

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

МЕЖДИСЦИПЛИНАРНЫЙ  
ПОДХОД



Москва  
2006

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

МЕЖДИСЦИПЛИНАРНЫЙ  
ПОДХОД

ИИНТЕЛЛ

Москва

2006

УДК 100.32

ББК 32.816

И 86

Под редакцией д.ф.н., проф. Д.И. Дубровского и  
члена-корреспондента РАН, д.ф.н. В.А. Лекторского

И 86 **Искусственный интеллект: междисциплинарный подход.**  
**Под ред. Д.И. Дубровского и В.А. Лекторского – М.:**  
**ИИнтелЛ, 2006. – 448 с.**

Книга представляет собой один из первых в отечественной научной литературе опытов междисциплинарного подхода к проблематике искусственного интеллекта. В ней рассматриваются философские, методологические, общетеоретические и социокультурные аспекты данной проблематики, обсуждаются актуальные задачи моделирования искусственного интеллекта в связи с рядом логических и математических вопросов и под углом соотношения искусственного интеллекта с естественным и современных разработок проблемы «сознание и мозг». Авторы статей – философы, психологи, специалисты в области компьютерных наук, логики, математики, биологии, нейрофизиологии, лингвистики.

*Издание осуществлено при финансовой поддержке  
Российского гуманитарного научного фонда (РГНФ)  
проект № 05-03-16056 д*

ISBN 5-98956-005-2

© ООО «ИИнтелЛ», 2006 г.

---

## ОТ РЕДАКТОРОВ

Развитие информационного общества выдвигает на передний край науки проблему искусственного интеллекта (ИИ), придает задачам развития новых интеллектуальных информационных технологий первостепенное теоретическое и практическое значение. Успехи в этой области способны существенно повлиять на развитие экономики страны, о чем неоднократно говорил в последнее время Президент России. Основательная разработка методологических и теоретических аспектов проблематики ИИ – необходимое условие достижения высокого уровня научных исследований в этом направлении и создания таких интеллектуальных информационных технологий, которые отвечают насущным потребностям нашей страны, способны обеспечить ей достойное место на мировом уровне.

Эта существенная роль методологической работы ясно осознается многими специалистами в области информационных технологий, что продемонстрировала проведенная в январе прошлого года Всероссийская конференция «Философия искусственного интеллекта». В ней приняло активное участие значительное число представителей компьютерных наук, робототехники, информатики, сотрудников соответствующих Институтов РАН, а так же специалистов высших учебных заведений и различных научных учреждений из 44 регионов России. На основе рекомендаций этой конференции был создан Научный Совет РАН по методологии ИИ, который возглавил известный отечественный математик, академик В. Л. Макаров (член Президиума РАН, Академик-Секретарь Отделения общественных наук РАН, директор Центрального экономико-математического института РАН). Задачи Совета: содействие интегративным процессам в области существующих направлений научных исследований и технической деятельности в области ИИ, организация и повышение эффективности междисциплинарного диалога по проблемам ИИ и его соотношений с естественным интеллектом; качественное методологическое обеспечение разработок проблематики ИИ и улучшение методологической подготовки специалистов в системе высшего образования. Стратегической задачей Совета является всемерное содействие развитию новых направлений интеллектуальных информационных технологий в нашей стране.

За истекший год Советом проведен ряд значительных мероприятий (созданы Региональные отделения Совета в крупных научных центрах страны, секции и рабочие группы Совета, секция по проблематике искусственного интеллекта в Экспертном Совете по инновациям и интеллектуальной собственности Государственной Думы РФ и др.). Систематически проводятся заседания ежемесячного теоретического семинара Совета.

---

На нынешнем этапе вопросы междисциплинарного сотрудничества составляют одну из наиболее актуальных задач в области разработок проблем ИИ. Здесь как раз и возникают наибольшие методологические трудности. Они связаны с необходимостью теоретически корректного соотнесения языков различных, часто далеко отстоящих друг от друга научных дисциплин, с выработкой таких концептуальных структур, которые бы позволили создавать продуктивные способы междисциплинарного общения и взаимопонимания. В преодолении этих трудностей пока еще не видно серьезных достижений.

Но прежде всего следует, конечно, ясно обозначить основной круг вопросов, предъявляемых к проблематике ИИ теми научными дисциплинами, которые вовлечены в разработку этой проблематики. Одно дело вопросы логики и математики, другое – вопросы психологии с ее результатами изучения естественного интеллекта или вопросы тех комплексных научных дисциплин, которые исследуют информационные процессы в головном мозгу и в живых системах. Но есть еще лингвистика, от которой во многом зависят успехи моделирования естественного языка и которая предъявляет свои вопросы и свои требования к проблематике ИИ; не говоря уже о других дисциплинах, имеющих свои интересы в этой области.

Иными словами, на первом этапе необходимо выслушать участников междисциплинарного диалога, их вопросы, сомнения, предложения, касающиеся перспектив разработки проблем ИИ. Важно выделить и поддержать те локальные интегративные процессы, которые наблюдаются в отдельных областях многомерной проблематики естественного интеллекта и дали уже положительные результаты.

Именно такого рода вопросам посвящена предлагаемая книга. Она представляет собой одну из первых попыток систематизированного в первом приближении междисциплинарного подхода к проблематике ИИ. Большинство статей, составляющих книгу, это специально доработанные авторами тексты их докладов и выступлений на теоретических семинарах Совета и на Круглом столе «Философско-методологические проблемы когнитивных и компьютерных наук», проводившемся в течение двух дней в рамках IV Российского Философского Конгресса (май 2005 г.).

Естественно, в ряде случаев те или иные соображения авторов носят остро дискуссионный характер, нуждаются в дополнительном обосновании. Однако в целом, по нашему убеждению, книга дает достаточно репрезентативную панораму нынешнего состояния междисциплинарного подхода к разработке проблематики ИИ, и она способна служить стимулом для дальнейших методологических и теоретических исследований в этой области.

*Д.И. Дубровский, В.А. Лекторский*

---

# МЕТОДОЛОГИЧЕСКИЕ И ЭПИСТЕМОЛОГИЧЕСКИЕ ВОПРОСЫ

---

---

## ПОЛУЧЕНИЕ НОВОГО ЗНАНИЯ МЕТОДОМ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

*В.Л. Макаров*

Со времен Френсиса Бэкона, а, может быть, и раньше ученые осознали, что новое знание получается методом «чтения книги жизни». Однако вопрос о том, как же конкретно ее читать, остается.

Приборы для чтения книги жизни многообразны и подвержены классификации подобно наукам. Физические приборы, включая миллиардной стоимости синхрофазатроны, – это один тип приборов. Им подобны астрономические, химические, биологические приборы (телескопы, микроскопы, котлы, камеры и пр.).

Другой тип приборов используется, в основном, общественными науками. Это различные способы измерения происходящего в обществе: опросы, статистические конторы, СМИ, институты голосования, ИНТЕРНЕТ-обработка и др.

При использовании приборов чтения книги жизни возник феномен вмешательства в явление в процессе его измерения. Это принцип неопределенности Гейзенберга в физике, влияние на мнение населения при проведении и публикации опросов и др.

В настоящей статье я обсуждаю проблему получения нового знания с помощью прибора, который изобретен сравнительно недавно, а именно, с помощью компьютерного моделирования.

Компьютерная модель объекта, явления, процесса может быть разной степени сложности и разной степени адекватности. Возможна ли точная копия объекта или нет? Что говорит мысленный эксперимент? Какие парадоксы здесь возникают? Другое понятие машины времени. Возможные миры и реализуемые миры. Я пытаюсь обсудить эти вопросы в теоретическом плане.

Далее я привожу несколько примеров действующих компьютерных моделей и результатов экспериментов с ними. На этих примерах обсуждается вопрос, можно ли считать полученное в эксперименте новым знанием, знанием относительно реальности, а не только относительно того, что называется виртуальной реальностью.

Чем отличается виртуальная, в том числе компьютерная реальность, от «настоящей» реальности. И есть ли «настоящая» реальность?

Понятно, что, как всегда, вопросов больше чем ответов.

### **Типы реальности**

Первичная реальность. Это нечто, что познается человеком непосредственно, с помощью его органов чувств. То – есть человек с ней, с этой реальностью, соприкасается непосредственно. Для человека 20 века автомобили, самолеты, радио, телефон, телевидение являются первичной реальностью, подобно тому, что для человека 19 века таковой реальностью являются лошади, телеги, мушкетеры и т. д.

Вторичная реальность появляется из рассказов о первичной реальности другими людьми. То – есть первичной реальностью является сам рассказ, а внутри рассказа есть другая реальность, которую человек не может воспринимать как первичную. Епископ Джозеф Беркли, основатель солипсизма, признавал только первичную реальность. Полученная информация с помощью, например, телевизора, есть информация вторичная.

Вопрос: как понимать реальность, полученную человеком не через органы чувств, данные Богом, а через вспомогательные приборы, стоящие между органами чувств и предметами реальности? Например, слепой пользуется аппаратом, заменяющим ему зрение. Астроном пользуется телескопом, биолог – микроскопом. Снимки обратной стороны Луны или Марса получаются с использованием целой совокупности приборов.

Вроде бы современный человек смирился с тем, что получаемая с помощью приборов информация также первична. Ибо информация, полученная ощупыванием рукой или палкой (то – есть прибором), так сказать, одного порядка.

### **Работа мысли**

Получив информацию о реальности, человек формирует представление о мире с помощью мыслительных процедур. В частности, с помощью мыслительных процедур устанавливаются причинно – следственные связи. Все мы знаем, что мыслительные процедуры также можно усиливать с помощью приборов. Усиление мыслительного процесса с помощью приборов также имело и имеет своих скептиков, как и получение первичной информации, о которой речь шла выше. Любой калькулятор является таким прибором. Я лично имел возможность встречаться с таким скептиком. В семидесятые годы 20

века в Сибирском Отделении Академии Наук работал бухгалтер высочайшей квалификации, который доверял счётам (прибору), но не доверял электронным калькуляторам (другому прибору), получившим распространение в то время.

Появившиеся во второй половине 20 века компьютеры почти сразу стали рассматриваться как усилители мыслительных процедур. Проведение с помощью компьютеров математических выкладок, логических заключений привело к тому, что сложные и утомительные математические доказательства стало возможным поручать компьютеру. Опять появились неверующие: считать или не считать теорему доказанной, если часть работы проделал компьютер.

Главная цель мыслительных процедур: из фактов главным образом первичной, но не только, реальности создавать картины мира и его частей, создавать теории. Эти картины, эти теории помогают ориентироваться в жизни. Именно они создают у человека ощущение, что он *понимает*, что происходит. Субъективное ощущение понимания возникает именно тогда, когда в голове построена теория, построена модель. И когда появляющиеся новые факты укладываются в эту теорию, в эту модель.

Ясно, что мыслительный процесс неоднозначен. На одних и тех же фактах можно строить, и реально построены разные теории. Отсюда бытовые и научные споры.

### **Компьютерные модели**

Создание моделей реальных объектов является едва ли не главным в процессе познания. И модели создавались всегда, коль они неотделимы от процесса познания. Модели разные по своему инструментальному содержанию. Лев Толстой создал художественную модель войны 1812 года. Брюллов – модель гибели Помпеи. Историк, излагая исторические факты, вольно или невольно придерживается какой то модели, иначе изложение не будет понятным. Модель царствования Петра I у славянофила и у западника будут принципиально различаться, хотя и используют одни и те же факты.

Модели создаются не только для познания, но и для облегчения творчества. Особенно любят модели архитекторы. Прежде чем возводить дворец, они делают его модель. Распространены модели кораблей, самолетов, городов.

Основной тезис моего выступления состоит в том, что компьютеры совершили переворот в процессе создания и использования моделей. Модели получили новое качество, и мы пока не осмыслили всех последствий этого нового качества для процесса познания.

Возможно, что здесь появляются новые логические и даже физические парадоксы.



## **Компьютерная машина времени**

Чтобы представить себе, о чем идет речь, рассмотрим мысленный эксперимент. Построим компьютерную модель какого-нибудь реального объекта, например, модель Советского общества 80 – х годов 20 века. Как всякая модель, она является лишь некоторым приближением действительности. Однако модельеры знают, что любую модель можно «совершенствовать», делать ее все более точной. Например, если в исходной версии набор действий, осуществляемых (компьютерным) человеком состоял из 10 позиций, что в следующей версии он может состоять из 100 позиций и т. д. То – есть отражение действительности в модели становится все более точным.

Теперь спросим себя, где находится предел для описанной последовательности моделей, в которой каждая последующая модель более точно отражает действительность. Пределом является *точная копия действительности*. Если бы предел был достижим, что мы бы имели не что иное, как машину времени в предположении, что компьютерное время быстрее реального. Причем компьютерная модель устроена так, что двигаться во времени можно в двух направлениях, и вперед, и назад.

Специалист по математической логике сразу скажет, что предел недостижим хотя бы потому, что возникает логический парадокс. А именно, собственная часть оказывается равной целому. Компьютерная модель действительности в точности равна самой действительности, собственной частью которой является эта модель.

## **Экспериментальное знание против математического**

Традиционно считается, что математическое знание имеет, так сказать, высший рейтинг. Это знание высокого качества. Оно неленно, оно навсегда, оно не подлежит пересмотру.

Ясно, однако, что математический способ получения нового знания весьма ограничен. Он, как известно, состоит в следующем. Формулируется математическая модель объекта, процесса, явления в виде набора исходных предположений (аксиом). А далее доказываются утверждения относительно свойств данной модели. Например, для знаменитой модели рыночной экономики Эрроу – Дорбе доказывалось существование и оптимальность рыночного равновесия.

Стремление сделать математическую модель более совершенной, более приближенной к реальности приводит к ее переусложнению. Когда математическая модель перегружена деталями, затруднительно или невозможно получить результат математическим путем. Другими словами, математическое моделирование имеет весьма низкий порог сложности, даже в предположении, что часть выкладок будет производиться компьютером.

Компьютерная модель лишена этого недостатка. Можно строить сколь угодно сложные модели, которые будут все более точно отражать действительность. Но что дальше делать с построенной моделью? Последнее время развивается методология получения нового знания с помощью вычислительных экспериментов на компьютерных моделях. И естественно возникает вопрос, насколько знание, полученное из вычислительного эксперимента, может считаться знанием относительно реальности (а не относительно искусственного объекта – компьютерной модели).

### **Вычислительные эксперименты**

Поначалу, как это обычно бывает, вычислительные эксперименты рассматривались как забавные игрушки, как нечто, не очень серьезное. Например, эксперименты с конечными автоматами, в частности с клетками Фон Неймана, см. Цейтлин [2]. Сюда же относятся разного рода модели эволюции.

Модель эволюции представляет собой заданную в начальный момент популяцию (искусственных) существ, правила их рождения и смерти, правила взаимодействия между собой. При этом некоторые операции могут носить вероятностный характер. Далее запускается вычислительный процесс, имитирующий эволюцию, и смотрится, к чему он, в конечном счете, приведет. Интрига в том, что заранее трудно или невозможно предсказать результат. Получить математически результат эволюции тоже, как правило, чрезвычайно трудно, если возможно вообще. Меняя те или иные правила можно получать, естественно, разные результаты эволюционного процесса. С рафинированной научной точки зрения результат вычислительного эксперимента нельзя рассматривать как нечто, достоверно полученное. Ибо он зависит от конкретных чисел. Возьмем другие числа, и результат может получиться иным. Ибо это просто одна из случайных реализаций. Такая критика легко снимается правильной методикой обработки экспериментальных данных. Но вот сомнения относительно ценности результата по отношению к действительности, а не искусственно построенной действительности остаются.

Наверное, здесь должна убеждать практика. Хороший пример – компьютерная имитация ядерных взрывов. Ядерные взрывы в природе запрещены, дорогостоящи, загрязняют окружающую среду. Жизнь показала, что компьютерные ядерные взрывы дают ответы на некоторые вопросы по совершенствованию данного оружия.

Многие, вероятно, помнят открытое методом компьютерного моделирования явление, названное «ядерной зимой». Можно сомневаться или не сомневаться в достоверности результата, полученного таким способом, но согласитесь, что в свое время он произвел впечатление на научное сообщество, да и не только на него.

В качестве убедительного примера, для меня по крайней мере, «результата из компьютера» укажу на быстро развиваемое в настоящее время направление инкорпорирования политических процессов в экономические теории. Это, так называемая, новая политическая экономия. Математические модели здесь мало – что дают из-за вынужденных упрощений. А результаты компьютерных калькуляций вполне заслуживают внимания. Мы со своим коллегой Данковым взяли компьютерную модель голосования, предложенную в Kollman, Ken, John H. Miller and Scott E. Page (1997) [3] и провели на ней ряд вычислительных экспериментов. См. А.Н. Данков, В.Л. Макаров. (2002) [1]. В частности мы выяснили, что политическая партия, склонная к большей гибкости, имеет и большие шансы на победу. Это, в общем – то печально, ибо побеждает беспринципность, но убедительно с точки зрения эксперимента. В данной модели платформа партии представляет собой булев вектор, где единица показывает, что партия «за» данную альтернативу (монетизация льгот, война в Чечне и пр.), а ноль – соответственно «против». Граждане голосуют несколько туров и партии имеют право корректировать свои платформы в зависимости от результатов предыдущих туров. Корректировка разрешается в окрестности платформы. Так вот, чем шире эта окрестность, тем больше шансов на выигрыш. В жизни, как известно, это так и не так. «Единая Россия», проявляя гибкость, (или беспринципность, кому как нравится) выигрывает, а СПС, наоборот, проиграла.

### **Искусственные миры**

Итак, наука идет в направлении создания искусственной, виртуальной, компьютерной действительности. Называйте, как хотите. Но эта рукотворная реальность есть некая другая картина мира. Не та, которая сидит в наших головах или описана в книгах, а именно другая, чего раньше не было в арсенале человечества. Это принципиально иной прибор для познания действительности, для добывания нового знания.

Правда, нельзя сказать, что это совершенно новый, неожиданный скачок в инструментах познания реальности. Промежуточным шагом можно считать всем хорошо известные игры. Игры, которые были всегда, игры, в которые играют и животные. Игры имитируют жизненные ситуации. Слово «имитируют» означает, что в игре речь идет о модели реальности. Например, в футболе обучаются ловкости, в бизнес – играх обучаются бухгалтерскому, брокерскому делу управлением компанией и т. д. Распространение компьютерных игр среди детей свидетельствует о том, что виртуальная реальность специально делается симпатичнее настоящей реальности.

Всё большее распространение получают, так называемые, ситуационные комнаты. Там играют важные дяди, облеченные большой ответ-

ственностью. Первая ситуационная комната, как известно, была сделана по приказу Мак-Намары, министра обороны США в то время. В ней разыгрывались варианты военных действий во Вьетнаме в зависимости от тех или иных решений командиров. «Что будет, если...» (What – If) анализ, инструментарий которого взят на вооружение всеми развитыми странами, представляет собой не что иное, как создание и использование искусственных миров. Понятно, что чем ближе искусственный мир к реальному, тем лучше. Тем достовернее предсказания, тем реалистичнее прогнозы.

Уже наклёвывается естественное разделение труда. Одни создают модели искусственных миров, а другие проводят с ними эксперименты.

Не за горами то время, когда инструмент искусственного мира работает на полную мощь и станет доминирующим способом получения новых знаний в общественных науках. Каковы последствия существования электронного зеркала действительности, вопрос, на который пока нет определенного ответа.

#### ЛИТЕРАТУРА

1. *А.Н. Данков, В.Л. Макаров.* (2002) «Межтерриториальная и межпартийная конкуренция: сравнительный анализ влияния политических институтов». Препринт Российской экономической школы.
2. *Цетлин М.Л.* Исследования по теории автоматов и моделированию биологических систем. М., Наука, 1969.
3. *Kollman, Ken, John H. Miller and Scott E. Page,* (1997), Political Institutions and Sorting in a Tiebout Model. *American Economic Review.* 87: 977-992.
4. Computational Economics Editor-in-Chief: H.M. Amman ISSN: 0927-7099 (print version) Journal no. 10614 Springer US
5. The Journal of Artificial Societies and Social Simulation. JASSS, ISSN 1460-7425, V. 8, Issue 2
6. Программирование искусственного интеллекта в приложениях / *М. Тим Джонс;* Пер. с англ. Осипов А.И. – М.: ДМК Пресс, 2004. – 312 с.: ил.

---

# ФИЛОСОФИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНАЯ НАУКА

*В.А.Лекторский*

Познание всегда было одной из основных тем философских исследований. Было время, когда эта проблематика вообще стала центральной в философии (после так называемой «эпистемологической революции» в 17 столетии) и даже начала претендовать на единственность – ряд философских направлений конца 19 и начала 20 веков: махизм, неокантианство и др. Широко распространилось мнение о том, что познание во всяком случае остаётся прерогативой философского изучения (среди отечественных философов в 50-ые и 60-ые гг. прошлого столетия была популярна идея о том, что теория познания и есть предмет философии).

Однако это мнение стало выглядеть не столь непререкаемым уже со второй половины 19 века, ибо не только философия стала претендовать на изучение познания. В это время возникает экспериментальная психология, которая начинает научными методами исследовать процессы чувственного познания (ощущение, восприятие). В 20 веке психологи стали изучать процессы мышления: вюрцбургская школа, гештальт-психология, работы Ж. Пиаже. К середине 20 века возникают профессиональные история науки, социология науки. В 60-ые и 70-ые гг. прошлого столетия происходит так называемая «когнитивная революция» в психологии и возникает когнитивная психология. Особенность последней не в том, что психология впервые начинает изучать познавательные процессы (она делала это и раньше в рамках тех направлений, которые я только что упоминал). Специфичность когнитивной психологии в другом: во-первых, она понимает когнитивные, т.е. познавательные процессы как лежащие в основе всех психических функций (в том числе эмоций, мотивации, волевых проявлений), а, во-вторых, в том, что познание начинает истолковываться как процесс переработки информации, аналогичный тому, который имеет место в компьютере. На возникновение и развитие когнитивной психологии серьёзно повлияло появление в 60-ые гг. 20 века и последующее бурное развитие направления исследований, которое получило название искусственного интеллекта.

Исследователи в области искусственного интеллекта (во многом стимулированные идеями А.Тьюринга<sup>1</sup>) занялись созданием таких

---

<sup>1</sup> См. знаменитую статью *А.Тьюринга* “Computing Machinery and Intelligence”, *Mind*, vol. LIX, No. 236, 1950

программ, работая по которым, компьютер мог бы решать те задачи (доказательство теорем, планирование операций, игра в шахматы и др.), которые человек осуществляет с помощью своего естественного интеллекта. Важно при этом заметить, что речь идёт не о простой имитации компьютером интеллектуальных операций человека. Компьютер может решать ту или иную интеллектуальную задачу иными способами, чем это делает человек, при этом во многих случаях гораздо эффективнее. Поэтому, если перед разработчиками в области искусственного интеллекта стоят чисто практические задачи – решение компьютером определённого типа проблем, создание роботов, выполняющих некоторые действия за человека, в тех случаях, когда последнему было бы трудно действовать в тех или иных условиях, – тогда бессмысленно претендовать на то, что работа в области искусственного интеллекта может дать что-то новое для понимания человека и его познания.

