом виде такой способ оказывается практически неприменимым. Основная причина заключается в том, что в рамках такого способа никогда не будет известно, все ли возможные решения системы уже найдены. Поэтому, в частности, поиск решения для противоречивой (не имеющей ни одного решения) системы окажется бесконечным. Кроме того, такой способ чрезвычайно малоэффективен. Даже для относительно простых систем утверждений поиск решения будет вестись очень долго. Поэтому используемый человеком механизм разрешения основывается не на слепом подборе подходящих решений, а на применении логического вывода. С помощью такого вывода система утверждений постепенно преобразуется в алгоритм (в общем случае алгоритм перебора), который по значениям входящих в эту систему известных величин строит ее решения, т. е. наборы удовлетворяющих этой системе значений неизвестных. Процесс разрешения систем утверждений может быть, таким образом, условно разбит на следующие два этапа: 1) синтез по имеющейся системе утверждений алгоритма, строящего все решения этой системы, и 2) реализацию этого алгоритма, т. е. построение самих решений. Поскольку выполнение второго этапа естественно считать функцией языкового интерпретатора (см. раздел 8), то нас в основном будет интересовать лишь этап синтеза (вывод) алгоритма. Механизм, осуществляющий этот этап, мы называем *механизмом вывода* или просто *выводом*.

Мы не будем здесь описывать устройство механизма вывода — этому будет посвящена отдельная статья. Остановимся лишь на некоторых проблемах, которые возникают на пути его создания. В простейшем случае, когда система утверждений имеет единственное решение, функцией этого механизма является преобразование такой системы в алгоритм построения этого единственного решения. Однако при попытке создания универсального и эффективно работающего механизма вывода даже для такого частного случая приходится столкнуться с серьезными трудностями. В частности, механизм вывода, работающий с утверждениями, записанными на языке, введенном в разделе 7, пока достаточно эффективен лишь в тех случаях, когда, во-первых, программы, соответствующие этим утверждениям, не имеют циклов, и, во-вторых, алгоритмы, строящие решение, также могут быть записаны без циклов. Для циклических же утверждений (даже достаточно простых) этот вывод часто строит очень громоздкие и

малоэффективные (долго работающие и требующие большой памяти) алгоритмы. Низкая эффективность является, по-видимому, принципиальным недостатком необучаемого, т. е. использующего фиксированный набор правил, механизма вывода. Единственный путь преодоления этой трудности лежит, с нашей точки зрения, в том, чтобы искать закономерности в самом процессе вывода и на этой основе строить новые, более эффективные правила вывода, применимые в конкретных условиях и выбираемые из памяти с учетом важности этих условий.

Напомним, что в тех случаях, когда имеющаяся система утверждений противоречива, механизм вывода должен иметь возможность устанавливать этот факт, чтобы не предпринимать поиск несуществующих решений. В силу известной теоремы Гёделя, решение такой задачи при фиксированном наборе аксиом и правил вывода оказывается не всегда возможным. Однако обучаемость механизма вывода позволяет преодолеть и эту трудность. Если закономерности искать и в самих выражениях языка описания закономерностей, то аксиоматика механизма вывода будет постоянно пополняться, расширяя тем самым класс разрешимых утверждений.

12. О представлении задач. Следует заметить, что человек пользуется выводом не только для разрешения найденных им утверждений, но и в процессе самого поиска этих утверждений, получая с помощью вывода рекомендации относительно проверки тех или иных конкретных высказываний или выполнения каких-либо преобразований. Так с помощью вывода выясняется тавтологичность утверждений или зависимость новых утверждений от уже имеющихся. Иногда, пользуясь выводом, удается выяснить, каких именно дополнительных данных было бы достаточно, чтобы система утверждений имела единственное решение (например, однозначно определяла правильный ответ). Таким образом, в механизме сознания человека процесс вывода по существу оказывается неотделимым от процесса поиска закономерностей, и естественно считать, что реализуется он с использованием того же самого механизма обращения внимания и тех же самых эвристик. При этом легко встать на точку зрения, когда любую задачу можно рассматривать как задачу вывода. Действительно, решение любой задачи можно представить как поиск значения некоторой неизвестной величины, связанной системой утверждений с известными понятиями. Такой неизвестной может быть, очевидно, не только слово (например, ответ), но свойство, алгоритм или закономерность (в задаче поиска закономерностей). Эта точка зрения особенно привлекательна тем, что позволяет рассматривать механизм вывода как единственный механизм решения любых задач, однако при этом требуется существенная перестройка внутреннего языка описания закономерностей.

Литература

1. *А. Тьюринг.* Может ли машина мыслить? Физматгиз, 1960.
2. *М. М. Бонгард, М. Н. Вайнцвайг.* Об оценках ожидаемого качества признаков.— Проблемы кибернетики, 1968, № 20.
3. *Н. Нильсон.* Искусственный интеллект. М., «Мир», 1973.
4. *Дж. Слейгл.* Искусственный интеллект. М., «Мир», 1973.
5. *Р. Л. Гудстейн.* Математическая логика. М., ИЛ, 1961.
6. *А. А. Марков.* Теория алгоритмов. М., Изд-во АН СССР, 1954.
7. *М. М. Бонгард, И. С. Лосев, М. С. Смирнов.* Проект модели организации поведения —«Животное».— Наст. сб.
8. *Е. А. Либерман.* Молекулярная вычислительная машина клетки.— Биофизика, 1972, т. 18, № 5.
9. *М. Н. Вайнцвайг, Е. А. Либерман.* Молекулярная вычислительная машина.— Биофизика, 1973, т. 18, № 5.

ОПЕЧАТКИ

Стр.	Строка	Напечатано	Следует читать
138	9 сн.	$ГГ$	$ГТ$
150	2 св.	$\Delta_{i,\ i'}$	$[\Delta_{i\ i'}$
150	5 св.	$\mathbf{F}_p$	$\mathbf{F}$
225	10 св.	xc	$\overleftarrow{xc}$
228	3 св.	ответ u	ответ y