Однако если придерживаться идеи (которую разделяет когнитивная психология), что процессы переработки информации человеческим мозгом принципиально схожи с процессами, имеющими место в компьютере, то нужно признать, что исследования в области искусственного интеллекта могут быть важным способом понимания естественного интеллекта, да и вообще всех познавательных процессов человека. Только в этом случае приходится считаться с тем, что на процессы переработки информации у человека существенные ограничения накладывают биологически обусловленные (строение органов чувств и структура мозга) его особенности. Если это учесть, то можно создавать такие компьютерные модели, которые с точки зрения когнитивной психологии можно рассматривать как теоретический ключ к пониманию соответствующих психических процессов человека. Именно такого рода модель зрительного восприятия предложил, например, в начале 80-ых гг. Д.Марр.<sup>2</sup> В этой связи возникает идея о том, что будущее когнитивной психологии вообще связано с её превращением в вычислительную, или компьютерную психологию (computing psychology) и что исследования в области искусственного интеллекта могут играть роль теоретического базиса экспериментальных психологических изысканий.<sup>3</sup>

В середине 70-х гг. в США возникает междисциплинарное движение, получившее название когнитивной науки. Последняя в качестве своих частей включила исследования в области искусственного интеллекта, когнитивную психологию, когнитивную лингвистику (обязанную своим возникновением идеям выдающегося американского

---

<sup>2</sup> D. Marr. Vision. San Francisco, 1982.

<sup>3</sup> Статья Д. Деннета «Искусственный интеллект как философия и психология» (D.Dennett. Artificial Intelligence as Philosophy and as Psychology. In: D.Dennett. Brainstorms. Cambridge (M), 1986, pp. 109-126.

психолингвиста Н.Хомского, сыгравшего важную роль в «когнитивной революции») и те философские исследования, которые давали в рамках нового направления теоретическое понимание ряда важных проблем познания и сознания.<sup>4</sup> Важно при этом отметить, что объединение разных дисциплин в единую науку было связано не только с наличием единого для всех предмета изучения – познавательных процессов, – но и с тем, что участники когнитивной науки исходили из некоторых общих идей. Это по крайней мере, следующие: 1) познавательные процессы были поняты как вычислительные в широком смысле, в этой связи было предположено существование в сознании процедур, аналогичных вычислительным алгоритмам; 2) было предположено, что в сознании существуют ментальные репрезентации, аналогичные компьютерным структурам данных; к ним относятся, в частности, логические пропозиции, схемы, понятия, правила и т.д. Ментальные репрезентации являются средствами хранения информации в сознании и той основой, которая обеспечивает использование вычислительных процедур в познавательных процессах.

В рамках такого понимания когнитивной науки философия в тех её разделах, которые имели дело с исследованием познания и сознания (эпистемология и философия сознания), входит в новую науку как её часть и приобретает тем самым новый статус. Такое понимание можно было рассматривать как реализацию известной идеи крупнейшего американского философа и логика середины 20 века У.Куайна о «натурализации эпистемологии».<sup>5</sup> Куайн, правда, связывал «натурализацию эпистемологии» с превращением её в раздел психологии, но это можно объяснить тем, что в конце 60-ых гг, когда была сформулирована куайновская идея, ещё не возникла когнитивная наука, а сама «когнитивная революция» во много была впереди.

Нужно сказать, что развитие исследований в рамках когнитивной науки осуществлялось довольно интенсивно. Это выразилось, в частности, в публикации множества книг, статей, в многочисленных конференциях и конгрессах, в острых дискуссиях, в которых участвовали представители разных дисциплин, до недавних пор развивавшихся обособленно друг от друга.

Философы играли особую роль в этом движении. В самом деле. Если человеческое познание осуществляется посредством менталь-

---

<sup>4</sup> В числе этих философов были прежде всего Дж. Фодор, Д.Деннетт, М. Боден. См.: *J.Fodor. The Language of Thought. Hassoks, 1976; J.Fodor. Psychosemantics. Cambridge (M), 1986; D.Dennett. Brainstorms. Cambridge (M), 1986; D.Dennett. The Intentional Stance. Cambridge (M), 1987; M.Boden. Artificial Intelligence and Natural Man. Succex, 1976; Minds and Mechanisms. Succex1981.*

<sup>5</sup> *W.V. Quine. Epistemology Naturalised. In : W.V.Quine. Ontological Relativity and Other Essays, N-Y, 1969*

ных репрезентаций, то возникает вопрос о том, какова структура этих репрезентаций и как относятся входящие в них компоненты, с одной стороны, к вычислительным процедурам, а с другой, к переработке информации нейронами в головном мозгу. Как соотносятся принятые в философии понятия пропозиции, категории, пропозициональные установки (мнения, желания, интенции и др.) с понятиями, которыми оперируют исследователи в области искусственного интеллекта (схема, фрейм, скрипт и др.)? Как всё это относится к соответствующим вычислительным процедурам? И к процессам в мозгу на межнейронном и внутринеуронном уровне? В частности, как можно совместить процесс логического вывода (а такого рода выводы играют важную роль в познавательных процессах), подчинённого правилам и логическим нормам, с теми взаимодействиями между нейронами, которые имеют место при переработке информации головным мозгом – ведь нейроны могут действовать только причинным образом?

До недавних пор в философии считалось аксиоматичным, что причины и аргументы относятся к разным мирам. Волна критики психологизма в эпистемологии и логике в начале 20 века была связана именно с обвинениями последнего в том, что он смешивал то и другое. Но если мы пытаемся показать, как физическое взаимодействие в мозгу на нейронном уровне обеспечивает осуществление нормативного рассуждения, мы должны каким-то образом совместить эти разного типа процессы. Как соотносится переработка информации в головном мозгу, которая предполагает чисто физическое взаимодействие между нейронами (передача электронных импульсов от одного к другому) с передачей содержания информации? Какое влияние на содержание информации (т.е. на то, что для нас выступает в качестве познания и сознания) оказывает, с одной стороны, взаимодействие организма посредством органов чувств с внешним миром, а с другой, взаимодействие между нейронами, т.е. то, что имеет место внутри когнитивной системы?

Как видно, без ответа на эти вопросы нельзя понять познавательные процессы. Сами по себе модели, предлагаемые разработчиками искусственного интеллекта, эмпирические данные и теоретические идеи когнитивной психологии не дают на них ответа. Вместе с тем можно сказать, что эти вопросы – лишь новая форма старых философских проблем: каково взаимоотношение между мозгом и сознанием (психофизическая проблема), между сознанием и внешним миром, между миром сущего и миром должного и т.д. Всё дело однако в том, что новая форма старых вопросов, ставшая возможной в рамках когнитивной науки, даёт возможность не только более точной их формулировки, но и нового рода аргументации в защиту того или иного предлагаемого решения – аргументации, принимающей во внимание не только вычислительные модели из области искусствен-



ного интеллекта и факты, добытые в когнитивной психологии, но и то, насколько плодотворными оказываются те или иные философские идеи в качестве ядра конкретных исследовательских программ.

В связи с обсуждением вопросов, о которых шла речь, был предложен ряд философских идей, оказавших значительное влияние на последующее развитие когнитивного движения. Вот некоторые из них.

Прежде всего это сформулированная Х.Патнэмом концепция функционализма: содержание того или иного психического явления определяется его отношением к поведению, с одной стороны, и к другим психическим явлениям, с другой, т.е. выполняемыми им функциями.<sup>6</sup> У человека психические явления осуществляются при помощи мозга. Но содержание их от мозга не зависит. Психику можно уподобить компьютерной программе, а психические явления вычисляемым функциям. Так как содержание программы отличается от её воплощения «в железе» (т.е. в том или ином компьютере), те же самые психические функции могут выполняться (вычисляться) не только мозгом человека, но и каким-то иными их воплощениями: например, совершенно иначе устроенными телами инопланетян или же искусственно созданными роботами (тезис о «множественности реализуемости» психического). Тезис функционализма был истолкован как опровержение прямой редукции психических явлений к мозговым процессам и как обоснование независимости психологических исследований от нейрофизиологии.

Джерри Фодор предложил гипотезу о существовании языка мысли как способа осуществления мыслительных (и, возможно, других познавательных) процессов<sup>7</sup>. Мышление не сводится к оперированию символами того или иного конкретного языка, так как, во-первых, можно мыслить и помимо конкретного языка, во-вторых, одна и та же мысль может быть выражена разными разговорными языками (например, русским, английским, французским и др.). Но поскольку мышление – это осуществляемое через вычислительные процедуры оперирование символами, то приходится допустить, что символы воплощаются в ментальных репрезентациях, а оперирование ими определяется правилами закодированного в мозгу (и врождённого) универсального языка мысли.

Дэниэл Деннетт предложил особый способ понимания интенциональности (т.е. направленности на что-то), которая со времён Ф.Брентано считалась большинством философов специфической особенностью психического. Согласно Деннету интенциональность – это не столько особое состояние когнитивной системы, сколько спо-

---

<sup>6</sup> *H. Putnam. The nature of Mental States. In: Readings in Philosophy of Psychology. Ed. By N.Block. L., 1980. Vol. I, pp. 223-231.*

<sup>7</sup> *J. Fodor. The Language of Thought. Hassoks, 1976.*

соб понимания её деятельности (будет ли это человек, другое живое существо или робот) – интенциональная установка (*intentional stance*).<sup>8</sup> Мы приписываем интенциональность тем системам, которые в своих действиях обнаруживают рациональность. Если такая рациональность в действиях не обнаруживается, мы отказываем таким системам в интенциональности. В этой связи Деннетт различает физическое (или органическое) устройство системы, её функциональное назначение и возможность её понимания посредством понятия интенциональности.

Эти идеи (и ряд других<sup>9</sup>) породили большую дискуссию, которая во многом определила теоретическое содержание когнитивных исследований на первом этапе.

Но можно считать, что в 80-ые гг. наступает второй этап в развитии когнитивной науки, в ряде отношений отличный от первого как проблематикой, так и идеями. Он был связан с осознанием ряда трудностей функционалистского подхода к пониманию психики и познания и с признанием того, что предлагавшиеся вычислительные модели переработки информации являются сильно упрощёнными и не учитывают ряд важных особенностей работы мозга<sup>10</sup>. Критикуется тезис о «множественности реализуемости» психических процессов, ибо реальные (а не мнимые) психические процессы не только воплощаются в работе нервной системы человека: существенные особенности первых определяются особенностями второй. Предлагается серьёзно учитывать эти особенности и в этой связи внимательно относиться к данным нейронаук.<sup>11</sup> Последние теперь включаются в когнитивную науку как важная её часть и даже возникает новая дисциплина – нейрофилософия.<sup>12</sup>

---

<sup>8</sup> *D. Dennett. The Intentional Stance. Cambridge (M), 1987.*

<sup>9</sup> Например, идея Дж. Фодора о «методологическом солипсизме» как стратегии исследования в когнитивной психологии. Дж. Фодор считал, что для изучения психических процессов достаточно иметь дело только с способами переработки информации, а не с содержанием последней, определяемой взаимодействием организма с внешним миром, т.е. не с семантикой информации, а лишь с синтаксисом - способами вычисления. *J. Fodor. Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. In: J.Fodor. Representations. Cambridge (M), 1981.*

<sup>10</sup> Возникли и серьёзные вычислительные проблемы для разработчиков искусственного интеллекта, в частности, в связи с т.н. проблемой фрейма. См. *The Robot's Dilemma. Ed. by Z.Pylyshyn. Norwood, 1987.* Особенно интересна в этой книге статья Д.Деннетта. (*D.Dennett. Cognitive Wheels: the Frame Problem of AI. Pp. 41-64.*)

<sup>11</sup> В свою очередь когнитивная наука повлияла на исследования в области нейронаук: возникли когнитивная и компьютерная нейронауки.

<sup>12</sup> *P. Churchland. Neurophilosophy. Cambridge (M), 1986.*

Однако определяющим для развития когнитивной науки на втором этапе было прежде всего возникновение нового подхода в компьютерном моделировании процесса переработки информации человеческим мозгом, подхода, подучившего название коннекционизма, или параллельно распределённой переработки информации. Модель коннекционизма или PDP (Parallel Distributed Processing) противопоставляется классической модели искусственного интеллекта (или, как её называют «доброму старомодному искусственному интеллекту» – GOFAI, Good Old Fashioned Artificial Intelligence), исходившей из последовательной переработки информации (как это имеет место в обычных компьютерах) с помощью алгоритмических процедур, применяемых к символическим структурам в виде ментальных репрезентаций.<sup>13</sup>

Коннекционизм придерживается другого представления, которое он считает гораздо более соответствующим реальной работе человеческого мозга. Мозг моделируется как переплетение многослойных искусственных нейронных сетей. Единицы этих сетей – отдельные нейроны, которые являются простыми процессорами, при этом взаимоотношения между ними обладают разной силой связности (сила связности моделирует действие синапсов, связывающих в мозгу человека один нейрон с другими). Обработка информации осуществляется параллельно, т.е. одновременно в разных пунктах переплетения сетей, и при этом процесс распределён, т.е. одна индивидуальная связь участвует в хранении разной информации.<sup>14</sup>

Но тогда возникает вопрос о том, как же следует в этом случае понимать ментальные репрезентации. Если для сторонников классического подхода последние были символическими структурами, подобными структурам языка, то для последователей коннекционизма репрезентации должны быть поняты как активация некоторого паттерна, который распределён в нейронных сетях (как в пространстве, так и во времени) и не может быть понят по аналогии с языком. Гипотеза о существовании языка мысли тем самым отвергается. Ряд сторонников классического подхода (прежде всего Дж. Фодор и З.Пылишин) не приняли коннекционистских новаций. Дискуссия между теми и другими продолжается до сих пор.

Для сторонников коннекционизма появляется проблема, которая не существовала в случае классического подхода: если переработка информации осуществляется мозгом параллельно и распределённо, то как тогда понять существование связности и последовательности в

---

<sup>13</sup> P.Smolensky, On the Proper Treatment of Connectionism // Behavioral and Brain Sciences, 11 (1988): 1-74.

<sup>14</sup> Эксперименты с моделями нейронных сетей в рамках коннекционистского подхода продемонстрировали способность такого рода моделей к обучению распознавания лица, чтению и решению ряда других задач.

познании (особенно в мышлении) и единства сознания? Были предложены разные решения этой проблемы, самым оригинальное из которых принадлежит Д.Деннетту: понимание сознания не как некоего состояния, а как конкуренции разных «интерпретативных структур» и понимание «Я» как условного «центра нарративной гравитации».<sup>15</sup>

Живое обсуждение идей коннекционизма продолжается до сих пор. Тем не менее можно говорить о том, что когнитивное движение вступило в 90-ые гг. в третий этап. Этот этап связан с появлением так называемого динамического подхода в понимании когнитивных систем.<sup>16</sup>

Ранее главный упор в исследовании познания как переработки информации делался на анализе процессов внутри когнитивных систем, на роли в этой связи ментальных репрезентаций как носителей вычислительных действий. Внешняя среда рассматривалась как импульс внутренних процессов и как то, на что система реагирует в результате осуществления внутренней деятельности. Подобное понимание рассматривается сегодня многими исследователями как изоляционизм (отделение когнитивной системы от внешней среды) и на смену картезианства – отсюда и возможность обсуждения в своё время идей «методологического солипсизма». На современном этапе развития когнитивной науки всё более популярным становится другой подход к познанию: оно не сводится к тому, что происходит в мозгу или даже в биологическом теле, а включает постоянное взаимодействие организма и его окружения, познающего и мира. При этом выдвигаются идеи о том, что сама граница между организмом и внешней средой, между «внутренними» и «внешними» процессами условна. Когнитивную систему следует рассматривать как включающую в качестве своих необходимых аспектов мозг, тело и внешнее окружение. Интегратором этой системы является её деятельность. При этом в случае человека внешнее окружение, входящее в этот «расширенный субъект», содержит в себе как естественные объекты, так и культурные артефакты (в том числе язык, миф, науку и т.д.); соответственно познание и сознание должны быть поняты в рамках именно этого «расширенного» субъекта, а не индивидуального организма.<sup>17</sup> Формулируется идея о том, что в ряде случаев необходимо анализировать коллективные познавательные процессы, предполагающие взаимодействие нескольких познающих агентов между собой и с их окружением – как природным, так и культурным. В этом

---

<sup>15</sup> D Dennett. *Consciousness Explained*. L, 1991.

<sup>16</sup> M.Boden. *The Philosophy of Cognitive Scienc*. In: *Philosophy at the New Millennium*. Cambridge, 2001, pp. 209-226.

<sup>17</sup> A.Clark, D.Chalmers. *The Extended Mind* // *Analysis*, 58, 1998, pp. 7-19

случае в качестве носителей этих процессов следует рассматривать соответствующие коллективные организованные системы.<sup>18</sup>

При этом важно заметить, что в новом подходе речь не идёт об отказе от идеи ментальных репрезентаций – просто они понимаются по-новому, принимая во внимание, с одной стороны, результаты изысканий коннекционистов и, с другой, учитывая тот принципиальный факт, что репрезентации производны от взаимодействия организма с окружением. Не идёт речи и об отказе от понимания когнитивных процессов как вычислительных – просто вычисления становятся гораздо более сложными, происходящими не только внутри организма, но и в процессах его взаимодействия с окружением, при этом внешний мир может рассматриваться как аналогичный компьютерной памяти.<sup>19</sup>

На становление нового подхода повлияли практические разработки в области ситуативной робототехники и зрения аниматов.<sup>20</sup> В теоретическом плане сильным было воздействие известной «экологической теории восприятия» Дж. Гибсона и неогибсонистов.<sup>21</sup> В философском плане признаётся влияние феноменологических идей М.Мерло-Понти<sup>22</sup>. Ряд исследователей считают, что на третьем этапе когнитивная наука перешла от парадигмы Декарта к парадигме Аристотеля. Отечественный исследователь легко заметит серьёзную перекличку данного подхода с тем пониманием познания и сознания, который разработан в нашей психологии и философии как деятельностный и культурно-исторический, включая идею «расширенного» коллективного субъекта.

\* \* \*

Я попробую подвести определённый итог сказанному и сделать некоторые выводы.

Прежде всего «натурализация эпистемологии», включение философии в когнитивные исследования во взаимодействии с искусственным интеллектом, когнитивной психологией, когнитивной лингвистикой, нейронауками не только не элиминировали традиционные философские проблемы, а даже сделали их более острыми и выявили их

---

<sup>18</sup> *A. Clark. Embodiment and the Philosophy of Mind // Current Issues in Philosophy of Mind. Cambridge, 1998, p. 49*

<sup>19</sup> См. там же. Р. 40. Вместе с тем среди сторонников этого подхода есть и те, кто готов вообще отказаться от идеи ментальных репрезентаций, так же как от компьютерного подхода к пониманию познания, См. *T. Van Gelder. What Might Cognition Be, If Not Computation? // Journal of Philosophy, 1995, 92/7, pp. 345-381*

<sup>20</sup> *D. Ballard. Animate Vision // Artificial Intelligence, 1991, 48, pp. 57-86*

<sup>21</sup> *Дж. Гибсон. Экологическая теория зрительного восприятия. М., 1988*

<sup>22</sup> *M. Merleau-Ponty. La Structure de Comportment. Paris, 1942; М. Мерло-Понти. Феноменология восприятия. СПб, 1999*

значимость не только для самой философии, но и для специальных исследований познавательной деятельности. При этом философские вопросы приобрели новую форму постановки и обсуждения, а предлагаемые решения оказались способными генерировать конкретные исследовательские программы.

Развитие когнитивной науки, которое в философском плане было стимулировано идеями аналитической философии, привело к необходимости ассимиляции идей, наработанных в других философских направлениях: в частности, в феноменологии, культурно-историческом анализе познания и сознания.

Взаимодействие философии с когнитивными исследованиями оказалось весьма плодотворным как для первой, так и для последних. Если принять во внимание то важное обстоятельство, что во многом определяющим фактором современной цивилизации является бурное развитие новых информационных технологий и переход к информационному обществу (обществу знаний), то можно полагать, что развитие когнитивных исследований (которые сегодня захватывают всё более широкий круг научных дисциплин, включая когнитивную этнологию, когнитивную теорию биологической эволюции и др.) будет ставить всё новые проблемы, важные и для философии, и для специального исследования когнитивных процессов. Когнитивная наука открыла новое поле для философской деятельности. В обозримой перспективе можно ожидать только его расширения.

Вместе с тем важно подчеркнуть, что философия, даже в тех её разделах, которые исследуют познание и сознание, не может быть понята только как раздел когнитивной науки. Во-первых, философская проблематика выходит за пределы той, которая характерна для когнитивной науки в узком смысле слова: в частности, эпистемология не может уйти от анализа нормативных проблем, т.е. от переосмотра и переосмысления существующих норм познавательной деятельности, чем эпистемология исторически всегда занималась, будучи своеобразной «критикой познания». Во-вторых, анализ познания осуществляется сегодня в эпистемологии также в социальном плане (социальная эпистемология), а в философии науки в связи с анализом структуры и динамики развития научного знания, а также в связи с историей научного знания (историческая философия науки). Возможно, что в будущем будут установлены связи между этими способами философского анализа познания и когнитивной наукой в узком смысле слова. Но это дело будущего.

---

## РОЛЬ ФЕНОМЕНОЛОГИИ В ОРГАНИЗАЦИИ МЕЖДИСЦИПЛИНАРНЫХ ИССЛЕДОВАНИЙ В ОБЛАСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА\*

*А. Ф. Зотов*

Пока что наиболее продвинутым орудием предварительного критического анализа состава любого знания является *феноменологический метод*. Поэтому я намерен продемонстрировать роль этого метода в организации взаимодействия ученых разных дисциплин, включая философов, в комплексе исследований по *искусственному интеллекту*, которые, согласно господствующему в современном научном сообществе мнению, призваны сыграть едва ли не ведущую роль в понимании феномена жизни и человеческого мышления.

*Искусственный интеллект* как исследовательская программа представляет собой сегодня конгломерат разных дисциплин, связанных общей задачей. Однако, как показывает практика конференций и семинаров по данной программе, сначала ученые, объединившиеся для выполнения такой работы, объединены не столько задачей, сколько названием, термином, смысл которого участники конференции, семинара или даже начавшей свою работу исследовательской лаборатории понимают очень по-разному, и проходит довольно долгое время, пока складывается такая ситуация, когда группа *заинтересованных* людей, собравшихся вместе, обсуждая вопросы, по-видимому, имеющие отношение к объявленной теме, начинают понимать друг друга. В итоге возникает осмысленный, содержательный дискурс, определяется предмет обсуждения, формируются группы *единомышленников*, которые могут иметь *разные мнения*, разные позиции, но касаются эти мнения одного и того же предмета.

Это – предварительная, подготовительная фаза любого эффективного совместного исследования, когда его тема действительно новая.

Когда мне, человеку, профессионально занимающемуся философией, предложили принять участие в конференции по проблемам искусственного интеллекта, в составе оргкомитета которой числятся весьма известные в философском сообществе специалисты, к тому же работающих в Институте Философии РАН – один по теории познания, другой – по проблеме сознания, и к тому же предложили срочно назвать тему выступления, то первое, что прошло мне в голову – это по-

---

\* Этот текст подготовлен в рамках исследовательского проекта РГНФ. № гранта 04-03-00355а

заимствовать название у М. Хайдеггера, который именно так назвал книгу, где были собраны тексты курса, прочитанного им в 1951-52 гг. в университете г. Фрейбурга. Книга имела немалый успех, поскольку с той поры несколько раз была переиздана. Кроме того, им была опубликована небольшая статья с таким же названием, текст которой имеется и в русском переводе. К тому же сам я получил исследовательский грант РФНФ по теме «Введение в феноменологическую теорию разума» и уже осознал, что проблема разума вообще и рациональности в частности настолько непростая, что итогом моей работы может быть разве что *введение*, провизорный обзор исследований, которые должны быть проведены, чтобы можно было говорить о каком-то общем понимании этого предмета, с первого взгляда кажущегося столь легко доступным (ведь каждый из нас, нормальных представителей рода человеческого, «по определению» существо разумное). К тому же этот предмет так разносторонне исследован логиками, психиатрами, невропатологами, психологами, физиологами и прочими представителями весьма уважаемых и признанных научных дисциплин, что не только его наличие или отсутствие, но и степень развитости или недоразвитости, не говоря уже о патологических отклонениях от «нормы», измеряют, видимо, с высокой степенью точности – и в учебных заведениях, и в клиниках, и при поступлении на работу, и даже «на – глазок», когда судачат о знакомых, начальниках, политиках, деятелях литературы и искусства. Так что, помимо всеобщей осведомленности об этом предмете, налицо широчайший спектр профессиональной информации о нем. Так что тем, кто занимается *искусственным* интеллектом как технической проблемой, кажется, грех жаловаться на недостаток сведений о *естественных* прообразах проектируемых и создаваемых ими в лабораториях и конструкторских бюро, а затем и на заводах интеллектуальных устройствах. Вопрос лишь в том, *насколько хорошо* воспроизводится «естественное» в «искусственном», удалось ли человеку приблизиться к природному (расцениваемому как предел совершенства искусственного), или даже *превзойти* природу. Тогда психолог, физиолог или философ выступали бы в роли судей (или критической инстанции) по отношению к мастеровым людям, которые примеривают на себя личину тульского Левши.

Собственно, так и было на исходе средних веков в Европе, когда рождалось промышленное (механическое) производство, когда часовщики создавали поразительные приборы – не только для измерения времени (вспомним механического павлина из Эрмитажа), когда многими из них овладела страсть создавать механические подобия человеческих существ («Пианистка», «Танцовщица», «Шахматист» и т.п.). Дело это в те времена в Европе было довольно опасным – творцы, вместе со своими творениями, могли быть преданы суду святейшей инквизиции и даже приговорены к сожжению на костре. Причем (нам важно иметь это в виду в связи с нашей темой), как су-



ды, так и создатели таких искусственных существ надеялись на успех в этом (греховном) соревновании с Господом. Такая – мировоззренческая – позиция нашла выражение в искусстве и литературе, как, разумеется, и в философии: книги «Человек – машина» Ламетри и «Изложение системы мира» Лапласа – только наиболее яркие проявления этой всепроникающей идеологии. Есть поучительные примеры из области, видимо, более близкой научно-техническому аспекту проблемы искусственного интеллекта. В книге «Великое искусство», относящейся к началу 13 века, испанский богослов-католик, лингвист и логик Раймонд Луллий изложил принципы построения «мыслящей машины», которая позволяла воспроизводить логические рассуждения. Машина представляла собой соединение особого алфавита, в котором буквы представляли понятия, сочетания фигур силлогизма с помощью семи концентрических колец, вращение которых создавало комбинации терминов, которые, как считал автор, исчерпывали все абсолютные предикаты мироздания. Это идея пользовалась успехом среди ученых вплоть до 18 века, хотя Лейбниц в 1666 г. подверг ее критике, назвав «слабой тенью подлинного искусства комбинаторики». Затем, уже в 19 веке, английский математик Чарльз Бэббидж изобрел аналитическую машину для выполнения математических вычислений разного рода, в которой была «память», программируемое и вычислительное устройство из рычагов и шестеренок, и в которую программа вводилась с помощью перфокарт. Правда, Бэббидж уже понимал, что он не проникает таким образом ни в тайны мышления, ни в глубины устройства мироздания, тем более не моделирует устройство мозга – он понимал, что создает не аналог мыслящего субстрата, а всего-навсего средство для решения расчетных задач в любых областях, где они могут пригодиться. Поэтому он не видел качественной разницы между этими своими изобретениями и другими новациями, вроде страховых таблиц, тахометра или устройства, которое сбрасывало случайные препятствия с рельсов перед паровозом.

Однако и на заре *современной* кибернетики, в работах Н. Винера, по сути, была снова воспроизведена более архаичная мировоззренческая схема, правда, скорректированная и дополненная более поздними поправками. Казалось бы, европейское мировоззрение и методология науки давно избавились не только от механистической формы редукционизма даже в том его виде, который был присущ классификации форм движения материи в изложении Ф. Энгельса, где механическое движение признавалось не только простейшим, но и базисным («физика есть механика атомов»). *Субстанциалистская* натурфилософия сменилась сначала «энергетизмом», а потом была потеснена *функционализмом*, принципы которого стали чуть ли не общепризнанными в научной картине мира после публикации книги

Э. Кассирера «Понятие субстанции и понятие функции». Тем не менее, почти в неприкосновенности сохранилось, хотя бы в качестве объяснительного принципа, стремление видеть в «сложном», в соответствии с этимологией этого слова, нечто «сложненное» из «простых» («более простых») элементов, а его специфику понимать как результат структурных связей, отношений и пр. между элементами. Методологическими терминами, которые обозначали эти «добавки», стали «организация» с ее «уровнями», «система» с ее «степенями сложности», «управление» с его «прямыми» и «обратными» связями, и, наконец, «информация». Последняя даже приобретает качества прежней «субстанции». Однако и образ «машины» не исчез: как вселенная в целом, так и мозг предстают в сознании не только «простых людей», но и большинства ученых как «машина» (правда, теперь *информационная*). Стоит напомнить, что Н. Винер, по его воспоминаниям, занимаясь проблемами технического обеспечения ПВО Британии, отметил в качестве важного факта, что он *случайно* заметил аналогичность структуры участка нервной ткани мозга, связанного с работой рецепторов, со структурой электронного устройства, управляющего огнем зенитной батареи. А ведь за этим (или параллельно этому) развернулось и множество исследований по созданию электронных моделей нервных клеток и «нервных сетей», а также множество дискуссий о том, «может ли машина мыслить». Причем та машина, о которой шла речь, была именно ЭВМ! Так что не только компьютер представал как «электронный мозг», но и живой мозг трактовался как «биологический компьютер» – и это было нечто большее, чем просто метафора. Соответственно, и новая наука – кибернетика – обретает некоторые черты философии; конечно, не в средневековом толковании последней, как «метафизики», а близком духу позитивизма позднего Просвещения – как «науки о наиболее общих законах развития природы, общества и мышления». Эти черты очевидно проступают в винеровском определении кибернетики как «науки об управлении и связи в организме, машине и обществе». Поэтому как раз наши философы быстро прореагировали на ее появление: одни резко негативно (прежде всего, видимо, потому, что пришла она к нам «из за бугра»), другие позитивно; сразу вспомнили о сочинении А. Богданова «Всеобщая организационная наука (тектология)», универсалистскую установку и философский характер которой подчеркнул сам автор в предисловии к немецкому изданию своей книги (1923 г.). Философы не случайно составили ядро Института Системных Исследований, ставшего, пожалуй, главным центром по разработке теоретических проблем кибернетики в нашей стране.

Такова духовная (культурная) атмосфера, в которой рождался класс проблем, которые сегодня объединяет термин «искусственный интеллект». В итоге возникло то, что я назвал бы «семантической ловушкой» (это близко по значению к термину «эпистемологическое

препятствие» у Г. Башляра): те, кто в самом начале занимались конструированием вычислительных устройств наряду с другими, как правило, необычными и сложными, техническими задачами, которые потом были включены в проблематику «автоматики», «распознавания», теорию моделей и технику моделирования, а также организации сложных (в частности, человеко-машинных) систем, включавших компьютеры. Все эти (как правило, сложные) системы получили название «умных машин», «интеллектуальной техники», а *технический* компонент, который обеспечивал их взаимодействие с объектом (распознавание, моделирование, адаптацию – например, в форме обеспечения «гомеостазиса» системы) – *искусственного интеллекта*.

Разнообразие, дифференциация этого класса проблем, *т.е. прикладных технических задач*, с самого начала приводили к глубокой специализации разработчиков интеллектуальной техники и имели результатом, в конечном счете, избавление от «бионических» априорных предпосылок, которые, начиная с принципов, лежавших в основании подхода Винера, выразились в его определении кибернетики. Сегодня они имеют ценность скорей для историков кибернетики, чем для тех, кто связан с практическими, да и теоретическими разработками. Соответственно, большинство последних не оглядывается в своих работах и на достижения физиологов. Как, собственно, не видят они особой пользы и от философских новаций в теории познания, которая тоже избавилась и от попыток редукции содержания знания к ощущениям, и от трактовки ощущений как «преобразовании энергии внешнего раздражения в факт сознания». В общем, и определение познания как «отражения», даже с многочисленными поправками, не касается ядра современной эпистемологии.

Между прочим, и определения логики как «науки о правильном (в смысле «подлинном», хорошем) мышлении» вряд ли годятся для обозначения разработок в области искусственных (формальных) языков и логических исчислений (что, кстати, как раз прямо связано и с теорией программирования, и практикой разработки компьютеров и других «умных» машин).

Значит ли это, что область сотрудничества, полезного взаимодействия между философами и теми, кто называет предмет своих исследований «системами искусственного интеллекта», попросту исчезла, и пришло время сначала определиться хотя бы в том, что содержание и структура тех предметов, которые называются словами, звучащими очень похоже, но настолько различны, что вспоминается старый анекдот о немецком майоре, который надеялся полечить детей у своего подчиненного, вольноопределяющегося, доктора философии, поскольку-де «доктор есть доктор».

Нет, конечно, и первая область, которая мне приходит на ум, где сотрудничество просто необходимо – это характер и перспективы тех

изменений в культуре, которые принесла компьютеризация, тех изменений в образовании, научной работе, характере общения, повседневной жизни, человеческой психике, места и функций человека в больших человеко-машинных системах, как в современных условиях, так и в перспективе.

Что же касается мышления, попробуем, для начала, разобраться с тем, что значит «мыслить», т.е. какой смысл придают этому термину философы, психологи, физиологи высшей нервной деятельности и специалисты по «искусственному интеллекту», а также так называемая «образованная публика», и, прежде всего, разведем друг от друга как *очень разные* предметы «технику» решения «интеллектуальных задач» в нашу эпоху, и трактовку этих проблем в культуре на разных ступенях ее развития, т.е. от того, что Хайдеггер называл «вопрошанием, обращенным к бытию».

Приведу некоторые важные тезисы Хайдеггера касательно сути мышления, как его трактует этот мыслитель:

- Мышление не ведет ни к какому знанию, подобному тому, которое содержится в науках.
- Мышление не доставляет никакой мудрости, полезной для жизни.
- Мышление не разрешает никаких мировых загадок.
- От мышления нет никакой непосредственной пользы для деятельности.

Этимология термина «искусственный интеллект» отсылает к общему историческому прошлому тех дисциплин, которые некогда отпочковывались от некоего общего «ствола», подобно тому, как все современные науки возникли в процессе «отпочкования» от философии в изначальном значении этого термина. Этим «стволом» была попытка понять, что значить «мыслить», что такое разум. В ходе поиска ответа на этот «изначальный» вопрос, исследователи стремились, описывая этот *естественный* феномен как специфическое свойство человека, *воспроизвести* («смоделировать») его частично или полностью в разного рода *искусственных*, т.е. созданных самим человеком, устройствах. Эти устройства могли быть как идеальными (логические формулы и математические алгоритмы), так и вещественными (машины и механизмы различной сложности, в том числе счетные машины и человекоподобные роботы). Достигнув и в том и в другом внушительных успехов, ученые и конструкторы далеко ушли от первоначальной задачи и сформировали множество ставших самостоятельными дисциплин, которые сегодня объединены главным образом только глубокими историческими истоками и общим названием. Но это состояние оказывается источником появления тупиковых исследовательских проектов и попыток искать помощь в других дисциплинах, которые занимаются совсем другими предметами. Поэтому первое, что следовало бы сделать – это четко *сформулировать*

ключевые термины и выразить с их помощью предметы и задачи каждой из дисциплин, т.е. «нормализовать» язык каждой из них и, насколько это возможно, сформулировать «метаязык», в терминах которого обозначаются имеющие место междисциплинарные контакты. В процессе такой работы, важные методологическим принципом которой была бы «феноменологическая редукция», можно избавиться от эпистемологической ловушки, которая состоит в том, что использование термина «интеллект» незаметно побуждает специалистов *разных* дисциплин искать ответы на *свои* вопросы, обращаясь к тем, кто обозначает *собственные* предметные области тем же словом – прежде всего, к философам, психологам и логикам, чей предмет, и в самом деле, ближе к «изначальному» смыслу слова «разум». (То же самое, кстати, можно сказать и о философам, которые полагают, что специалисты по «умным машинам» помогут им понять тайны человеческого мышления). На мой взгляд, исторический и семантический анализ сложившейся ситуации способен избавить от напрасных надежд и тупиковых поисков, а с другой – определить зоны плодотворных контактов в том сообществе, которое сегодня объединено не столько понятийно, сколько фонетически.

Любая исследовательская программа, объединяющая усилия ученых разных специальностей, которые изначально работают в разных областях науки и занимаются разными предметами, предполагает, что ее участники, приступая к совместной работе, уже понимают, что задача, которую они намерены решать совместно, предстает как *новое поле исследований*, которое не является суммой, совокупностью или даже «пересечением» предметных областей, где они уже являются специалистами. Иначе говоря, *образуется новый предмет*. Нередко понятие междисциплинарного исследования отождествляется с понятием комплексной программы, или даже технического проекта, что приводит к немалым трудностям, прежде всего практического (организационного) плана, хотя серьезные комплексные программы, получившие форму конкретного технического проекта, часто в свой состав включают научные коллективы, которые выполняют исследовательские программы. Хотя они и инициированы проектами, но их цели, «логика» их функционирования и способ организации существенно иные. В числе таких исследовательских программ могут быть и междисциплинарные. Общеизвестными примерами комплексных (по сути, технических) программ является знаменитый Манхэттенский проект, или такие проекты НАСА, как программа высадки человека на Луну или создания многоразовых космических кораблей. Как правило, такого рода программы и проекты начинаются с формулировки задачи, которая не вырастает в науке имманентно, «внутри» процесса научного познания, а ставится перед учеными «извне»; ученые же, в качестве исследователей, только *привлекаются* для выполнения та-

кой работы, которая не может быть выполнена уже известными в «прикладной» науке средствами организаторами и вдохновителями проекта. Последние выступают в роли заказчиков, а ученые – исследователи или целые научные коллективы, соответственно, оказываются *исполнителями*.

В *современной* науке общепринято ее разделение на «дисциплинарные уровни», согласно общему принципу «разделения труда» – на фундаментальные исследования, прикладные исследования и разработки. Это отобразено, в частности, в организационной структуре современной науки. Например, в номенклатуре научных званий; в разных странах существуют научные коллективы различных ведомств – например, в России имеется так называемая «академическая наука», развитая система академических институтов с собственными библиотеками, лабораториями, вычислительными центрами и пр., которая, по традиции, ориентирована, прежде всего, на «фундаментальные» исследования; в США эти исследования сосредоточены, главным образом, в крупнейших Университетах. «Прикладные» исследования – в центре внимания научных центров, обслуживающих разные сферы производства, политические учреждения, оборону и т.п. Обычно они занимаются и *разработками*, хотя теперь, как правило, это либо задача заводских *лабораторий*, либо временных коллективов, решающих конкретную задачу, либо, наконец, «многопрофильных» учреждений, которые берут на себя организацию таких временных коллективов и обеспечение их работы всем необходимым. Сегодня организация научной и исследовательской деятельности, во всех ее аспектах, стала особым видом бизнеса, причем очень быстро растущим. Все компоненты этой структуры меняются в зависимости от системы и иерархии социальных ценностей, поскольку, во-первых, «научный мир», при всей его автономии, не изолирован от остальной культуры, как не изолированы друг от друга и его регионы; во-вторых, структура каждого из вышеназванных регионов оказывается аналогичной структуре «разделения труда» в науке как целом (в составе «фундаментальной» науки имеются «прикладные» исследования и «разработки»), которые обеспечивают исследовательскую деятельность; соответственно, в составе развитых «прикладных» наук возникает то, что с полным правом можно назвать их «теоретической саморефлексией» – а также «разработки», обеспечивающие их собственный технический инструментарий. То же самое (хотя, пожалуй, в меньшей мере) справедливо и относительно «разработок». Эта структурная аналогия в различных образованиях современной науки, по-видимому, обеспечивает их взаимодействие, образуя в каждом из них нечто вроде «зон максимального благоприятствования», в которых *все* «защитные механизмы», обеспечивающие функционирование «автономного организма» научной дисциплины, оказываются ослабленными.

Междисциплинарные исследования – сравнительно поздний продукт развития знания, когда наука в целом достигает состояния зрелости – т.е. отделяется от изначального «нерасчлененного» знания о мире и человеке, получает особое место в составе культуры и добивается если не полной самостоятельности, то широкой автономии в сфере культуры, а также обретает дисциплинарную организацию. В свою очередь, научные дисциплины, входящие в состав науки, достигнув зрелости, тоже формируют собственные предметы и соответствующие этим предметам методы. Все эти моменты развития науки и научной деятельности находят отражение в организации специализированных исследовательских и образовательных учреждений (сначала НИИ – в составе академий, и факультеты – в составе Университетов, которые сначала образуют иерархию, но потом организуются как подвижные структуры с широкой автономией и взаимными связями, которые соответствуют общей предметной организации научного знания и научной деятельности, а также установившейся системе социальных ценностей).

Важным компонентом этого процесса достижения наукой состояния зрелости являются процессы в *языке науки*. Сначала роль такого языка выполняет «университетская латынь». По мере развития науки и углубления специализации, функции *общего* языка мирового научного сообщества начинает исполнять математика, и одновременно образуется множество специализированных языков, *словарный* состав которых может быть каким угодно (слова могут черпаться из разных «живых» и «мертвых» национальных языков, в зависимости от конкретной ситуации – например, основной состав языка информатики – английский), а понятийный состав коррелятивен предмету каждой конкретной науки. Об этом свидетельствует наличие множества профессиональных толковых словарей, и также профессиональных сленгов.

Все эти характеристики образуют некое единство, в котором происходят перемены в случае междисциплинарных исследований (что отнюдь не обязательно, когда речь идет о комплексных программах). Проведение комплексного исследования сначала требует от каждого из участников некой предварительной подготовки, т.е. *осведомленности* касательно специфики предметов, методов и результатов, уже достигнутых в каждой из соответствующих дисциплин. (Это можно назвать курсом по преодолению профессиональной ограниченности, которая сопутствует до сих пор профессиональной компетенции, хотя выражение «идиотизм профессионала» сегодня общеизвестно, а возвращение к «университетской парадигме» в системе высшего образования, в противоположность к недавней еще парадигме узкого профессионального образования, которое давали технические институты – в несколько меньшей степени гуманитарные – говорит сам за себя). По ходу комплексного исследования образуется *новый предмет*, разрабо-

тывается соответствующая этому предмету группа методов (некоторые из которых могут быть видоизмененными методами из арсенала «базовых» наук и тех дисциплинарных областей, из которых пришли участники междисциплинарной исследовательской программы). Формируется и язык, соответствующий новому предмету – это уже нечто иное, чем «промежуточное» семантическое образование – язык междисциплинарного общения. Далее повторяется все то, что уже происходило в каждой науке при достижении ею зрелого состояния.

В той мере, в какой и данная междисциплинарная программа, и каждая из дисциплин, представители которой в ней участвуют, являются историческими феноменами, т.е. существуют в конкретном культурно-историческом контексте, ее эффективность зависит от понимания *неявных предпосылок*, которые влияют на стратегию исследования, его организацию, его планирование, оценку ближайших и отдаленных перспектив. Эти предпосылки, будучи неосознанными, являются непеременимыми условиями научного поиска. Но они же часто выступают и в роли *эпистемологических препятствий*, создавая своего рода шоры на глазах теоретиков и экспериментаторов. Поэтому осознание *предпосылочности знания*, как и включений в его состав методологических установок и интерпретаций, которые воспринимаются как органическая часть «позитивного результата», способно сыграть немалую роль в повышении эффективности научного поиска.



---

# ВОЗМОЖНОСТИ И ТРУДНОСТИ В МОДЕЛИРОВАНИИ ИНТЕЛЛЕКТА

*А.П. Огурцов*

А.П. Огурцов

Основной замысел данной статьи состоит в том, чтобы указать на фундаментальную опасность, которая грозит компьютерному моделированию и нейрокомпьютерингу – соблазн принять новую терминологию за новый подход, повторно описать в этой терминологии те феномены и процессы, которые уже долгое время изучались, но заключить, что найдено объяснение тех феноменов, которые уже описаны в этой терминологии.

Сравнение интеллекта с машиной возникло вместе с формированием классической науки. Правда, это сравнение ограничивалось теми полуавтоматами, которые были построены в то время, – с часами. Это сравнение присуще не только Ламеттри, но и Р.Декарту, Лейбницу и многим другим философам и ученым нового времени. Конечно, каждый из них понимал не тождественность человека и машины, понимал по-своему, усматривая нередуцируемые немашинобразные функции и способности человека то в страстях, то в самосознании, то в духовности «монад» и т.д.. Но эта метафора позволила классикам новоевропейской науки найти модели, ставшие средствами «симуляции» движений организма и актов интеллекта, «очистить» его работу от всех «механических» функций с тем, чтобы оставить то, что выходит за рамки сопоставления человека с машиной, не поддается (как сказали бы сегодня) машинному моделированию.

Помимо эмоциональной сферы и самосознания философия обращалась к воображению (И.Кант), к воле А.Шопенгауэр), к бессознательному (З.Фрейд, К.Юнг), к интенциональности сознания (с которой Ф.Брентано связывал дуализм ментальных и физических процессов, а Э.Гуссерль, наоборот, их имманентность, трактовка интенциональности Д.Деннетом как установки на приписывание рациональности поведению и предсказания действий объекта), к боли, к чувству как к тем способностям, которые не могут быть приписаны компьютерам, поскольку выходят за рамки сугубо «машинобразных» функций человека.

Иными словами, философия все время расширяла сферу ментальных способностей человека, выявляла в ней новые слои и такие потенции, которые не могли быть «актуализированы» ни в машинах, ни в компьютерах, а выражены на весьма «неточном», полным метафор, нередко экстравагантном и экзотическом философском языке.

Создание компьютеров привело к тому, что вновь возникло сравнение работы мозга с работой компьютера и оно стало весьма рас-

пространственным. Так, профессор психологии и экспериментальных наук, директор государственной программы Канадского института современных исследований Зенон Вальтер Пилишин, отождествляя интеллект с процессом вычисления, писал: «Возможность описания вычислительной работы и познания в одних и тех же абстрактных терминах устраняет причину, которая вынуждала бы нас рассматривать соответствующее сопоставление как только метафору в противовес его буквальному толкованию. Несмотря на широкое распространение компьютерной терминологии (например, такие термины, как «запоминающее устройство», «процесс», «операция») большая часть такого словоупотребления связана хотя бы с некоторой метафорической окраской. Наблюдается нежелание трактовать вычислительную работу как буквальное описание мыслительной деятельности, а не просто как эвристическую метафору. Отказ от буквального толкования компютации открыл возможность для весьма широкой трактовки разнородной деятельности под рубрикой «теории инфорационных процессов», что в значительной части представляет собой существенный отход от того, что я рассматриваю как ядро компьютерной теории мышления». И чуть ниже: «Если мы будем рассматривать вычисление более абстрактно как знаковый процесс, трансформирующий формальные выражения, которые в свою очередь интерпретируются в терминах той или иной области репрезентации (например, в числах), мы увидим, что представление о ментальных процессах как о вычислении может пониматься так же буквально, как и представление о том, что делают компьютеры системы IBM, адекватно рассматривается как вычисление».<sup>1</sup>

Итак, для Пилишина интеллект – это вычисление, которое представлено в компьютерах «программному обеспечению», а биологии и нейрофизиологии мозга соответствует то, что представлено в «аппаратном обеспечении» компьютера. Этот ход мысли уже существовал в философии Э.Кондильяка, для которого аналитичность мышления и искусство рассуждения представлены в языке, который и является аналитическим методом, осмысление искусства рассуждения сводится к хорошо построенному простому языку, поддающемуся исчислению. Вместе с тем Кондильяк отнюдь не отрицал иных способностей человека – воображения, памяти, созерцания (он называл их действиями души, лишь одно из которых можно выразить с помощью исчисления).

Область Искусственного интеллекта (далее –ИИ) насчитывает более 30 лет. В ней есть большие достижения. Вместе с тем и компьютерное моделирование, и нейрокомпьютинг столкнулись в модели-

---

<sup>1</sup> *Pylyshin Z.W.* Computation and Cognition: Issues in the Foundation of cognitive Science // *The Behavioral and Brain Science*. Vol2.1980. P. 114-115 (111-169). Эта же мысль проводится и в его книге: *Computation and Cognition: Toward a Foundation for Cognitive Science*. Mass. 1984.

ровании актов интеллекта с рядом трудностей, которые показывают недостаточные эвристические возможности современной философии и психологии сознания (в том числе интеллекта).

Создатель исследовательского направления и самого термина «искусственный интеллект»<sup>2</sup> Джон Маккарти настаивал на законности приписывания мыслительной способности компьютерам: «Приписывать определенные «мнения», «знания», свободу воли», «намерения», «сознательность», «способности» или «желания» машине или компьютерной программе – дело вполне законное, если такое приписывание выражает относительно машины ту же информацию, какую оно выражает относительно человеческой личности. Такое приписывание полезно, если оно помогает нам понять структуру машины, ее прошлое или будущее поведение или же облегчит ее исправление либо усовершенствование. Подобный взгляд, быть может, никогда не является логически необходимым даже по отношению к человеческим существам, он просто выражает в достаточно компактной форме представление, какие ментальные (или изоморфные ментальным) свойства может потребовать известное нам состояние машины в конкретной ситуации... Приписывание ментальных свойств машинам известной нам структуры, таким, как термостат или действующие компьютерные системы, не составляет труда, но в высшей степени полезным оказывается его применение к сущностям, структуры которых известна нам в весьма недостаточной степени»<sup>3</sup>.

Как мы видим, Маккарти, хотя и приписывает термостатам определенные мнения о состоянии температуры в комнате, но все же его утверждения гораздо более корректны, чем рассуждения Пилишина: речь идет о *приписывании свойств* известных нам устройств тем сущностям, структура которых нам неизвестна, – мозгу. Такое приписывание полезно и законно, ведь речь идет о изоморфности ментальных свойств человека и процессов компьютера. Как назвать процессы, происходящие в компьютере и изоморфные способностям человека, по мнению Маккарти, не важно. Важно лишь, чтобы информация относительно машины и человека была той же самой. Но что значит «той же самой»? Может быть, уподобление компьютерных процессов и способностей человека – это лишь метафора? Но тогда, где границы этой метафоры? И насколько корректна экстраполяция процессов, происходящих в компьютере (или в термостате), на ментальные способности и действия человека? Таковы те вопросы, которые возникают у каждого, кто познакомился с программными статья-

---

<sup>2</sup> Кондильяк Э.Б. Об искусстве рассуждения и Язык исчисления // Сочинения. Т.3. М., 1983.

<sup>3</sup> McCarthy J. Ascribing mental qualities to machines // Philosophical Perspectives in Artificial Intelligence. Ed. by M.Ringle. New York, 1979. P. 61

ями основоположников новой области компьютерного моделирования – искусственного интеллекта.

Хотя я уже около 15 лет работаю на компьютере, но до сих пор остаюсь «чайником». У меня нередко компьютер «зависает», что вызывает у меня взрыв негативных эмоций. Задумавшись над тем, почему же зависает компьютер, я пришел к выводу, что, очевидно, я вызываю команды, альтернативные друг другу, «не стыкующиеся», не коррелирующие друг с другом. Действительно, компьютерные программы представляют собой алгоритмические предписания, которые построены иерархическим, последовательным (или, как сказал бы Деннет, сериальным) образом и осуществляют дедуктивные операции (нельзя «перепрыгнуть» за тот шаг в дедуктивном следовании операций, который уже запрограммирован). «Не стыкующиеся» команды вызывают альтернативные программы и затем «сбой» в работе компьютера, поскольку фильтры не могут допустить неожиданный «скачок» в вызове команд, не допускают «наложения» программ одна на другую или нарушения их иерархического порядка. Пошаговое выполнение программных команд означает, что единицы языка программирования организованы иерархически (сериально). Аналогичным образом мыслится и устройство нейронных сетей. Это весьма «сильное» допущение, на котором строится композиция программ, всего процессора и работа мозга. Нейрофизиология уже осознала системный и многоуровневый характер нейронных сетей, их нередуцируемость лишь к линейной цепочке сигналов.

Лингвистика уже давно осмыслила не сводимость языка к дискурсивной цепочке утвердительных и отрицательных пропозиций. Но, именно в этой редукции всего богатства языка с его модальностями, многозначностью, метафоричностью, оттенками смысла к линейной цепочке пропозиций, то есть к языку машины Тьюринга, и основывалось моделирование актов интеллекта в теории искусственного интеллекта. Речь идет о богатстве языка, даже не о речи, о богатстве потенций языка, далеко не всегда актуализируемых в речи с ее интонационными, невербальными, коммуникативными характеристиками. Из всей многослойности языка компьютерное моделирование интеллекта «выбрало» лишь одну проекцию – проекцию синтаксиса пропозиций, в которой просто невозможно выразить все многоцветие и все богатство смысловых потенций языка. А если учесть различие языка и речи, которое стало общепризнанным после работ Ф. Соссюра (мы исходим из той трактовки этого соссюровского различия, которое дано Э.Косериу: язык как система потенций, а речь – совокупность их актуализаций), то компьютерное моделирование вообще не приступало к построению симуляционных моделей речи, речевого сообщества, взаимоинтенциональности участников речевой беседы. Здесь еще работы непочатый край.

Поворот к анализу языка, который был осуществлен в 20-е годы XX века в логическом позитивизме был поворотом к языку пропозиций, к их синтаксису. Именно этот поворот обеспечил успехи в моделировании интеллектуальных актов, выраженных в пропозициональном языке. В наши дни совершенно отчетливо осознана нередуцируемость языка интеллекта к этой проекции языка. Поэтому в центре внимания лингвистической философии – дискурс как надфразовое единство, надфразовая целостность. Если же учесть, что дискурс выражен в речевом общении, то задача моделирования существенно усложняется – необходимо моделировать взаимоинтенциональность речевых актов, их понимание и взаимопонимание, их явные и скрытые смыслы, их нагруженность метафорами естественного языка, недомолвками, жестами и прочее и прочее. Диалог между человеком и машиной – это не диалог между представителями речевого сообщества, даже принадлежащих к одному речевому сообществу. Тем большие трудности возникают при моделировании актов коммуникации между представителями разных речевых сообществ с различными культурными установками, ценностными ориентациями и языковыми концептами. Такова одна из важнейших трудностей, встающих на пути моделирования интеллектуальных актов. И в этом отношении концептуальный аппарат структурной и постструктурной лингвистики, в которых проведено различие между акторами и нарратором, между действующими лицами и повествователем, между автором и акторами и т.д., позволил достичь многомерной интерпретации и текстов, и дискурсов<sup>4</sup>.

Помимо того, что в концепциях ИИ язык представлен как синтаксическая цепочка пропозиций, отмечу еще одну трудность моделирования интеллектуальных актов – необходимость поворота внимания к семантике значения, что сразу же включает в область моделирования *интерпретатора*, а вместе с ним различие внешнего и внутреннего планов анализа языка. Различение внешней и внутренней формы слова – хорошо известное завоевание лингвистики еще со времен В.Гумбольдта и ставшее общепризнанным после работ Г.Г.Шпета. Но это различие стало в психолингвистике после исследований порождающей грамматики Н.Хомского различием внешних и внутренних (глубинных) структур языка.

Хочу привести большую цитату из блестящей книги Дагласа Хофштадтера, который провел различие между синтаксическими и семантическими характеристиками формы: «Наш разум оснащен переводчиками, производящими на основе двухмерных схем многомерные значения, такие сложные, что мы не можем их описать сознательно...Субъективно может показаться, что механизм, извлекающий

<sup>4</sup> См. статью *Неретиной С.С., Огурцов А.П.* Дискурс в кн.: Теоретическая культурология. М., 2005. С.252-255; 247-252.

внутреннее значение, совершенно отличен от механизма, проверяющего наличие или отсутствие некоего определенного качества, такого, например, как правильно-сформированность строчек. Возможно, это потому что внутреннее значение – это что-то, что проявляется со временем. Из этого следует, что в схемах, которые мы анализируем, можно говорить о двух видах формы. Прежде всего, там существуют такие качества, как правильно-сформированность, наличие которой можно определить с помощью *предсказуемо конечных текстов*... Я предлагаю называть это *синтаксическими* характеристиками формы... С другой стороны, *семантические* характеристики формы не могут быть проверены с помощью предсказуемо конечных тестов: для них требуются *открытые* тесты... Анализ извлечения значения из строчки означает, по сути, установление всех связей данной строчки с остальными строчками, и это, в свою очередь, выводит нас на бесконечную дорогу»<sup>5</sup>. Итак, даже в монологической речи помимо правильно сформированной синтаксической цепочки фраз (строчек, пропозиций) следует вычлениить более глубокие процедуры выявления семантического *значения*, что предполагает как интерпретатора, так и осознание включенности этой фразы в целостность дискурса – актуального и виртуального, наличного и исторически представленного, а тем самым включение параметра времени, уходящего в бесконечность интерпретации. То обстоятельство, что каждая пропозиция может быть понята лишь в контексте целого (текста, теории), уже осознано логиками (в этом существо тезиса Дюгема-Куайна), но как смоделировать временные, глубинные семантические структуры, скрытые за внешними, пространственными цепочками пропозиций, – это (может быть, пока!) загадка.

Я не принадлежу ни к компьютерофилам, ни к компьютерофобам. Для меня важно понять достижения в компьютерном моделировании интеллекта и трудности, стоящие перед ним. Для меня не приемлемы те программы, которые отождествляют работу компьютера с интеллектом и тем более с сознанием. Не секрет, что в наши дни такого рода отождествления на волне успехов компьютеризации и нейрокомпьютеринга весьма распространены. Так, Деннет писал: «Осознающие себя человеческие сознания являются более или менее сериальными виртуальными машинами, имплантированными – непосредственно – на параллельном hardware («жесткий» диск, на котором запрограммирована программа, или мозг – авт.), который нам предоставила эволюция»<sup>6</sup> По его интерпретации, человеческий мозг – виртуальная машина, порождающая сознание.

<sup>5</sup> *Хофштадтер Д.* Гедель, Эшер, Бах: эта бесконечная гирлянда.. Самара, 2001- С.544.

<sup>6</sup> *Dennet D.* Consciousness Explained. London.1991.P.218 (См. также *Юлина Н.С.* Головоломки проблемы сознания. Концепция Д. Деннета. М.2004).

Одна из трудностей ИИ состоит, по моему, в том, что в различных концепциях ИИ речь идет о мозге, но не о теле. Правда, в философии XX века, начиная с А.Бергсона, Э.Гуссерля, М.Мерло-Понти и кончая французскими постмодернистами (Ж.Делез, Ж.Деррида и др.) обратили внимание на роль *телесных схем* в концептуализации мира, в восприятии его пространственных, временных и вообще предметных характеристик. Способность восприятия наших органов чувств адаптирована к определенному оптическому, звуковому и т.д. диапазону волн в отличие от ряда животных, адаптированных, например, к ультрафиолетовому (например, бабочки) или ультразвуковому диапазону (например, дельфины).

Иными словами, мир задан в телесных схемах. Он представляет собой мезокосм, конструируемый с помощью телесных схем и когнитивной активности. «Мы не можем выйти за пределы той области, которая определена возможностями нашего тела и нашей нервной системы. Не существует никакого иного мира, кроме того, о котором мы узнаем через эти процессы, - через процессы, которые поставляют нам данные и из которых мы устанавливаем, кто мы есть. Мы находимся внутри некоей когнитивной области, и мы не в состоянии выпрыгнуть из нее или установить, где она начинается и как мы ее обрели»<sup>7</sup>.

Итак, уже при анализе чувственного фундамента человеческой активности не следует ограничиваться активностью мозга, а необходимо осознать объективные границы нашего восприятия, позволяющие адаптироваться к окружающей среде и перестроить ее. Кроме того, следует учитывать биологическую коадаптацию и коэволюцию органов чувств человека. Тем более, следует учитывать расширение диапазона наших восприятий, которое осуществляется благодаря достижениям техники. Именно в технических изобретениях и приборах П.А.Флоренский увидел органопроекции нашего тела, его продолжение, не просто амплификацию органов чувств человека, но амплификацию, расширяющую их диапазон<sup>8</sup>. Спустя столетия Г.Андерс написал книгу «Антикварность человека», в которой он проводил мысль о том, что телесная организация человека не адекватна современным техническим устройствам, что человек сталкивается с целым рядом трудностей при адаптации к требованиям современной техники (скорость реакции на те или иные вызовы со стороны технических устройств, возможности ответов оператора на те или иные технические параметры и т.д.). В «человеческом факторе» усматривают одну из причин техногенных катастроф. Этот мотив антикварности

---

<sup>7</sup> Varela F. Der Kreative Zirkel. Skizzen zur Naturgeschichte der Ruckbezuglichkeit // Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Munchen. 1998. S.306.

<sup>8</sup> Флоренский П.А. Сочинения. Т.3(1). М., 1999. С.374-422.

биологической основы интеллекта становится решающим при сравнительном обсуждении робототехники и биологии человека. Как сказал Х.Моравец, «протеин – не идеальный материал. Он устойчив только в узком диапазоне температуры и давления, очень чувствителен к радиации и представляет помеху для многих конструктивных решений и компонентов. Человек, созданный генетической инженерией, – это всего лишь второсортный робот, в который вмонтирован неисцелимый дефект – его конструкция задана протеиновым синтезом под руководством ДНК»<sup>9</sup>.

В осознании «антикварности биологической основы» человеческого сознания и заключен исток поворота к функционализму в интерпретации ИИ. Для функционалистских концепций ИИ существенен отказ от поиска субстанции интеллекта, от всех субстанциалистских вариантов (мозг, нейросети, тело) интерпретации интеллекта. Отсюда – один шаг к тому, чтобы трактовать сознание как виртуальную машину, где мысли сами себя мыслят. Иными словами, возвратиться к *самосознанию* как основному определению сознания, к самосознанию, которое не определимо в пространственных характеристиках, а предполагает имплантацию на той телесной основе, которая предоставлена эволюцией. Такова позиция Д.Деннета.

Если специфика сознания связывается с актами самосознания (а именно этим объясняется и поворот к картезианству, присущий и лингвистике, и философии сознания), то возникают новые трудности перед концепциями ИИ. Прежде всего, самосознание не может понято вне контекста взаимоотношений Я и Другой, вне контекста коммуникации, вне анализа сообщества (речевого, трудового и пр.). В конце XX века философия обратилась к исследованию коммуникативной природы сознания, значимости коммуникаций и сообщества в научных исследованиях, в освоении языка и др. Здесь возможны по крайней мере две точки зрения: одна из них усматривает в коммуникации последнее основание для выявления социальных характеристик языка, общества и т.д., а другая – саму коммуникацию выводит из коммуникативного действия. Представитель первой – Н. Луман, представители второй позиции – Ю. Хабермас и К.-О.. Апель. Для Лумана «не действие, а коммуникация является неразложимой (социальной) социальной операцией, и к тому же, именно той операцией, которая неизбежно включается всегда, когда образуются социальные ситуации»<sup>10</sup>.. В противовес ему Хабермас и Апель исходят из первоначальности коммуникативного действия, обращаясь при обос-

---

<sup>9</sup> *Moravec Hans. Mind Children: The Future of Robot and Human Intelligence. Cambridge.1988.P.108.*

<sup>10</sup> *Луман Н. Что такое коммуникация? // Социологический журнал. 1995, № 3.С.114-125.*



новании этой позиции первый к социальному действию, второй – к речевой коммуникации.

Итак, трудности моделирования интеллектуальных актов различны – ограничение актов интеллекта языковыми выражениями, взятыми в одном (пропозициональном) измерении, оставляя вне поля зрения те многообразные аспекты языка, которые связаны с формированием дискурса, а тем более с речевым диалогом, всегда взаимоинтенционального, ориентированного на взаимопонимание и включающего в себя различные модальности речи – от вопроса до сомнения, от долженствования до приказа. Кроме того, концепции ИИ, даже обращающиеся к нейрокомпьютингу, все же ограничиваются функциональными изоморфизмами между работой мозга и компьютера. Свести проблему сознания к проблеме соотношения мозга и сознания, к функциональному изоморфизму между работой нейронных сетей и сознанием вряд ли удастся, хотя бы потому, что и нейронные сети, и сознание не поддаются описанию в линейных, пошаговых схемах, хотя такого рода модели строятся. И работа нейронных сетей, и тем более работа сознания всегда не линейны, предполагают не просто возможность, но и действительность интегративных, целостных систем с различными уровнями обработки сигналов.

Отметим трудности нейрокомпьютинга и построения искусственных нейросетей. Прежде всего все компьютерные модели далеки от биологической основы работы интеллекта. Более того, все сильнее слышатся голоса о том, что тело и мозг далеко не адекватная биологическая основа для сознания. Достижения робототехники трактуются как способ преодоления «антикварности человека», телесно-субстанциального воплощения сознания.

Кроме того, очевиден рост интеллекта, его функций и форм. Поэтому возникает истолкование эволюции как роста научения и обучения. Иными словами, биологическая эволюция трактуется в терминах обучения. В связи с этим и возникает поиск новых алгоритмов, позволяющих осмыслить эволюцию интеллекта. Хотя в настоящее время и разрабатывается модель генетических алгоритмов, или эволюционного исчисления<sup>11</sup>, однако в ИИ сохраняется трактовка и алгоритма как точного, однозначного предписания, и интеллекта как алгоритмически вычислимого и предсказуемого детерминистического следования однозначному алгоритму в то время, как *сознание и мышление прежде всего заключается в изменении и трансформации правил, в нелинейной организации и самоорганизации процессов, в саморефлексии и самоизменении исходных оснований и принципов организации, т.е. в самодетерминации и саморазвитии*. Нелинейность актов интеллекта может

---

<sup>11</sup> Goldberg D.E. Genetic algorithms in search, optimization and machine learning. Massachusetts. 1989.

быть продемонстрирована хотя бы в процедуре абдукции. Однако моделей такого рода процессов пока не найдено, хотя их можно увидеть в идеях аутопойесиса Матураны и Варелы<sup>12</sup>.

Но для меня столь же не приемлемы как программы, которые полагают, что история компьютеризации и нейрокомпьютинга – это история постоянных триумфов, так и те, которые не видят успехов в компьютерном моделировании работы интеллекта.

Итак, в компьютерных системах существенна роль программного обеспечения, которое представлено в логических структурах иерархического, дедуктивного порядка и выражено в бинарном языке («да» или «нет»), причем это язык пропозиций (или предложений) утвердительных или отрицательных. Вместе с тем компьютер – это определенного рода «железки», техническая аппаратура определенной сложности, физические процессы в которой «переводят» логические характеристики в двухзначный сигнал или в физические состояния аппаратуры («есть сигнал» или «нет сигнала»). Иными словами, этот «перевод» логических состояний в физические может быть представлен как конструирование изоморфизма между ними и соответственно между состояниями технической аппаратуры и действиями интеллекта.. Именно в этом изоморфизме логических и физических состояний и коренится основная сложность в понимании компьютерного моделирования. Существенно то, что архитектура процессора и всего компьютера не допускает «наложения» одних программ на другие, их корреляции друг с другом, а пропозициональное исчисление строится как *логика тождества*. Между тем нейрофизиологические исследования мозга показали его системную архитектонику, существование в нем различных уровней – иерархических, параллельных, «кольцевых», циклических, вероятностных, а не только последовательных процессов.

Исходная позиция – та, которую можно назвать слабой версией ИИ: возможна симуляция работы мозга и интеллекта, но невозможно их дублирование по причинам, которые будут отмечены чуть позже.

Достижения в компьютерном моделировании актов интеллекта есть ничто иное, как моделирование вербальных актов интеллекта, т.е. лингвистических и лингво-философских отчетов о работе интеллекта. *Тем самым компьютерное моделирование и ИИ представляют собой не органопроекции нашего интеллекта, а органопроекции наших представлений об интеллекте и его актах. Наш интеллект гораздо богаче, чем те представления о нем, которые амплифицируются в программном обеспечении, вероятностных связях и в компьютерных моделях. Более того, богаче и представления о структуре интеллекта и его нейрофизиологических основаниях.*

---

<sup>12</sup> Maturana H.R., Varela F. The Tree of Knowledge. Doston, London. 1988.

«Машина Тьюринга» представляет собой «амплификацию» представлений о мышлении как вычислении, «органопроекцию» программы анализа актов интеллекта как элементарных операций счета.

Логические операции пропозиционального исчисления, развитые в алгебре логики Д.Буля и возведенные в один из канонов стандартной концепции науки в логическом атомизме 20-х гг. XX века, стали идейным резервом построения компьютерных моделей, обрабатывающих двухзначный сигнал. Поэтому утверждения о том, что «мы представляем собой машины Тьюринга»<sup>13</sup> (такова позиция не только Х.Патнэма, а многих философов и психологов особенно бихевиористского направления) базируются на идее функционального изоморфизма двух систем, хотя физическая реализация ментальных состояний может быть различна, а сама машина Тьюринга может быть представлена просто как последовательность состояний, команд для осуществления простых операций, а не как физическая система. В таком случае мозг оказывается «цифровым компьютером». Впрочем, сам Патнэм отмечает, что сопоставление работы мозга и машины Тьюринга является упрощением<sup>14</sup>, хотя и отстаивает позицию, согласно которой психология есть лишь программное обеспечение мозга как компьютера, переводя эту проблему в проблему функционального тождества состояний мозга и состояний сознания, их функционального изоморфизма.

Мною уже были отмечены корреляции между: 1) методом перебора вариантов и концептуальной схемой бихевиоризма; 2) Общим Решателем Проблем и концептуальными схемами гештальт-психологии; 3) программой Бэкона и концептуальными схемами индуктивизма<sup>15</sup>.

Ограничилось ли компьютерное моделирование «двухзначной логикой» кодирования? Нет, и здесь уже давно выявлены новые возможности. Моделирование трехзначного кодирования в машине «Сетунь», разработанной в МГУ под руководством Н.М.Брусенцова в 1965 г.<sup>16</sup> предполагает построение трехзначной логики (Д.Н. Юрьев, А.С. Карпенко), что открывает новые перспективы в технической разработке нейрородных элементов и искусственных нейросетей.

Столь же впечатляющи и достижения ИИ после формулирования первой программы Д. Маккарти. Назовем лишь некоторые, наиболее известные и общепризнанные результаты. К ним принадлежит, конечно, теория фреймов М. Минского. Обоснована возможность моде-

<sup>13</sup> Патнэм Х. Философия сознания. М., 1999, с.89.

<sup>14</sup> Там же, с.74.

<sup>15</sup> См.: Огурцов А.П. Междисциплинарные исследования творчества: итоги, поиски, перспективы // Междисциплинарный подход к исследованию научного творчества. М., 1990, С.39.

<sup>16</sup> Брусенцов Н.М. Опыт разработки трехзначной вычислительной машины // Вестник МГУ. Серия 1: Математика, механика. 1965 № 2.

лирования интеллектуальных актов по продуктивному правилу «Если...то». Создание языка программирования «Prolog» – первое приближение к более широкому моделированию вербализуемых актов интеллекта. Не менее впечатляющие и успехи ИИ в построении ряда специализированных программ. Например, построение различных моделей медицинской диагностики, в том числе программы EMYSIN (Ван Мелл) – универсальной системы диагностики, усиленной механизмом логического вывода на основе продуктивного правила «Если...то». Создание ряда специализированных программ, например, программы Brightware с элементами самообучения для автоматической генерации ответов на вопросы пользователей Интернета, для автоматических служб телефонных кампаний и др.

Большую известность приобрели исследования, связанные с построением моделей мышления как актов распознавания (М.М.Бонгард, А.С.Кронрод, Д.С.Чернавский, Н.М.Чернавская, В.П.Карп, А.П.Никулин) и как системы различных «блок-схем» (банк объектов, преобразование признаков, обучения, построения решающего правила, его оценки, внимания, использования, предварительного диагноза).

Определены пути построения моделей эволюции научения (компьютерные модели Darwin 2 и 3), на базе которых выдвинута концепция «нейронного дарвинизма» Эделмана. Само собой разумеется, что это лишь некоторые из достижений компьютерного моделирования интеллекта и ИИ, которые мне известны. Очевидно, специалист в этих областях укажет гораздо большее число достижений современных исследований.

Итак, ни у кого не вызывает сомнений то, что компьютерное моделирование представляет собой важную часть многообразных «приставок-амплификаторов», входящих в экстрацеребральную и транссиндвидную «интеллигибельную материю», создаваемую людьми в социокультурном развитии.

Известно, что современная философия широко использует компьютерную терминологию (программы, язык программирования, продуктивные правила, формат, фреймы и пр.), но мало кому известно, что компьютерное моделирование и нейрокомпьютинг амплифицируют в своих схемах и «железках» те представления о мышлении, которые уже развиты в философии и психологии, что язык философии и психологии (может быть не современной, а XIX и XX веков) стал языком, на котором программируются и конструируются компьютерные модели. Поэтому трудности современного компьютерного моделирования – это прежде всего трудности философии и психологии сознания, а перспективы компьютерного моделирования и нейрокомпьютинга – это перспективы философии и психологии сознания. Именно ими будет создан новый язык анализа сознания, в том числе интеллекта, который найдет свою амплификацию в компьютерных моделях.

Поэтому было бы целесообразным сопоставить то, как мыслится интеллект в психологии и в исследованиях по ИИ, насколько адекватны язык психологии и ее модели достижениям и трудностям компьютерного моделирования интеллекта и ИИ. Сразу же оговоримся, что в психологии интеллект отождествляется с мышлением, а в исследованиях по ИИ – с сознанием (Mind). Такого рода отождествления означают, что исследования интеллекта как мышления или как сознания сразу же нагружены неоправданными метафорами и допущениями. В область искусственного интеллекта сразу же включаются те смыслопорождающие и смысловоспринимающие области мышления и тем более сознания, которые в интеллект не следует включать, поскольку они гораздо более широкие и более сложные по организации, чем сфера интеллекта, включают в себя воображение, волю, эмоции и др. Когда мы говорим об *интеллекте*, мы имеем в виду *мышление, взятое в одной проекции – проекции последовательности операций при постановке и решении задач и выраженных в языке пропозиций*. И это различие интеллекта и рациональности (ratio) осуществлено уже в средневековой схоластике.

Особенностью психологии мышления является то, что она трактовала его как *процесс* или как *деятельность*. Именно по этой линии проходит демаркация между различными исследовательскими программами психологии мышления. Кроме того, психология мышления всегда обращалась к индивидуально-личностным факторам формирования мышления. Так, для У. Джемса мышление – умственный процесс, обеспечивающий ориентацию в новых данных опыта и включающий в себя анализ, отвлечение, личный интерес и пронципальность. А.Адлер, развертывая трактовку психологии как индивидуальной психологии, трактует мышление как один из процессов жизни, для которого характерны движение к поставленной цели, упорядочивающей и обеспечивающей целостность всех функций души. Ж.Пиаже, проводя различие между логической и биологической природами интеллекта, усматривает в интериоризации внешних операций способ обеспечения устойчивого и подвижного равновесия между универсумом и мышлением.

В зарубежной психологии обращение к схемам деятельности характерно для гештальт-психологов.. О.Зельц проводит различие между репродуктивным и продуктивным видами духовной деятельности; для первой характерна рутинная актуализация средств, а для второй – открытие новых методов решения задач. В этой школе мышление стало трактоваться как процесс или как деятельность решения задач, а задача психологии стала усматриваться в определении факторов их решения. Так, Р.Крачфилд, Д.Креч выделили ситуационные (стимул, временная организация, эмоциональное и мотивационное состояния) и личностные (уровень знаний, тип личности и др.) факторы.

И в отечественной психологии мышление интерпретировалось как процесс и как деятельность. Причем нередко эти две трактовки понимались как тождественные. Такова позиция С.Л.Рубинштейна, который вычленил три процедуры мышления – анализ, синтез, обобщение. П.Я.Гальперин истолковывал мышление как совокупность умственных действий, вычленил пять их видов: от ознакомления с задачей и образцами действий до акта внутренней речи. Для А.Н.Леонтьева мышление – сплав внутренних и внешних действий, причем решающее значение имеет интериоризация внешних действий. В работах Н.Г.Алексеева начинается проводиться различие между схемой процесса и схемой деятельности применительно к анализу мышления, причем схема деятельности позволяет осмыслить один из наиболее существенных механизмов мышления – рефлекссию собственных средств и оснований, ревизию прежних способов решения задач, способов организации и уровней, считавшихся фундаментальными и исходными.

Исследования в области ИИ обратились к старому, известному еще со времен схоластики понятию *интенциональность*, которое якобы позволит найти *differentia specifica* сознания. В интенциональности стали усматривать внутреннее свойство сознания (Дж. Фодор, Д. Серль, Т.Нагель и др.) или внутреннюю установку сознания (У.Куайн, Р.Порти, М. Минский, Д. Деннет).

Если задуматься о том, релевантен ли язык современной психологии исследованиям в компьютеризации интеллекта и ИИ, то сразу же надо сказать о его нерелевантности. Хотя психология в XX веке достигла многого в изучении мышления – обратила внимание на то, что существуют различные типы мышления (эмоциональное, наглядно-действенное, практическое и т.д.), на антропологическую, индивидуально-личностную «природу» мышления и даже на коммуникативность сознания, однако категориальные и методологические средства психология мышления оказались далеки от проблем компьютерного моделирования интеллекта и ИИ.

Возникновение когнитивной психологии возродило надежды на то, что именно в ее языке может быть найден язык релевантный компьютерному моделированию интеллекта и ИИ. Дело не только в том, что когнитивная психология широко использовала компутационный язык, говоря, например, о фреймах как рамках соотнесения, предвосхищающих получение и отбор приоритетной информации из избыточного массива информации, о формате как системе приема определенного вида информации, но и предложила свой язык для описания перцептивных схем (планов сбора информации) и когнитивных карт. Язык когнитивной психологии скорее всего окажется перспективным при моделировании мышления как распознавания образов, при выделении специфических блок-схем именно этих психических процессов и операций но далеко не во всех областях компьютерного моделирования и ИИ.

Уже неоднократно отмечали явные упущения когнитивной психологии в трактовке мышления. Так, один из основателей когнитивной психологии У.Найссер отметил, что, анализируя мышление как процесс переработки информации, когнитивная психология элиминирует из актов познания их культурно-историческую определенность и многообразие культурно-исторических контекстов функционирования и развития мышления<sup>17</sup>. Кроме того, когнитивная психология (особенно на первых этапах своего развития) элиминировала из всего круга своих проблем нагруженность мышления языком. Хотя философской базой когнитивной психологии была аналитическая лингвистическая философия, однако в ней не были «ухвачены» ни формообразующая роль языка как источника категоризации бытия, ни выражение интеллекта в языке пропозиций. В когнитивной психологии личность трактуется как самость (Self), как та точка осознания, суждений и действий, которая обеспечивает устойчивость самоидентификации человека. И в истолковании личности, и мышления когнитивная психология исходит из идеи тождества личности, будучи направлена на поиск способов самоидентификации человека. И в этом акценте на тождестве когнитивных процессов личности это направление психологии целиком и полностью находится в классической системе отчета, которая кладет в основание *логику тождества в противовес* неклассическому способу мысли, для которого важна логика различий<sup>18</sup>.

Теоретики компьютерного моделирования и ИИ вынуждены были взять на себя осмысление того, что же они моделируют, каковы особенности сознания. В этой области философско-психологической интерпретации сознания с позиций ИИ сложились два направления, которые можно обозначить как «мягкую» и «жесткую» программы. К первой принадлежат К.Поппер, Дж.Экклз, Дж.Р.Серль, Т.Нагель, ко второй – Д.Деннет, Р.Рорти и др. Различия между этими программами касаются многих проблем таких, как отношение сознания и тела, существа сознания, возможности актуализации сознания в различных физических структурах – от тела до физической аппаратуры и т.д.

В нейрокомпьютинге основная функция мозга усматривается в том, чтобы давать представления о внешнем мире. С этим связаны интерпретации и когнитивных карт, и процессов научения, и следов памяти, подчиненных адаптации организмов к окружающей среде в

---

<sup>17</sup> Он писал: «Недостаточная экологическая валидность, безразличие к вопросам культуры, отсутствие среди главных характеристик восприятия и памяти, как они проявляются в повседневной жизни, способны превратить такую психологию в узкую и неинтересную область специальных исследований» (Найссер У. Познание и реальность. М., 1981, с. 119).

<sup>18</sup> Именно в этом направленность такой работы Ж.Дилеза, как «Различие и повторение» (СПб., 1998).

то время, как решающей характеристикой человеческого отношения к миру является конструирование искусственной среды.

Компьютерное моделирование имеет дело с данными, уже включенными в память. Трудность заключается в том, что сам выбор этих данных осуществляется человеком и не может быть (может быть, пока?) передан компьютеру, который предстает как «амплификатор» ряда достаточно простых операций человеческого сознания.

Существующий семантический пробел, т.е. различие между естественным языком и языками программирования, не может быть преодолен до тех пор, пока не будут созданы новые средства анализа естественного языка и интеллекта, которые позволили бы осмыслить его метафорические («тропологические») истоки, его укорененность не в однозначном вербальном выражении («пропозициях»), а в фигурах речи, предполагающих не только способы аргументации, но и молчание (паузы с различным смыслом), невербальные акты («жесты», «миимику» и др.), интонацию и т.д. Важнейшие шаги на этом пути сделаны лингвистами (Н.Д.Арутюнова, Ю.С.Степанов, А.Вежбицкая), которые сделали предметом своего исследования концепты «родного языка» и их сравнительный анализ. Большую трудность представляет собой проблема обучения компьютеров и моделирование на компьютерах процессов развивающегося обучения.

Ныне надежды возлагаются на теорию «нечетких множеств» Л.Заде (1965) и на построение нечетких логических выводов («нечеткой логики»). С этим связывается переход от классических способов представления знаний с помощью продуктивных правил к нечетким способам представления знаний (такovým является язык программирования Prolog). Следует обдумать все следствия из данного подхода, в том числе и новые трудности, которые возникают при нем, прежде всего относительно алгоритмизации «нечетких логик».

И алгоритм, и продуктивное правило, и способы представления знаний трактуются в компьютерном моделировании и в нейрокомпьютинге как логический интеллект, т.е. как вынесение суждения на основании однозначного алгоритма, имеющего наиндивидуальную, общезначимую ценность в то время, как мышление всегда индивидуально. Поэтому элиминация субъективной ментальной реальности – сознательного субъективного опыта, Самости – Self (по английски), Selbst (по-немецки) влечет за собой не столько утверждение «научного материализма», сколько гальванизацию идей метемпсихоза, т.е. воплощения души в различные существа (в этом суть позиции Д.Деннета, по словам которого «могла бы существовать сознательная Самость, чьим телом был бы робот, а мозгом – компьютер»<sup>19</sup>). Позиция К.Поппера и Д.Экклза, согласно которым «мы не в состоянии построить электрон-

---

<sup>19</sup> Dennett D. *Consciousness Explained*. L., N.Y. 1991, p.43



ные компьютеры, наделенные сознательным субъективным опытом»<sup>20</sup>, мне представляется гораздо более осторожной и более корректной.

Однако противопоставление в интроспекционизме сознания, замкнутого в тождественности Я, всем достигнутым формам моделирования актов интеллекта, все же недостаточно. Оно просто не продуктивно. В отечественной философии и психологии сознания развита концепция, которая настаивает на постижении субъективности в единстве с объективностью «многомерных надстроек природных возможностей человека», вместе с «приставками-амплификаторами», образующими мир «интеллигибельной материи», в которую субъективность включена в качестве внутреннего элемента структуры<sup>21</sup>. Этот подход, по моему мнению, позволяет выйти за пределы и дуализма, и материализма, сколь бы научным он не репрезентировал себя. «Вселения» Самости в любые материальные структуры, дает возможность найти путь объединения развернутых ныне форм моделирования актов интеллекта с достижениями психологии и объединить анализ сознания, самосознания и их объективации в социокультурных амплификаторах. Новые перспективы в ИИ открывают нанотехнологии, в том числе молекулярный структурный дизайн и синтез супрамолекулярных архитектур (Н.А.Бульенков)<sup>22</sup>.

Итак, философия в XX веке выработала концептуальный аппарат, который с успехом используется и в психологии сознания, и в компьютерном моделировании интеллектуальных актов (самосознание, интенциональность, внутренние установки и др.). Однако далеко не все возможности в исследовании сознания, предоставляемые современной философией, используются в компьютерном моделировании интеллекта, которое (надеюсь, пока!) основывается на моделях Тьюринга.

---

<sup>20</sup> *Popper K., Eccles J. The Self and its Brain. Berlin, N.Y., London. 1977, p.208.*

<sup>21</sup> *Зинченко В.П., Мамардашвили М.К. Проблемы объективного метода в психологии // Вопр. философии . 1977 № 7. С.21-35.*

<sup>22</sup> *См. статью Н.А.Бульенкова в сб.: «Методология науки: статус и программа». М., 2005.*

---

## МЕЖДУ ЗНАНИЕМ И НЕЗНАНИЕМ – НАИВНАЯ ТОПОГРАФИЯ 2

*А.С. Нариньяни*

Пространство оппозиции «Знание--Незнание» – ключевая мета составляющая Системы Знания, намного более сложная, чем простое противопоставление, определяющее границы нашей Картины Мира. Даже самый примитивный аппарат грубой оценки компонентов знаний и комбинации нескольких мета утверждений позволяет выявить в этом Пространстве черновые контуры структуры, представляющей собой первый – хотя и очень поверхностный – но все-таки достаточно наглядный набросок его многомерной «карты».

### ВВЕДЕНИЕ: В НАЧАЛЕ ЗНАНИЯ

Инженерия знаний, также как и вся тематика Искусственного интеллекта, продолжают, несмотря на почти полувековую историю, развиваться на стадии эмбриона. Поскольку эта область с самого начала позиционирует себя как прикладная, авторы многих ad hoc формализмов не только пытаются строить их теоретически, но и готовы смело пускать в ход при создании реальных систем, что не может не быть «чревато».

Для того чтобы избежать опасности такой поспешной псевдоформализации, существует только один путь: достаточно тщательное до-формальное исследование изучаемой области. Этот путь является азбукой экономики, физики, лингвистики, да и любой другой сформировавшейся науки и вряд ли инженерия знаний может претендовать тут на какой-то свой исключительный путь.

В частности, пора попробовать разобраться и в сложной системе отношений между Знанием и Незнанием, поскольку без этого вообще трудно определить, чем собственно занимается эта наша очень молодая наука. От попытки рассмотрения области размежевания и взаимодействия Знания и Незнания, пусть даже совсем на первых шагах наивной, уйти невозможно.

Поэтому, взявшись за эту непростую задачу, я буду ограничиваться уровнем именно до-формального исследования, используя для начальной ориентации на этой мало изученной территории самые простые средства.

Данная статья является новой, радикально переработанной версией доклада [1]. Она выстроена следующим образом. Первый раздел посвящен отправной точке темы – рассмотрению оппозиции «Знание--Незнание». Во втором разделе дается набросок аппарата ориентации в

пространстве этой оппозиции, проецирующемся на нашу Систему Знаний. Третий связан с двумя полюсами исследуемого пространства: Абсолютным Знанием и Абсолютным Незнанием. В четвертом и пятом разделах обсуждаются два основных полушария нашей наивной карты – Территория Знания и Территория Незнания, а в шестом – расположенная между этими полушариями Территория Неопределенности. В Заключении подводятся итоги проведенного эксперимента.

## 1. ЗНАНИЕ И НЕЗНАНИЕ КАК ОППОЗИЦИЯ

Бинарные оппозиции являются базовыми составляющими нашей Системы Знаний (далее СЗН), относящимся к разным уровням конкретности: от «прямой» – *тепло* \ *холодно*, *тихо* \ *громко* и многие десятки других, до «мета» – *простой* \ *сложный*, *субъективный* \ *объективный*, *истинный* \ *ложный* и т.п.

1.1. За каждой из таких оппозиций чаще всего стоит многомерное семантическое и прагматическое пространство, в котором, как правило, выделяются области, относящиеся к полюсам данной оппозиции, и некоторая центральная зона, компоненты которой к полюсам отнесены быть не могут.

Для многих простых оппозиций центр ассоциируется с нормой, а аппарат ориентации в этом пространстве включает одну или несколько шкал оценок. При этом и норма, и большинство шкал являются контекстно-зависимыми. Например, для пространства *хорошо* \ *плохо* его центр (*нечто среднее*) и шкала (от *замечательно* до *хуже не бывает*) определяются спецификой области приложения и контекстом данной оценки.

Наша Система Знаний напоминает пирамиду, низ которой состоит из бесчисленных объектов, фактов, процессов и т.п., а верхняя часть относится к ее *мета* составляющей, расщепляемой, в свою очередь, на «восходящие» слои – Метазнания, Мета-метазнания и т.д. При этом понятийная оппозиция «Знание--Незнание», являющаяся одной из базовых в нашей Картинае Мира, относится к вершине этой пирамиды, поскольку охватывает все ее этажи

Рассматриваемое пространство оппозиции «Знание–Незнание» (будем называть его далее *Пространством*) – одно из самых сложных. Конечно, его можно было бы попытаться упростить до предела, представив в виде линейной шкалы, в которой центром-нормой было бы «обычное» знание, нижняя часть представляла «знание все ухудшающегося качества», вплоть до *Полного Незнания*, а верхняя относилась ко «все более высокому», в пределе – к *Абсолютному Знанию*.

Однако, такая схема была бы слишком проста даже для той Наивной Топографии, которая строится в этой работе. В исследуемом Пространстве нет ни центра-нормы, ни общей линейной шкалы, однако есть определенная структура, эскиз которой я и попытаюсь представить.

Перед тем, как приступить к делу, необходимо оговориться, что объект нашего рассмотрения – *Система Знаний* – в этой работе значительно упрощается:

В СЗн включается только «концептуальное» отражение нашей Картины Мира, созданное нашей цивилизацией и отображаемое в языке. При этом из нее исключается ее «вегетативная» часть, формируемая органами восприятия, и являющаяся в первом приближении общей для всех животных, относящихся к высшим. Именно поэтому эта часть Картины Мира и Системы Знаний «выносится за скобки» и не принимает участия в дальнейшем обсуждении.

Таким образом, под СЗн ниже имеется в виду множество различных утверждений как конкретных (факты), так и обобщенных (мета), вплоть до утверждений о самой СЗн. К понятию «Знание» относится все, что может быть вербализовано, т.е. содержание любого текста, как бы его смысл не оценивался с точки любых возможных критериев.

Понятие же «Незнание» в контексте этой статьи обозначает то Знание, о существовании которого в СЗн ничего не известно; иначе говоря, «Незнание» здесь не имеет никакого отношения к оценке высказываний, которые чем-то плохи (неадекватны) в рамках данной СЗн.

*В рамках нашей упрощенной схемы подразумевается, что выводы и заключения делаются любой СЗн всегда корректно. Это предполагает, что несовместимость в СЗн появляется только между новыми компонентами и уже включенными ранее. Однако, очевидно, что это более чем сильное допущение: сама возможность не максимальных оценок достоверности предполагает выводимость из них несовместимых альтернатив. При этом значительная, если не основная, часть знаний, оценка которых в СЗн неадекватна в силу влияния самых разных факторов человеческой природы.*

Таким образом, в связи с до-формальным характером нашего эксперимента все используемые далее понятия заведомо не формальны и ни на какую строгость не претендуют. Можно сказать, что если они на что-то и претендуют, то на *нестрогость*.

1.2. Перед тем, как перейти к собственно рассмотрению, следует обратить внимание читателя на важную методологическую проблему, которая может мешать в дальнейшем нашему общению на «общем языке».

Дело в том, что в искусственном интеллекте (ИИ) при исследовании нового понятия ему часто присваивают конкретную лексическую *этикетку* (слово или словосочетание), позаимствованную из естественного языка. Это нередко порождает путаницу, поскольку смысл изучаемого явления и сопоставленного ему словосочетания – далеко не одно и то же. Словосочетание может иметь целый спектр значений, не всегда различаемых четко, а нередко и с большим трудом, в достаточно узких контекстах. Исследуемое же явление или объект опреде-

ляются их прагматикой, т.е. их ролью в реальности, причем границы этой роли тоже далеко не всегда (для нашей темы заведомо) являются достаточно четкими.

Поскольку в нашей топографии таких этикеток будет много, то попытка ориентироваться в содержании обозначаемого, исходя из лексической семантики обозначающего, может только помешать читателю. В частности, вводимые ниже названия зон Пространства являются лишь этикетками, выбранными (заведомо не слишком удачно) ради мнемонической ассоциации с обозначаемыми ими классами компонентов СЗн.

## 2. ВЫБОР АППАРАТА ОРИЕНТАЦИИ

Понятно, что адекватной следовало бы считать такую модель Пространства, которая отражала бы сумму наших сегодняшних мета знаний о нем. Однако разработка такой модели – задача, явно выходящая за рамки как объема данной статьи, так и возможностей автора, по крайней мере, сегодня.

Оставаясь в этих рамках приходится учитывать, что в начале движения в новом направлении лучше ограничиваться наброском, который должен помочь нащупать концепцию, не усложняемую многочисленными деталями. Это обычно позволяет упростить задачу и автору и читателю, а главное, обеспечить более высокую продуктивность обсуждения. Таким образом, в данной работе я прибегну к «лубку» – предельному упрощению, все компоненты которого представляют собой своего рода иероглифы с достаточно размытой семантикой, не позволяющей определить их более конкретно, чем это удастся сделать ниже.

Я называю представляемый ниже аппарат *лубком*, поскольку он похож на первые географические карты, пытавшиеся отразить на плане представление о мире своего времени: что-то, в той или иной степени знакомое, изображалось по памяти и здравому смыслу, а то, что было за этой гранью, домысливалось в соответствии с признанной тогда Системой Знаний.

Примерно так же придется действовать и нам, строя нашу топографию на основе пока еще очень наивной системы координат. Эту попытку оправдывает то, что и первые карты были нужны, хотя бы для того, чтобы следующие появились с учетом их ошибок, и это, в свою очередь, позволило формироваться более адекватному аппарату, который и в картографии сложился тоже далеко не сразу.

**2.1.** Надо предупредить, что принцип нашего подхода определяется тем, что любая СЗн существует или, по крайней мере, проявляет себя лишь в когнитивно-коммуникативном процессе. Таким образом, исследовать ее структуру можно только «со стороны», поскольку даже при анализе собственной СЗн участвуют две ее проекции: СЗн

КАК объект изучения, и реальная СЗН субъекта-исследователя, достаточно неадекватно отображаемая в изучаемой СЗН хотя бы потому, что сама эта реальная СЗН далеко не совершенна.

Поскольку рассмотрение СЗН в качестве объекта может происходить только в Пространстве СЗН субъекта, то, даже упрощая картину, отметим, что в этом процессе участвуют как минимум четыре СЗН:

*объект* – СЗН<sub>1</sub>,

*внешняя* СЗН<sub>2</sub> субъекта-исследователя, как он ее видит сам, и

*внутренняя* СЗН<sub>3</sub> субъекта-исследователя, какова она «на самом деле», т.е. как она выглядит с точки зрения некой *идеальной* СЗН<sub>и</sub>.

Однако исследователь в общем случае (т.е. тогда, когда он не занимается специально углубленным самоанализом) отождествляет свою СЗН<sub>2</sub> с идеальной СЗН<sub>и</sub>, так что в нашей наивной модели вполне можно ограничиться двумя Системами Знаний – СЗН<sub>1</sub> и СЗН<sub>2</sub>.

**2.2.** Отношение внешней СЗН<sub>2</sub> к объекту СЗН<sub>1</sub> принимает одну из трех известных форм:

*Сверху-вниз*, при которой СЗН<sub>2</sub> оценивает себя как *высшую, охватывающую* по отношению к СЗН<sub>1</sub>, а саму СЗН<sub>1</sub> как *вложенную, более простую* (традиционное отношение учителя к ученику, высшей культуры к более низкой и т.п.).

*Снизу-вверх*: отношение, обратное первому – СЗН<sub>2</sub> оценивает себя как *низшую, вложенную, более простую* в сравнении с *высшей, охватывающей* СЗН<sub>1</sub>.

*Равное*: отношение равного к равному по квалификации коллеге, партнеру, параллельной школе или культуре и т.п..

Принимаясь исследовать Пространство, мы будем считать, что в этой работе наша СЗН<sub>2</sub> выступает к обозреваемой СЗН<sub>1</sub> в отношении *Сверху-вниз*.

Это позволяет отразить в используемом аппарате расщепление в СЗН<sub>2</sub> компонентов знаний и их оценок на *собственные* (взятые из СЗН<sub>2</sub>) и *изучаемые* (принадлежащие СЗН<sub>1</sub>). Таким образом, рассматриваемая карта СЗН<sub>1</sub> проецируется на внешнюю карту субъекта, причем «координаты» общих для СЗН<sub>2</sub> и СЗН<sub>1</sub> компонентов на этих картах совпадают далеко не всегда, поскольку их оценки в этих Системах Знаний могут быть совершенно различными.

**2.3.** Ниже аппарат ориентации базируется в основном на оценках компонентов Системы Знаний ею самой (таким образом, при этом она выступает на основе отношения «равный» одновременно в роли СЗН<sub>2</sub> и СЗН<sub>1</sub>), хотя в некоторых частях Пространства используются и оценки внешней СЗН<sub>2</sub>.

Всякий компонент X Системы Знаний представляет ниже фрагмент знаний, включающий в нашей упрощенной модели множество утвер-

ждений, относящихся к понятию  $X$ , выделяемому СЗн в общей Картине Мира.

Каждому компоненту  $X$  сопоставимы три шкалы оценок:

Оценка  $\Pi$  объема знаний  $X$  от *максимальной* до *минимальной*; стоит отметить, что крайнее верхнее значение условно, поскольку не имеет в реальной СЗн буквального смысла и может относиться только к особым абстрактным сущностям, например, к Аксиомам; взять в качестве нижней оценки  $\Pi$  *полное отсутствие* было бы семантически некорректно, поскольку означало бы, что  $X$  не существует в СЗн.

*Совместимость*  $X$  (обозначим ее  $K$  – конфликт) с текущей СЗн может варьироваться от *полной совместимости* до *полной несовместимости*; первое предполагает, что включение  $X$  в СЗн ведет только к ее расширению и не требует редакции; остальные означают необходимость соответствующей этой оценке перестройки  $X$  и/или СЗн от «косметической» до радикальной, нужной для снятия конфликта.

*Достоверность*  $D$  источника знаний  $X$ : от *полная достоверность* до *полная недостоверность*.

Очевидно, что для  $K$  и для  $D$  крайние значения также определяются соответствующей СЗн условно. Предельно упрощая эти шкалы, ограничим каждую из них тремя значениями:

$\Pi$  : мах (*максимум*), 0 (*минимум*),  $>0$  (*нечто между первым и вторым*),

$K$  : мах (*полная несовместимость*), 0 (*полная совместимость*) и  $>0$  (*частичная совместимость*),

$D$  : мах (*максимум достоверности*), 0 (*максимум недостоверности*) и  $>0$  (*частичная достоверность*).

В ряде случаев для какой-то из этих оценок конкретное значение может быть любым и тогда ей будет приписываться значение *неопределенность*, т.е. «любое из конкретных».

**2.4.** Эта система оценок нужна далее нам для того, чтобы каждой содержательной зоне на разрабатываемой карте Пространства «Знание–Незнание» сопоставить характеризующую ее комбинацию оценок. Таким образом, наш аппарат ориентации включает по три округленных значения для выбранных трех измерений, однако семантика предложенных шкал ограничивает определяемое ими число комбинаций этих значений, поскольку:

при «минимальном» знании ( $\Pi=0$ ) оценки  $K$  и  $D$  компонента  $X$  обычно имеют неопределенные значения;

при  $K = \text{мах}$  для данного  $X$  его  $D$  может быть либо 0, либо неопределенно, в зависимости от того, как интерпретировать *полную несовместимость*.

В отдельных случаях нам придется ниже дополнять три наших грубых оценки более тонкими: *почти мах*, *почти 0* и  $\gg 0$ .

Конечно, в аппарат ориентации можно было бы включить и общую оценку *истинности* И знаний X, однако, четвертая оценка явно вывела бы нашу упрощенную систему шкал за рамки статьи. Фактически, оценка Д принимает на себя часть функций оценки И, а сама оценка И используется только для определения статуса истинности выделяемых Утверждений, относящихся к Метазнаниям рассматриваемых СЗн.

**2.5.** Ниже каждому *Утверждению* набора Метазнаний приписывается один из трех статусов, определяемых оценкой его Достоверности (а тем самым и Истинности):

*Аксиома*, для которой  $D = \text{мах}$ ,

*Презумпция* – оценка Д близка, но не равна максимуму, и

*Гипотеза* – оценка Д положительна, но достаточно далека от мах ( $\gg 0$ ).

Видно, что уже здесь для Метазнаний мы расширили шкалу Достоверности, включив в нее два новых значения: почти мах и  $\gg 0$ .

Введенный аппарат настолько упрощен, что в рамках рассматриваемой топологии Пространства оказывается достаточным лишь частично. Выбранный жанр «лубка» позволяет нам некоторые из комбинаций значений игнорировать, оставляя их на нашей карте белыми пятнами, а для выделения других учитывать (а) «оттенки» значений некоторых оценок и (б) угол зрения не только СЗн<sub>1</sub> но и СЗн<sub>2</sub>.

**2.6.** Обозначим принципы, определяющие далее структуру Пространства СЗн.

**А.** Разнообразие рассматриваемых Систем Знаний группируется в соответствии с распределением статусов (*Аксиома*, *Презумпция*, *Гипотеза*), сопоставляемых в конкретной СЗн обсуждаемым ниже Утверждениям некоторого базового набора, относящегося к ее Метазнаниям.

**Б.** Основной единицей деления каждой СЗн является *зона*. Как правило, основанием для отнесения некоторого компонента СЗн к той или иной зоне является комплекс трех описанных выше оценок: *полнота*, *совместимость* (*конфликт с СЗн*) и *достоверность*. Такая система оценок представляет собой своего рода «мета-паспорт» (обозначаемый далее М-П) каждой зоны СЗн.

**В.** Комплексная оценка зоны может определяться как самой СЗн<sub>1</sub> объекта, так и СЗн<sub>2</sub> субъекта, в связи с чем зона получает паспорт от той СЗн, которая приписывает ей оценки.

(Везде ниже знак  $\Leftrightarrow$  маркирует иллюстрирующие примеры).

Например, в СЗн<sub>1</sub> компонент X считается ложным или сомнительным, а в СЗн<sub>2</sub> – вполне очевидным и точно установленным, что харак-



терно в случае новых для  $CЗН_1$  знаний и гипотез. Или наоборот, Аксиома  $CЗН_1$  может потерять этот статус в  $CЗН_2$  и даже стать ложной.

### 3. АБСОЛЮТНОЕ ЗНАНИЕ И АБСОЛЮТНОЕ НЕЗНАНИЕ

Обсуждаемый аппарат ориентации будет опираться на две ключевые зоны – *Абсолютное Знание*  $Z_{abs}$  и *Абсолютное Незнание*  $H_{abs}$ , представляющие полюса оппозиции «Знание--Незнание».

**3.1.** Для нашей будущей карты подходят две противоположные метафоры:

Одна из метафор исходит из функций полюсов  $Z_{abs}$  и  $H_{abs}$ , к которым примыкают *Территории Знания и Незнания*, разделяемые в свою очередь на зоны. Такая топография напоминает глобус с его полушариями, между которыми на нашей карте находится *Территория Неопределенности* – подпространство, включающее зоны, которые не удастся отнести к двум основным территориям.

Другая метафора опирается на то, что функции зон  $Z_{abs}$  и  $H_{abs}$  на нашей карте различны: соответствующая топография Пространства похожа не на глобус, а на типичную карту плоской Ойкумены с центром в  $Z_{abs}$  и безбрежным океаном  $H_{abs}$  за ее границей.

$CЗн$  по определению состоит из компонентов знаний. Но и за краем Пространства, т.е. в зоне  $H_{abs}$ , находятся тоже компоненты знаний, хотя и «латентные», т.е. такие, о которых  $CЗн$  не имеет никакого представления, т.е. которые никаким образом не относятся ни к форме  $CЗн$ , ни к ее текущему содержанию.

При этом зона  $H_{abs}$  однородна, едина и не разлагается на упомянутые латентные компоненты, поскольку каждый из них неразличим в текущей Системе Знаний. Вместе с тем, *Абсолютное Незнание*  $H_{abs}$  каждой  $CЗн$  индивидуально, поскольку зоны  $H_{abs}$  в  $CЗН_1$  и  $CЗН_2$  в общем случае не совпадают:

До какого-то времени человечество не подозревало о существовании радиоволн, атомной энергии, Америки, микробов и многого другого, что, естественно, сказывалось не только на объеме, но и на самой организации Системы Знаний.

Справедливо и обратное: о многих компонентах  $CЗн$  предыдущих культур, воспринимавшихся нашими предками как привычная часть их Картины Мира, мы сегодня не знаем ничего; более того, мы ничего не знаем даже о самом существовании многих культур.

**3.2.** Хотя по определению о компонентах Абсолютного Незнания ничего не известно, кроме того, что они – латентные компоненты  $CЗн$ , мы знаем кое-что о самой  $H_{abs}$  (Знание о Незнании).

Знание о  $H_{abs}$  включает *основное* Утверждение вершины Метазнаний, которое я назову *Аксиомой А*, поскольку оно имеет такой статус во всех рассматриваемых в данной работе Системах Знаний:

Аксиома А: «Незнание» является необходимой составляющей самого понятия «Знание», т.е. любая СЗн включает оппозицию «Знание--Незнание».

Таким образом, Аксиома А является ключевым компонентом Метазнания, необходимо включаемым в саму концепцию СЗн. И поскольку она является необходимой составляющей любой СЗн, то это определяет и существование в каждой СЗн ее зоны  $H_{abs}$ , включающей все (и только такие) компоненты, которые не выделены в текущей СЗн, т.е. не воспринимаются ею как различимые сущности.

Из Аксиомы А следует вывод:

*Познаваемость Мира – т.е. Система Знаний – всегда ограничена.*

Если бы это было не так, то были бы возможны СЗн, в которых  $H_{abs}$  могла стать пустой, что противоречит Аксиоме А.

Другими словами, зона  $H_{abs}$  разделяется на две соответствующие составляющие: знания *познаваемые* и *непознаваемые*.

**3.3.** К Аксиоме А примыкает Презумпция  $A^1$ :

Презумпция  $A^1$ :  $H_{abs}$  содержит не только знания, но и *Метазнания*.

Презумпция  $A^1$  не является необходимым следствием Аксиомы А, поскольку примитивные СЗн могут предполагать, что все Метазнания уже известны, и  $H_{abs}$ , таким образом, включает только факты.

Однако для любой СЗн, превысившей этот барьер (а мы здесь рассматриваем только такие), Презумпция  $A^1$  становится Аксиомой  $A^1$ .

К этим двум Аксиомам добавим еще Презумпцию  $A^2$ :

Презумпция  $A^2$ : *Кроме Метазнаний,  $H_{abs}$  содержит и Мета-Метазнания* (Системы Знаний, нам неизвестные).

Таким образом, для тех СЗн, в которых Презумпция  $A^2$  имеет статус Аксиомы, содержание  $H_{abs}$  структурируется по двум измерениям:

по одному оно делится на три составляющие: *факты, Метазнания и Мета-Метазнания*;

по другому – знания в этих трех составляющих расслаиваются на *познаваемые* и *непознаваемые*.

**3.4.** В этом месте приходится обратить внимание на своеобразие взаимосвязи СЗн с Абсолютным Незнанием:

с одной стороны все составляющие  $H_{abs}$  в текущей СЗн пребывают латентно, поскольку в явном виде в ней отсутствуют;

однако, из этих латентных компонентов  $H_{abs}$  *познаваемые* могут появиться в СЗн в процессе ее расширения и эволюции (т.е. включены в нее и перейти в другие ее зоны), а *непознаваемые* остаются в  $H_{abs}$  данной СЗн навсегда;

эти обнаруженные компоненты могут быть совместимыми или несовместимы с текущей СЗн (оценка К), последнее может привести к ее значительной или даже радикальной перестройке.

Это своеобразие отражается соответствующими дополнениями наших Аксиом. В частности, они могут быть дополнены Утверждениями (а) и (б), развивающими Аксиому  $A^2$ :

Утверждения: *Неизвестные нам Системы Знаний*, т.е. такие СЗн в  $H_{abs}$ , к которым наша СЗн может получить прямой или косвенный доступ, могут оказаться несовместимыми по содержанию:

(а) с нашей текущей СЗн,

(б) с нашей СЗн даже при ее развитии, т.е. в принципе.

Трудно представить, как первобытный человек мог бы включить в свою СЗн современный компьютер. Тем не менее, сколь бы значительно не различались СЗн каменного века и начала 21-го, тем не менее вторая произошла от первой, хотя и по очень длинной цепочке трансформаций, значительная часть которых была настоящими революциями.

Мы можем абстрактно «макетировать» ощущения, выходящие за рамки нашей системы чувств, но проинтерпретировать их восприятие в ее рамках не представляется возможным. Еще менее вероятной кажется возможность совместить в одной СЗн наш мир и мир элементарной частицы, окажись она живым или, тем более, мыслящим существом.

**3.5.** Статус Утверждений (а) и (б) в рассматриваемых СЗн позволяет разделить их на:

*открытые:* готовые к «встрече» с качественно новыми знаниями, которые не только могут потребовать радикальной перестройки текущей СЗн, но и в принципе быть с ней несовместимыми, – Утверждение (б) в статусе Аксиомы;

*приземленные:* признающие, что развитие СЗн идет по спирали и радикальные перестройки в ней возможны, но принципиально несовместимые с ней знания вряд ли мыслимы, – Утверждение (а) в статусе Презумпции;

*консервативные:* развитие СЗн в основном достигнуто и никакие радикальные перестройки в ней невозможны или очень мало правдоподобны, – Утверждение (а) недостоверно.

К Аксиомам и Утверждениям (а) и (б) стоит добавить Утверждения  $A^3$  и  $A^4$ :

Утверждение  $A^3$ :  $H_{abs}$  не только несоизмеримо больше  $Z_{abs}$ , но и гораздо сложнее, поскольку содержит более развитые СЗн. Это Утверждения  $A^3$  следует из Утверждения (а).

Утверждение  $A^4$ : *Процесс расширения Картины Мира (Системы Знаний) потенциально бесконечен.*

Может показаться, что последнее Утверждение является лишним, поскольку следует из уже заявленных Аксиом. Но это не так: вполне удовлетворяет Аксиомам такая конечная СЗн, которая способна дос-

тичь своего потолка развития и остановиться, сохранив  $H_{abs}$  непознаваемой. Однако такие нам тут вряд ли интересны, так что для их исключения Утверждения  $A^3$  и  $A^4$  переводятся в статус Аксиом.

Таким образом, при этом расширенном наборе Аксиом наши знания (точнее, Метазнания) о  $H_{abs}$  могут быть дополнены еще одним Выводом:

Вывод:  $H_{abs}$  содержит бесконечное число компонентов.

Ту составляющую Метазнаний СЗн (тот набор ее Аксиом), которые не подлежат изменениям в процессе ее развития, будем называть базовой (базовым). Аксиомы  $A - A^4$  относятся к базовому набору.

**3.6.** В рассматриваемом Пространстве зоне  $H_{abs}$  противопоставляется зона Абсолютного Знания  $Z_{abs}$ , для всех компонентов которой оценки П и Д равны max, а К – нулю.

Таким образом, полнота любого компонента X зоны Абсолютного Знания является законченной (по крайней мере, для текущей СЗн) при полной достоверности X и отсутствии конфликтов как в  $Z_{abs}$ , а также значимых для СЗн конфликтов с компонентами других приближенных к  $Z_{abs}$  зон. Иначе говоря, зона  $Z_{abs}$  включает только те компоненты текущей СЗн, которые признаны в ней «вечными неоспоримыми истинами», не подлежащими (на данный момент) дополнению или пересмотру.

Как мы уже видели, для разных СЗн содержание зоны  $Z_{abs}$  может быть различным, хотя в любой СЗн эта зона включает составляющую *Абсолютное Метазнание*, которая состоит, как минимум, из Аксиомы А.

Как уже отмечалось, более «продвинутые» СЗн включает в зону  $Z_{abs}$  переведенные в статус Аксиом некоторые или все обсуждавшиеся выше Утверждения. Статус (*Аксиома, Презумпция, Гипотеза*) остальных относящихся к Метазнаниям Утверждений, в частности, и приведенных ниже, определяется типом конкретной СЗн и влияет на формирование карты ее Пространства.

Естественно, что к  $Z_{abs}$  относятся и все те компоненты СЗн, которые выводимы из Аксиом с максимальной достоверностью. Присутствие же в  $Z_{abs}$  нижней составляющей, т.е. фактов – знаний, не относящихся к Метазнаниям, определяется статусом Утверждения В:

Утверждение В: *Помимо Аксиом существуют компоненты знания (факты), полнота и достоверность которых оцениваются как максимальные, а уровень конфликтов с другими компонентами  $Z_{abs}$  равным 0, т.е. такие «абсолютные» факты, которые не подлежат расширению или пересмотру в рамках данной СЗн.*

М-П.: П=max, К=0, Д=Max.

Поскольку мы знаем массу «неоспоримых» фактов, которые рассматриваются в нашей СЗн\* как соответствующие Утверждению В, то его очевидность в СЗн\* в статусе Аксиомы (и, следовательно, су-

существования нижней составляющей зоны  $Z_{abs}$ ) не подлежит сомнению.

**3.7.** С другой стороны, динамика развития рассматриваемого Пространства предполагает справедливость Утверждения В<sup>1</sup>:

Утверждения В<sup>1</sup>: Для любого, кроме базовых, компонента  $X$  зоны  $Z_{abs}$  данной  $CЗ_n$  существует явно или потенциально другая  $CЗ_n$ , в которой этот  $X$  относится к другим зонам, в частности, зонам Территории Незнания.

Данное Утверждение относится не только к тем  $CЗ_n$ , которые являются для  $CЗ_n$  альтернативным, но и к  $CЗ_n$ , принадлежащим фазам «траектории» компонента  $X$  в процессе развития  $CЗ_n$ :

*восходящая* – предыстория  $CЗ_n$ , относящаяся к появлению  $X$  из  $H_{abs}$  с дальнейшим переходом в зоны, приближающее его к зоне  $Z_{abs}$  и включению в нее;

*нисходящая* – дальнейшее развитие  $CЗ_n$ , при которой абсолютный компонент  $X$  теряет статус и переходит в зоны, более удаленные от  $Z_{abs}$ ; это может происходить за счет дальнейшего развития  $CЗ_n$  по спирали или появления новой «альтернативной»  $CЗ_n$ , выбирающей другой комплекс Аксиом, в той или иной степени конфликтных с  $X$ .

В качестве иллюстрации можно вспомнить о бесчисленных переписываниях истории не столько за счет объективного учета новых фактов, сколько из-за смены метода их интерпретации, т.е. перестройки комплекса Аксиом. Наиболее радикальный пример, – это концепция Фоменко. То же неоднократно происходило за последний век в физике. И не только в ней.

Это еще раз подтверждает спорность возможности существования в «продвинутых»  $CЗ_n$  абсолютных (вечных) Аксиом и фактов, кроме ограниченного набора базовых, равно как и определений таких понятий как *Абсолютные Знание, Истина* и *Информация* вне конкретного контекста.

Это замечательно выразил Г.В.Ф.Гегель своим известным афоризмом: *Истина рождается как ересь, а умирает как предрассудок.*

**3.8.** Таким образом, отсюда следует, что отражаемое в  $Z_{abs}$  Знание о Незнании является той наиболее достоверной и фундаментальной частью  $CЗ_n$ , на которую опирается вся Система Знаний, следовательно, и наш аппарат ориентации. Более того, здесь мы обнаруживаем некоторую парадоксальность во взаимосвязи зон  $Z_{abs}$  и  $H_{abs}$ .

С одной стороны  $Z_{abs}$  оказывается первичной, поскольку именно ее компонент Аксиома  $A$  является отражением существования оппозиции «Знание--Незнание», а, следовательно, и зоны  $H_{abs}$ .

С другой стороны:

**а.** в силу той же Аксиомы А знание об Абсолютном Незнании оказывается таким же *необходимым* компонентом любой СЗн вообще и ее зоны  $Z_{abs}$  в частности;

**б.** существование Абсолютного Незнания с его структурой не вызывает сомнений, а вот представление об Абсолютном Знании остается неясным, поскольку к нему пока нам удалось отнести только несколько Аксиом;

**в.** в силу базовых Аксиом  $H_{abs}$  не только несравнимо больше  $Z_{abs}$  по объему, но и сложнее по структуре;

**г.** содержание и сложность всей СЗн и зоны  $Z_{abs}$  растут за счет заимствований из  $H_{abs}$  и перемещения компонентов из зон более близких к  $H_{abs}$  в зоны, приближающиеся к  $Z_{abs}$ .

Как мы видели выше, к этому можно добавить и тот факт, что с развитием Системы Знаний наблюдается «девальвация» Абсолютных Знаний и их «утечка» из  $Z_{abs}$ . Некоторые Аксиомы теряют свой статус, а целые области фактов, которые в простых СЗн считаются очевидными как Абсолютные, перестают быть таковыми и даже переходят в зоны, относящиеся к Территории Незнания, или вообще выбывают из СЗн, «возвращаясь» в  $H_{abs}$ .

СЗн предшествующих цивилизаций включали в  $Z_{abs}$  многие знания, относившиеся к религии, науке, социальным отношениям, этике и др. По ходу развития и перестройки Системы Знаний одни из них переводились в статус несовместимых ( $K=\max$ ) с новой текущей СЗн, другие теряли в оценке полноты и достоверности, третьи сохраняли статус Аксиом, но он ограничивался каким-либо локальным контекстом.

Выше уже упоминалось, что обсуждаемый далее аппарат ориентации различает выделяемые в Пространстве зоны по степени их близости к  $H_{abs}$  и  $Z_{abs}$ , что позволяет в зависимости от этого сгруппировать их условно в Территории Незнания, Знания и Неопределенности. Ниже структура этих территорий будет обсуждаться подробнее.

#### 4. ТЕРРИТОРИЯ НЕЗНАНИЯ

Рассмотрим несколько примеров зон Пространства, наиболее близких к  $H_{abs}$  и поэтому отнесенных здесь к Территории Незнания.

**4.1.** Все зоны Территории Незнания могут быть разделены на сферы «*Осознаваемого*» и «*Неосознаваемого*» Незнания. Первая рассматривается ниже в пп.4.2 – 4.5 достаточно подробно, ко второй отнесем зоны «*Объективного*» Незнания и «*Псевдознания*».

*Объективное Незнание* включает все те компоненты охватываемой СЗН<sub>2</sub> (т.е. содержащиеся в любой ее зоне, кроме  $H_{abs}$ ), которые недоступны вложенной СЗН<sub>1</sub>, т.е. относятся к ее зоне  $H_{abs}$ .

Другими словами, зона Объективного Незнания содержит те знания, которые в СЗН<sub>2</sub> включены, а в СЗН<sub>1</sub> отсутствуют.

В качестве иллюстрации можно привести сколько угодно примеров современных знаний, совершенно не известных даже несколько десятилетий назад, т.е. отсутствующих в Пространстве СЗН того времени.

*Псевдознание*: к этой зоне относятся все компоненты СЗН<sub>1</sub>, оцениваемые в ней как истинные, но являющиеся ложными в охватывающей СЗН<sub>2</sub>. Это относится как к знаниям прямого восприятия, так и к концептуальным:

Естественное явление, воспринимаемое как чудо или «знак свыше».

Люди произошли от Адама и Евы.

Очень грубо зону Псевдознания можно разделить на «*анти-Знание*» и «*эрзац-Знание*». К первому относятся компоненты, оценка которых как достоверных может иметь деструктивные последствия для самой СЗН<sub>1</sub> и/или для действий, предпринимаемых на ее основе. Компоненты второго представляют собой «шум»: использование их как достоверных не имеет однозначно положительных либо отрицательных последствий. Априорно отнести тот или иной компонент Псевдознания к одной из этих частей вряд ли возможно, поскольку их последствия в общем случае определяются конкретным контекстом:

Мираж или фата-моргана, принимаемые за реальность.

Солнце вращается вокруг земли.

Цель оправдывает средства.

**4.2.** Если зоны сферы «*Неосознаваемого*» *Незнания* строятся на карте СЗН<sub>1</sub> Системой Знаний СЗН<sub>2</sub>, то разделение на зоны сферы «*Осознаваемого*» *Незнания* определяется рамками самой СЗН<sub>1</sub> (т.е. без учета оценок СЗН<sub>2</sub>). В связи с этим естественно рассматривать карту СЗН<sub>1</sub> в СЗН<sub>2</sub> как две взаимосвязанные проекции Пространства: так, например, компоненты СЗН<sub>1</sub>, отнесенные СЗН<sub>2</sub> к Псевдознанию, могут самой СЗН<sub>1</sub> относиться к различным зонам, в том числе и к Территории Знания, в частности, даже к  $Z_{abs}$

Базовый в любой религиозной СЗН<sub>1</sub> принцип Бога всякой атеистической СЗН<sub>2</sub> будет отнесен к *анти-* или к *эрзац-Знанию*.

Сфера *Осознаваемого Незнания* может быть разделена в СЗН<sub>1</sub> на ряд зон, различающихся степенью удаленности от зоны Абсолютного Незнания  $H_{abs}$  и причинами отнесения их к Территории Незнания.

Для того, чтобы определить виды знания, которые могут относиться СЗН<sub>1</sub> к ее Территории Незнания, рассмотрим их в рамках нашего комплекса оценок.

*Полнота*. Кажется очевидным, что оценка компонента знаний  $X$  именно по этому показателю вряд ли значительно влияет на отнесение  $X$  к Территории Незнания СЗН, разве что в сочетании с другими двумя показателями.

*Совместимость* и *Достоверность*. Оба эти признака заведомо играют роль в позиционировании X, причем в тесной взаимосвязи: отнесение совместимого компонента к Территории Незнания может быть связано только с его низкой достоверностью; и наоборот, поставив под сомнение принадлежность к Территории Незнания максимально конфликтного X может только максимальная же оценка его достоверности. Таким образом, сочетание именно этих двух оценок и определяет зональную структуру Территории Незнания, которую естественно распределять по некоторому – тоже примитивному – функционалу *релевантности* отнесения к категории Знание.

Ниже мы ограничимся несколькими примерами зон Территории Незнания, дающими общее представление о ее структуре без излишней детализации и претензии на полноту.

**4.3.** После сказанного выше кажется очевидным, что ближайшая к  $N_{abs}$  зона, которую назовем «Абсурд», включает компоненты с низшей оценкой релевантности.

*К зоне «Абсурд» относится всякий компонент X, за которым  $S_{Зн}$  отрицается какая-либо информативность и сама возможность на включение в другие зоны при любых вариантах ее развития.*

*Другими словами: по форме X «похож» на знание, но по содержанию ни коим образом считается им не может.*

Этого не может быть, потому что не может быть никогда.

Такую оценку можно получить только при минимальной совместимости и достоверности, т.е. зона Абсурд характеризуется следующим образом:

М-П: П= неопределенный (любой), К =max и Д = 0.

*Ахинея*: текст, из которого не удается извлечь никакого смысла.

Очевидная сто процентная ложь

Аксиома племени А, что все его члены произошли от оленя.

*Я выиграл первый приз в лотерею без билета после завершения тиража.*

Кроме этого, несовместимость с  $S_{Зн}$  может предполагать абсурдную противоречивость самого X: например, каждая его составляющая совместима с  $S_{Зн}$  в отдельности, но они несовместимы в рамках всего X (например, a & не a).

Сама принадлежность X к зоне Абсурд актуальна только в рамках  $S_{Зн1}$ , – известна масса примеров утверждений, относившихся к Абсурду в недавнем прошлом и ставшие с тех пор достоверным знанием. А также vice versa. В частности, к ахинеи относится любое нераскрытое зашифрованное сообщение.

Итак, с зоной Абсурда в первом приближении понятно: отсутствие содержания, а также сочетание максимальной несовместимости с



СЗн и нулевой достоверностью источника однозначно определяет ближайшее положение к полюсу  $H_{abs}$ .

Понятно также (в частности, и из наших примеров), что сама зона Абсурд распадается содержательно на достаточно разные подклассы.

**4.4.** К следующей зоне «Невероятность» можно отнести компоненты, которые отстоят от  $H_{abs}$  несколько дальше.

Зона «Невероятность» включает такие компоненты  $X$ , чья совместимость с СЗн и достоверность источника оценены как близкие к минимальным, но не настолько, чтобы СЗн полностью отрицала возможность перехода  $X$  в более релевантные зоны, хотя возможность этого допускается только при таких вариантах развития, которые связанных с радикальной перестройкой СЗн.

М-П: П= неопределенный (любой), К почти мах и Д почти 0.

*предсказания, могущие оказаться истинными с ничтожной вероятностью – я выиграл первый приз в лотерею еще до проверки номера.*

*утверждение вопреки общепринятой теории, не подтверждаемое фактами.*

*На самом деле зона Невероятности неоднородна, поскольку зависит от интерпретации всех основных компонентов определения:*

*совместимость с СЗн – важно, какие именно составляющие  $X$  с какими конкретно составляющими СЗн несовместимы; например, из несовместимости ключевого элемента с Аксиомой А следует, даже при совместимости всего остального, все высказывание попадает в зону Абсурд.*

*достоверность источника – необходимо уточнить, какова достоверность самой этой оценки достоверности (такие оценки могут принципиально отличаться).*

*радикальная перестройка СЗн – радикальность перестройки СЗн также может различаться принципиально в зависимости от того, какая часть ее меняется и на что, к каким последствиям это ведет и т.п.*

*Очевидно, что более детальная и проработанная модель Пространства превратит зону «Невероятность» в сложную многомерную структуру.*

**4.5.** Взятая в качестве третьего примера зона «Ложного Знания» включает компоненты СЗн, низкая релевантность отнесения которых к категории Знание определяется минимальной оценкой Д их источника:

В зоне «Ложного Знания» совместимость компонента  $X$  оценивается как не нулевая, однако, в виду максимальной недостоверности источника, СЗн отрицает возможность перехода  $X$  в более релевант-

ные зоны; возможность такого перехода допускается только при радикальной переоценке достоверности источника.

М-П:  $P > 0$ ,  $K > 0$ ,  $D = 0$ .

Конечно, низкая оценка достоверности  $X$  может определяться целым спектром разных причин, например:

**а.** Ложная информация на основе прямого восприятия:

(Осознаваемый) мираж или фата-моргана.

Распознанный ложный ход противника.

**б.** СЗн относит источник  $X$  к классу, достоверность которого очень низка.

Мы помним, что существование Трои отрицалось, поскольку источники информации о ней относились к чистой мифологии.

**в.**  $X$  противоречит  $X$ », достоверность источника которого оценивается выше.

Оценка показаний свидетелей на суде.

**г.**  $X$  противоречит другому  $X$ », помещенному СЗн в более релевантные зоны.

Многочисленные сообщения о дате нападения Германии на СССР отвергались ставкой поскольку противоречили ее собственным выводам.

Эти примеры, спектр которых можно расширить, показывают, что зона Ложного Знания, как и две предыдущие, неоднородна и должна стать в будущем объектом более детальной топологии.

Будем надеяться, что, несмотря на очевидную грубость используемого аппарата ориентации, рассмотренные примеры зон Территории Незнания дали о ней некоторое общее представление.

## 5. ТЕРРИТОРИЯ ЗНАНИЯ

Как уже говорилось выше, в Пространстве выделяются несколько зон, в той или иной степени близких к  $Z_{abs}$  и отнесенных поэтому к *Территории Знания*.

**5.1.** Начнем с вариантов дополнения самой зоны  $Z_{abs}$ .

Наиболее примитивный из них включает в качестве Аксиомы Утверждение  $V^*$ :

Утверждение  $V^*$ : *Вне текущей зоны  $Z_{abs}$  и ее прямых следствий нет компонентов, могущих быть отнесенными к ней в процессе дальнейшего развития.*

Такое Утверждение относительно зоны  $Z_{abs}$  характеризует достаточно **примитивную** СЗн, разделяющую *argioi* все знания на абсолютные и «прочие»:

с одной стороны, эта СЗн совершенно ригидна: любой  $X$ , хотя бы частично несовместный с абсолютными, оценивается в ней как эрзац, либо анти- знание;

с другой, в развитии остальных ее компонентов отсутствует перспектива достижения уровня  $Z_{abs}$  (что возможно в СЗн, исключая Утверждение В\*).

Такая СЗн соответствует обскурантизму и консервативной доктрине. Будем считать, что на интересующей нас части Пространства достоверность Утверждения В\* отрицается, и поэтому далее оно не упоминается.

Более гибкие СЗн допускают два варианта формирования  $Z_{abs}$ , при которых изменение  $Z_{abs}$  может происходить за счет:

только расширения, при котором все компоненты  $Z_{abs}$  сохраняют свой статус или

возможности корректировки, через редакцию множества компонентов  $Z_{abs}$ .

Первый вариант соответствует такому упрощенному типу СЗн, при котором текущее содержание зоны  $Z_{abs}$  остается в СЗн навсегда в качестве «вечных основ» и может лишь пополняться за счет совместимых с ним компонентов из других зон (в частности, и из  $N_{abs}$ ).

Второй вариант соответствует «нормальным» Системам Знаний, допускающим адаптацию текущей СЗн к новым знаниям, достаточно убедительным для того, чтобы поставить под вопрос статус Аксиом и/или абсолютных фактов, не относящихся к базовым.

Именно такими «нормальными» СЗн будет ограничено наше рассмотрение дальше.

**5.2.** Территорию Знания можно разделить на две основные сферы:

«Устоявшееся» Знание, вложенное в привычную сложившуюся СЗн и объединяющее пять зон, которые будут рассмотрены ниже в пп.5.3 – 5.7, и

«Спорное» Знание.

Компоненты Спорного Знания не включены в текущую сферу Устоявшегося Знания, но радикально не противоречит ей, т.е.:

Компонент X в сфере Спорного Знания характеризуется достаточной достоверностью, причем противоречие X со сложившейся СЗн не оценивается как неразрешимое.

М-П.:  $P > 0$ ,  $K > 0$ ,  $D > 0$ .

Достаточно правдоподобные, но не окончательно закрепившиеся гипотезы (например, из-за наличия альтернативных неустоявшихся или спорных гипотез).

Сфера Спорного Знания является буферной между Устоявшимся Знанием и Территорией Неопределенности: ее компоненты имеют неплохие шансы продвинуться к  $Z_{abs}$ , подняв свой уровень релевантности. Самые смелые гипотезы обычно зарождаются в зонах Территории Незнания и/или в сфере Спорного, переходя в процессе их подтверждения и развития в зоны Устоявшегося Знания.

**5.3.** Наиболее близкая к  $Z_{abs}$  зона «*Очевидного Знания*» ( $Z_1$ ) определяется следующим образом:

В зоне *Очевидного Знания* полнота  $X$ , его совместимость и достоверность оцениваются как близкие к максимуму; при этом текущее представление о  $X$  допускает расширение, но не редакцию и не снижение оценок совместимости и достоверности.

К зоне *Очевидного Знания* относятся компоненты, максимально близкие к переходу в статус Аксиомы и к включению в  $Z_{abs}$ . Их нахождение в  $Z_1$  может быть связано только с отсутствием некоторых второстепенных деталей, так что именно эта неполнота мешает им получить максимальные оценки совместимости и достоверности.

М-П: П почти мах, К почти 0, Д почти мах.

Презумпции (п.2.5)

Устоявшиеся теории, ставшие настолько привычными, что получают в общественном сознании статус Аксиом, хотя ими не являются.

Последний пример показывает, что «бытовые» СЗн не различают зоны  $Z_{abs}$  и  $Z_1$ . Более того, чем грубее и примитивней СЗн, тем меньше в ней зон и тем более размыты их границы.

**5.4.** Ближайшая к Очевидному Знанию зона *Знания2* ( $Z_2$ ) представляет те компоненты  $X$ , которые определяются менее жестким критерием, чем для  $Z_1$ :

Полнота  $X$  зоны *Знания2* и его достоверность оцениваются как близкие к максимуму, а его совместимость как достаточно высокая; текущее представление о  $X$  допускает его расширение и редакцию деталей, и не предполагает снижение оценок совместимости и достоверности.

Соответственно, паспорт зоны  $Z_2$  отличается от зон  $Z_{abs}$  и  $Z_1$  оценкой П и К:

М-П: П почти мах,  $K > 0$ ,  $D \gg 0$ .

Зона  $Z_2$  охватывает компоненты, близкие к переходу в зону Очевидного Знания, для чего требуется уточнение некоторых деталей, которые могут оказаться частично несовместимыми с другими второстепенными компонентами Территории Знания, имеющими высокие оценки совместимости и достоверности, т.е. требовать «косметической» адаптации  $X$  к СЗн, или их редакции для согласования.

отчет готов, осталось уточнить отдельные чисто технические пункты.

при расчетах в этом случае говорят: *допустимое округление решения*.

Вероятность, что «косметическая» адаптация  $X$  может закончиться не продвижением  $X$  в  $Z_1$ , а «ссылкой» в зоны более удаленные, невелика, но более вероятна, чем для компонентов зоны Очевидного Знания.

*округление решения* может оказаться недостаточно точным и потребовать корректирующего перерасчета.

**5.5.** Следующая зона Знания<sup>3</sup> ( $Z_3$ ) включает, естественно, те компоненты  $X$ , которые подчиняются еще менее жесткому и более широкому критерию, чем рассмотренные выше:

*Х в зоне Знания<sup>3</sup> данной СЗН занимает устойчивое, не вызывающее сомнений место, при отсутствии уверенности в максимальной полноте, непротиворечивости и \ или достоверности X; при этом X может допускать не только расширение и уточнение, но, возможно, и частичный пересмотр, при котором СЗН с высокой вероятностью не меняет основной части знаний X.*

М-П.:  $P \gg 0, K > 0, D > 0$ .

Я уверен, что Петров – человек порядочный. А что касается этих мелочей, то все мы не без греха. При расчетах определение *решение с точностью до такого-то знака* предполагает, что эта точность дает требуемое представление о результате; однако иногда задача может потребовать получение более высокой точности, которая с небольшой вероятностью, но может привести к некоторой корректировке принятой постановки задачи и/или метода ее решения (не меняя их в целом).

**5.6.** Естественным следующим шагом является выделение зоны «Привычного» Знания ( $Z_4$ ), которая занимает одно из центральных мест в Системах Знаний, так как ее компоненты  $X$  относятся к наиболее широкому и естественному типу устоявшихся знаний, составляющих костяк любой СЗН:

*Компонент Привычного Знания X в данной СЗН занимает устойчивое место, предполагающее высокую уверенность в достаточных полноте, непротиворечивости и \ или достоверности X; при этом X может подлежать не только расширению и уточнению, но, возможно, и пересмотру, при котором СЗН с достаточной вероятностью не меняет основной части знаний X.*

Некоторые компоненты этого определения выделены для указания на различия критериев для зон  $Z_3$  и  $Z_4$ , хотя упрощенность нашей системы оценок не позволяет их различать :

М-П.:  $P \gg 0, K > 0, D > 0$ .

Я считаю (почти уверен), что Петров – человек порядочный, хотя не мешало бы это еще раз проверить.

Как правило, туда можно доехать за полчаса.

**5.7.** Подведем итог обсуждению рассмотренной выше структуры Территории Знания, в которой выделено шесть зон. Пять из них были отнесены к сфере Устоявшегося Знания и включали:

зону Абсолютного Знания,

четыре зоны, к которым относятся категории знания, последовательно удаляющиеся от  $Z_{abs}$  по своей *полноте, непротиворечивости* и \ или *достоверности*, однако не выходящие в этих оценках за рамки, позволяющие относить эти знания к достаточно полным, устойчивым и надежным.

За границами Устоявшегося Знания следует зона Спорного Знания, содержание которой, хотя и претендует на включение в сферу Устоявшегося, но не настолько адаптировалось к СЗн, чтобы достичь необходимого для этого уровня базовых оценок по полноте, непротиворечивости и \ или достоверности.

Остается открытым вопрос о том, что кроме упоминавшихся Аксиом может входить в базовую – стационарную – часть Абсолютного Знания?

Представляется, что на текущем наивном уровне рассмотрения найти такие стационарные составляющие Метазнания в неузнаваемо усложняющейся Картинае Мира непросто: достаточно сравнить миры первобытного человека, жителя эллинской эпохи и разноразличной Систем Знаний нашего времени.

В пространстве современной Картины Мира лишь небольшой фрагмент занимают знания двухсот – трехсотлетней давности, основная часть которых быстро тает, навсегда уходя в историю. Трудно даже вообразить, какое место займет текущая, да и вся наша современная Картина Мира в общечеловеческой Системе Знаний через сто или даже пятьдесят лет.

**5.8.** В завершение обзора структуры Территории Знания стоит обратить внимание на тот факт, что рассмотренные сферы относились к знаниям, *осознаваемым* самой СЗн<sub>1</sub> и ею отнесенным к перечисленным зонам. Однако они естественным образом дополняются *внешней* сферой Территории Знания, которая включает все те компоненты, которые могут быть определены как относящиеся к этой территории внешней (эталонной) СЗн<sub>2</sub>, но в изучаемой СЗн<sub>1</sub> вообще отсутствуют или находятся вне этой территории. Например:

Уже сделанные опережающие открытия в восприятии рядовых современников, которые о них вообще не знают или их не понимают и\или не принимают.

В частности, к этой сфере относится «*Непризнанное*» Знание, для которого самооценка (оценка СЗн<sub>1</sub>)  $D = 0$ , при объективной оценке (оценка СЗн<sub>2</sub>)  $D = \max$ .

## 6. ТЕРРИТОРИЯ НЕОПРЕДЕЛЕННОСТИ

Обсуждение Территории Неопределенности ограничим несколькими примерами зон, не отнесенных к Территориям Знания и Незнания.

**6.1.** Зона «*Туманов*» включает компоненты, которым сопоставлены знания, само право на существование которых остается в СЗн неопределенным, что характеризуется следующим образом:

*Метазнания (Теория) позволяют выделить концептуальный компонент X, но его содержание и достоверность оцениваются как неопределенные:*

М-П::  $P > 0$ , К и Д равны неопределенности.

Возможность машины времени.

Иные, равные нам или превосходящие нас, формы разумной жизни.

Для данной зоны ключевым является полная неопределенность реального существования X.

**6.2.** К зоне «*Загадок*» отнесем компоненты, существование которых представляется достаточно достоверным, но содержательная информация о его сути отсутствует:

Компонент X зоны «*Загадок*» существует с максимальной достоверностью; при этом объем достоверных знаний в X равен 0, что означает, что совместимость X с СЗн может быть любой.

М-П:: П почти 0, К и Д равны неопределенности.

Приведем несколько примеров различных секторов зоны Загадок, различающихся источником информации и совместимостью X с СЗн:

**а.** «*Ненаблюдаемые*», но претендующие на достаточную достоверность, выводы из полученной информации и/или сложившейся Картины Мира:

М-П::  $P = \text{мин}$ , К = неопределенность,  $D > 0$ .

Античный мир понимал, что за Атлантическим океаном должно что-то быть, но что это (другая земля, край света или что-то еще) было неизвестно.

**б.** «*Воспринятые*» прямо, но необъяснимые, т.е. не вписывающиеся в сложившуюся СЗн, явления (ясно, что объяснение должно быть, но оно неизвестно):

М-П::  $P > 0$ , К и Д не определены.

Номера хорошего иллюзиониста для непрофессионального зрителя, знающего, что ему показывают фокусы.

Знакомство с паровозом для человека из первобытного племени.

Очевидно, что секторов в зоне *Загадок* больше, чем приведенные (а) и (б), однако для более детальной структуризации этой зоны требуется более разработанная топография.

**6.3.** Зона «Домыслы» включает компоненты СЗн, существование которых, как и в предыдущей зоне Загадок, представляется достаточно достоверным, однако между этими зонами есть и различия:

В зоне «Домыслы» компонент  $X$  существует с максимальной достоверностью при наличии достаточного объема знаний, однако их достоверность остается неопределенной, так как достоверность источника почти минимальна; т.о. совместимость  $X$  с СЗн может быть любой.

М-П.:  $P > 0$ ,  $K =$  неопределенность,  $D$  близка к 0.

Так же как и зона Загадок, зона Домыслы может быть разделена на несколько секторов, различающихся источником информации и совместимостью  $X$  с СЗн. Приведем несколько примеров:

**а.** «*Воспринятые косвенно 1*»:  $X$  включает достаточный объем сообщений, достоверность которых считается в СЗн неопределенной и/или низкой, поскольку информация о  $X$  противоречива и сами ее источники недостоверны; совместимость  $X$  с СЗн<sub>1</sub> не определена.

М-П.:  $P > 0$ ,  $K$  и  $D$  не определена.

Любой поток сообщений в СМИ на первых этапах сенсационного события, когда факты противоречивы и еще не уточнены, а опыт показывает, что, хотя для этого потока был повод, сами сообщения могут соответствовать реальности в очень незначительной степени.

**б.** «*Воспринятые косвенно 2*»: информация  $X$ , поддерживаемая достаточным объемом сообщений, источники которые считаются в СЗн недостоверными, в основном потому, что  $X$  выходит за ее рамки.

М-П.:  $P >> 0$ ,  $K > 0$ ,  $D$  не определена.

Гипотезы относительно явлений, стабильно наблюдаемых, но отличных от уже изученных: в частности, все паранормальные явления – полтергейст, НЛЮ, Бермудский треугольник и т.п.

**в.** «*Воспринятые косвенно 3*»: информация, поддерживаемая достаточным объемом сообщений из источника, которой считается заведомо дезориентирующим, но частично содержащим достоверную часть для «отвода глаз».

М-П.:  $P > 0$ ,  $K$  не определена,  $D = 0$ .

Дезинформация от противника.

**г.** «*Выводы*» из имеющихся недостоверных, противоречивых и/или неполных знаний:

Оправдан за недостатком улик.

**д.** «*Прогнозы*», не слишком достоверные, поскольку сделаны на недостаточно полных данных:

Любые недостаточно обоснованные предположения относительно нашей цивилизации через 100 или 1000 лет.

Выбор, имеющий 50% вероятности.



**6.4.** Даже на примере уже рассмотренных выше зон Территории Неопределенности очередной раз очевидна грубость используемой системы. Например: в случае демонстрации фокусов у компонента  $X$  есть как бы две составляющих  $\Pi$ :

«внешняя»: то, что воспринимается прямо и квалифицируется зрителем как заведомо ложная ( $\Pi=\max$ ), и

«настоящая»: обычно отсутствующее объяснение  $X$  ( $\Pi$  почти 0).

Эту ситуацию можно отнести к категории *Я знаю, что я не знаю о  $X$ , несмотря на то, что внешне кажется, что знаю все или почти все.*

В случае с хорошим фокусом *внешняя составляющая* определяет информацию о  $X$  как достаточно полную, (внешне) истинную и (внешне) достоверную, т.е. явно относящуюся к Территории Знания. *А настоящая составляющая*, формируемая ресурсами СЗн, утверждает, что подлинной, истинной и достоверной информации о  $X$  практически нет, поскольку технология фокуса для зрителя необъяснима.

Этот слоеный пирог может быть многоэтажным, – например, включающим три составляющих:

«внешняя»: то, что воспринимается прямо и заведомо ложно,

«псевдо – настоящая»: то, что объясняет воспринимаемое убедительно, но неправильно, и

«настоящая»: то, что объясняет воспринимаемое правильно.

Поскольку «настоящая» – правильная – информация в общем случае предполагает Истину, а та, как уже упоминалась, является контекстно-зависимой, то такой пирог начинает расплзаться не только в глубину, но и в ширину. Например, как правило, минимум две несовместимые или плохо совместимые истины фигурируют на суде (обвинитель и защитник), причем часто они обе могут претендовать по сути на функции «настоящих» и/или быть «псевдо – настоящими».

Эта дополнительная многомерность еще раз подтверждает правильность нашего решения ограничиться максимально упрощенным паспортом зоны. В противном случае увидеть пусть и грубую, но достаточно наглядную карту Пространства «Знание–Незнание» в границах данной статьи нам заведомо бы не удалось.

Будем считать, что приведенные в этом разделе примеры зон Территории Неопределенности, позволили как-то отразить ее специфику, что дает нам право завершить на этом эксперимент по наивной топографии исследовавшегося нами Пространства и перейти к подведению итогов.

## 7. ЗАКЛЮЧЕНИЕ

Задача этой работы состояла в попытке черного наброска понятийной структуры, названной в начале статьи Пространством оппозиции «Знание--Незнание».

В любой Системе Знаний такая структура эксплицитно и имплицитно присутствует, обладая своей топографией, которая позволяет как-то ориентироваться в СЗн на основе оценки ее компонентов в отношении полноты, достоверности и совместимости с самой СЗн.

Мы разделили весьма условно всю область исследования на три основные части: Территории Знания, Незнания и Неопределенности. Третья разграничивает первые две из них, которые концентрируются вокруг соответствующих полюсов Абсолютного Знания и Абсолютного Незнания и разделяются на зоны, различающиеся в основном удаленностью от этих полюсов.

Структура Пространства оппозиции «Знание--Незнание» в конкретной СЗн<sub>1</sub> определяется несколькими факторами:

Набором относящихся к Метазнаниям Утверждений и выбором их статуса, при котором часть Утверждений становятся Аксиомами, другие, напротив, отвергаются, третьим приписывается определенная степень достоверности. Именно эта система базовых Аксиом определяет для Системы Знаний ее собственное место в Пространстве «Знание--Незнание» и взаимосвязь в нем ее полюсов. А, следовательно, ее способность к развитию по спирали, т.е. к периодической радикальной перестройке  $Z_{abs}$  за счет восприятия новых Метазнаний из  $H_{abs}$ .

Системой координат, в которой каждой зоне сопоставляется ее паспорт, – набор огрубленных оценок *полноты*, *совместимости* с СЗн и *достоверности*, характеризующих компоненты знаний, относимые к данной зоне. Механизмы определения этих оценок и их взаимодействия, рассмотрение которых не уместилось в рамки статьи, регулируют динамику СЗн, т.е. баланс ее консервативной составляющей и способность к восприятию нового знания.

Соотношением оценок знаний СЗн<sub>1</sub> как ею самой, так и внешней СЗн<sub>2</sub>.

Основанная на таком примитивном аппарате Наивная Топология позволила выделить в СЗн более двадцати зон и потенциальных секторов, выполнив тем самым стоявшую перед нами задачу:

(а) сделать первый шаг в направлении до-формального исследования интересующего нас Пространства, и

(б) послужила в качестве черного наброска структуры этого Пространства, обеспечив определенный опыт для разработки в будущем более адекватной его карты.

При выполнении этой задачи мы очередной раз убедились, что всякое погружение в проблему «настоящего» знания подтверждает

вечную справедливость формулы Сократа *я знаю, что ничего не знаю*, которая в прямой интерпретации может быть перефразирована в нашей наивной Топографии как: *единственное абсолютное знание – это существование Незнания (Аксиома А), которое несравнимо больше, чем постоянно меняющаяся Территория Знания (Утверждение А<sup>3</sup>)*.

Можно обратить внимание на еще один из ее конкретных смыслов: *ничего «истинного» о том, что именно я знаю, выяснить невозможно* [2].

Таким образом, существование, хотя бы потенциальное, *идеальной Системы Знаний*, пусть и не во всем полной, но заведомо непротиворечивой и достоверной, остается под вопросом в принципе, поскольку такая СЗн возможна если ограничиться рассмотрением чисто формальных абстрактных теорий. Так что нам остается удовлетвориться уровнем текущего – полученного нами на данном этапе – достаточного контекстно-зависимого уровня взаимопонимания и согласия.

Тема этой статьи является лишь одной из составляющих общего проблемного поля взаимосвязи Знания и Незнания. Несколько лет назад она получила самостоятельность, отделившись от другой работой [2]. Надеюсь, что мне удастся доработать ее до достаточной полноты и порядка с тем, чтобы связать с остальными составляющими этого проблемного поля, в частности, с работами, посвященными тематике НЕ-факторов [3].

#### ЛИТЕРАТУРА

1. А.С. Нариньяни. Между Знанием и Незнанием – Наивная Топография. В сб. «Межд. Семинар ДИАЛОГ»2000» Протвино, Май 31-июнь 4, 2000, т.1, стр. 199 – 208.
2. А.С. Нариньяни. Знание, как особая форма Незнания (Что я знаю о том, что ничего не знаю), <http://www.futurerussia.ru/build/ch1/znan22.htm>
3. А.С. Нариньяни. НЕ-факторы 2004 // Труды 9-й национальной конференции по искусственному интеллекту с международным участием – КИИ»2004. –Москва: Физматлит, 2004. –Т.1. –С.420-432.

---

## II

# СОЗНАНИЕ, МОЗГ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

---

---

## СОЗНАНИЕ, МОЗГ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

*Д.И. Дубровский*

Разработка проблемы искусственного интеллекта (ИИ) необходимо связана с результатами исследования естественного интеллекта (ЕИ), который не ограничивается когнитивными функциями, представляет со-знательную деятельность в целом. Здесь мы имеем дело с тем, что обычно именуют проблемой сознания. Она – многопланова, ее анализ, помимо уточнения терминологии, предполагает теоретически корректное вычленение основных планов и последующее их сопоставление друг с другом. Не вдаваясь в эту задачу, мы выделим лишь один из основных планов этой проблемы – классический вопрос об отношении сознания к головному мозгу. Исследование этого вопроса способно расширить и углубить наше понимание специфики информационных процессов, протекающих в головном мозге, и тем самым стимулировать новые подходы в разработке ИИ.

Сознание обладает специфическим и неотъемлемым качеством *субъективной реальности* (СР). Мои мысли, мои ощущения зеленого, красного, запаха розы, не говоря уже о переживаниях боли или радости – всё это не менее реально и значимо для меня, чем стол, за которым я сейчас сижу. Но и сама реальность стола удостоверяется для меня моими психическими, осознаваемыми отображениями данного объекта. СР – это динамический континуум осознаваемых состояний человека, временно прерываемый глубоким сном или случаями потери сознания. Этот континуум в значительной степени центрирован нашим «Я», которое отдает себе отчет о конкретных явлениях СР и способно управлять некоторыми из них. Качество СР вы-

стует в различных формах (ощущение, образ, эмоция, мысль, чувство уверенности, волевое усилие и т.п.), оно способно выражать самое разнообразное содержание, но суть его – в субъективном переживании, в его как бы непосредственной данности индивиду (в аналитической философии это качество выражается разными терминами: «субъективный опыт», «ментальное», «квалиа» и др.).

Именно качество СР создает главные трудности для объяснения связи сознания с мозговыми процессами (и более широко – при попытках интегрировать феномены сознания в научную картину мира). В западной аналитической философии (там эти вопросы занимают центральное место и им посвящена колоссальная по объему литература) преобладает *редукционистский тип объяснения* в двух его основных вариантах: *физикалистском* (когда явления СР редуцируются к физическим процессам) и *функционалистском* (когда они редуцируются к функциональным отношениям). Сравнительно немногочисленные противники редукционизма (Т. Нагель, Дж. Серл и др.), высказывая убедительные критические соображения, не предлагают, однако, концептуального решения проблемы «сознание и мозг» (см.: 5, 6).

Такое решение должно представлять собой теоретически корректный ответ, по крайней мере, на два следующих вопроса: **I. Как связаны явления СР с мозговыми процессами, если первым нельзя приписывать пространственные и другие физические свойства, а вторые ими по необходимости обладают?** **II. Каким образом явления СР, которым нельзя приписывать физические свойства (массу, энергию), способны служить причиной телесных изменений, управлять ими?** С ними связан ряд других существенных вопросов (**III**), которые имеют тесное отношение к проблематике ИИ (о них речь пойдет ниже).

Для ответа на указанные два основных вопроса мною предлагается информационный подход (см. подробнее: 2, 3, 4). Его исходные посылки следующие:

(1) Информация необходимо воплощена в своем физическом (материальном) носителе, она существует лишь в определенной кодовой форме;

(2) Информация инвариантна по отношению к физическим свойствам своего носителя, т.е. одна и та же информация может кодироваться по-разному (сокращенно это именуется далее «*принципом инвариантности*») – ПИ);

(3) В самоорганизующихся системах информация может выступать причиной изменений в ее субстрате и структуре, служить фактором управления;

(4) Явления СР допустимо интерпретировать в качестве информации (например, мое восприятие дерева в данный момент, как явление моей СР, есть информация о соответствующем внешнем объекте). Если эти посылки (после критического анализа моими оппо-

нентами) всё же принимаются, то из них могут быть выведены искомые объяснения.

**I. Явление СР связано с соответствующим мозговым процессом как информация со своим носителем.** Последний представляет собой определенную мозговую нейродинамическую систему. Явлениям СР действительно нельзя приписывать длины, ширины и т.п. Говорить о пространственной локализации явления СР можно лишь в том смысле, что оно воплощено в определенном нейродинамическом коде и вне его не существует, а последний имеет сложную структуру и определенное расположение в головном мозге.

**I-а.** Связь явления СР со своим нейродинамическим носителем является *функциональной*, она представляет сложившуюся кодовую зависимость. Рассмотрим сравнительно простой случай. Переживаемый мной в данном интервале образ дерева (обозначим это явление СР через **О**) имеет своим носителем определенную нейродинамическую систему (обозначим ее **Х**). Связь между **О** и **Х** носит именно функциональный характер – это явления одновременные и однопричинные: **Х** есть кодовая представленность **О** или, короче, – код **О**. Основательное исследование подобных связей предполагает расшифровку кода. Задача расшифровки мозговых кодов психических явлений уже поставлена на повестку дня (вслед за расшифровкой генетического кода и генома человека).

**I-б.** Но что означает **расшифровка кода**, если информация всегда существует только в кодовой форме и от нее невозможно избавиться? Она может означать лишь одно: перевод неизвестного кода в известный. Для каждой самоорганизующейся системы существует два типа кодов. Назовем их *«естественными»* и *«чуждыми»*. Первые непосредственно «понятны» той системе, которой они адресованы, «прозрачны» для нее, не требуют операции декодирования (частотно-импульсный код на выходе сетчатки сразу «понятен» соответствующим мозговым структурам, слово «дерево» сразу понятно человеку хорошо знающему русский язык, ему не нужно специально анализировать физические и структурные свойства этого кодового объекта и т.п.). Декодирование требуется, когда система имеет дело с «чуждым» кодом, но оно означает лишь преобразование его в «естественный» код. После того как найден и закреплён способ такого преобразования «чуждый» код становится для самоорганизующейся системы «естественным», что знаменует акт ее развития.

**I-в.** Важно учитывать, что и «естественные» и «чуждые» коды могут быть для самоорганизующейся системы как *внешними* (например, сообщаемые человеку непонятные слова), так и *внутренними* (скажем, мозговые коды типа **Х**), что обуславливает специфику задач, связанных с их декодированием и перекодированием. Можно выделить **два вида задач расшифровки кода**: 1) «прямую», когда дан

кодовый объект и требуется выяснить информацию, которая в нем содержится (здесь мы имеем дело с «чуждым» кодом), и 2) «обратную», когда нам дана определенная информация и требуется установить ее носитель и его кодовую организацию; здесь перед нами «естественный» код и такая задача является, как правило, более трудной (см. подробнее: 3, гл. 5, с. 233 – 250).

**Г-г.** Мозговые коды типа **Х** являются внутренними «естественными» кодами. Воплощенная в них информация дана индивиду непосредственно в форме явлений его СР (чувственных образов, мыслей и т.п.). Причем, не только устройство такого кода, но даже наличие его в нашем мозгу нами совершенно не ощущается, не отображается. В явлениях СР нам дана информация как бы в «чистом» виде и способность оперировать ею. Таков кардинальный факт нашей психической организации, сложившейся в процессе биологической эволюции и антропогенеза, ибо живой системе для эффективного функционирования нужна информация как результат адекватного отображения внешних объектов, ситуаций, собственных действий и т.д. и в большинстве случаев не нужно отображение носителя информации (в силу принципа инвариантности – (2)).

У человека же на нынешнем этапе развития общества такая потребность возникает. Расшифровка мозговых кодов явлений СР – реальная научная задача (хот она является весьма трудной в силу того, что принадлежит к типу «обратных» задач). Ее решение способно вызвать судьбоносные для земной цивилизации последствия как позитивного, так и негативного характера (что особенно важно иметь в виду). Если мозговые коды явлений СР будут хотя бы частично расшифрованы, то это затронет фундаментальный принцип социальной самоорганизации – относительную «закрытость» внутреннего, субъективного мира личности. Будет нарушено ее неотъемлемое право: «открывать» другому свои мысли, намерения, чувства и т. п. по своей воле. Кто, как, зачем станет «открывать» другого? Каковы могут быть социальные и иные последствия этого? Подобные вопросы требуют тщательного рассмотрения как в теоретическом, так и в практическом плане (здесь необходимо специальное исследование).

**П. Явления СР способны служить причиной телесных изменений, способны управлять ими в качестве информационной причины. Психическая причинность есть вид информационной причинности.**

**П-а.** Отличие информационной причины от физической причины определяется принципом инвариантности (причинный эффект вызывается тут именно информацией, на основе сложившейся кодовой зависимости, а не самими по себе физическими свойствами носителя этой информации, которые, в принципе, могут быть разными). Всякое психическое причинение осуществляется в сознательно-бессоз-

нательном контуре информационных процессов, который представляет собой исключительно сложный объект для анализа. Трудности усугубляются еще и тем, что даже в случае ясно сознаваемого действия, необходимо учитывать в нем не только рефлексивное и актуальное, но также уровни арефлексивного и диспозиционального. К тому же возникает проблема речевого оформления явлений СР, языкового кода.

Однако в первом приближении правомерно все же выделить такую разновидность психической причинности как **произвольное действие**. Здесь можно сравнительно четко обозначить комплекс явлений СР, выражающих мое намерение совершить определенное действие и управляющих его реализацией. Возьмем простой пример. Я хочу включить свет настольной лампы и делаю это, нажимая кнопку. В данном случае мое желание, побуждение формирует программу действий и запускает цепь кодовых преобразований, хорошо отработанных в филогенезе и онтогенезе (имеется в виду последовательное и параллельное включение кодовых программ движения руки и сопутствующих ему других телесных изменений, а также кодовых программ энергетического обеспечения всего комплекса этих изменений, приводящих к достижению цели). Разумеется, произвольное действие требует более полного описания, мы ограничились лишь общими принципиальными моментами в его объяснении.

**П-б.** *Явления СР могут служить причиной не только телесных изменений, но и причиной изменения других явлений СР*, когда, например, одна мысль влияет на другую, влечет другую и т.п. Это повсеместный факт нашего опыта. Однако задача дискретизации континуума СР и вычленения отдельного явления СР вызывает серьезные теоретические трудности. Тем не менее в ряде простых случаев такая операция может быть корректно проведена. Тогда, если одна мысль (А) вызывает другую мысль (Б), то это равносильно преобразованию нейродинамического кода первой в нейродинамический код второй. Здесь также имеет место психическая причинность. Ведь внутренний «механизм» следования Б из А принципиально не отличается от тех процессов, когда явление СР вызывает определенное телесное изменение. Различны лишь контуры кодовых преобразований, те подсистемы головного мозга в которых они совершаются.

**П-в.** Когда мы говорим об отдельном явлении СР, то важно учитывать, что оно всегда принадлежит данному уникальному «Я» и несет на себе его печать, оно есть момент целостной СР, существующей только в конкретной личностной форме. Эта целостность, определяемая нашим «Я», представлена тем, что может быть названо **эго-системой головного мозга**. Будучи структурно и функционально подсистемой головного мозга, эта *эго-система* образует высший уровень мозговой самоорганизации и управления; именно на этом уров-



не функционируют кодовые структуры типа X. Она охватывает не только сферу сознательных, но и сферу бессознательных психических процессов, регулирует их взаимодействие. Только в контурах эго-системы информационные процессы приобретают качество СР, что связано со специфическими кодовыми преобразованиями. Эти кодовые преобразования отображают, в частности, и уникальные особенности эго-системы (и, значит, *личностные* особенности индивида), в том числе и такой личностный параметр как волеизъявление. И тут возникает традиционный вопрос о свободе воли, который всегда стоял в центре дискуссий по проблеме сознания и мозга.

**П-г.** *Совместим ли феномен свободы воли с детерминированностью мозговых процессов?* На этот вопрос можно дать положительный ответ. Здесь нет нужды вдаваться в подробный анализ феномена свободы воли. Для наших целей достаточно признать, что по крайней мере *в некоторых случаях* человек может сам совершать выбор действий, управлять движением своей мысли, переключать внимание, оперировать по своей воле теми или иными явлениями собственной СР (представлениями, интенциональными векторами), хотя в составе СР есть такие классы явлений, которые либо вообще неподвластны произвольному оперированию, либо поддаются ему с большим трудом. Но признание пусть частичной способности «Я» оперировать явлениями собственной СР (т.е. информацией в «чистом» виде), например способности переводить А в Б, равносильно признанию того, что я могу по своей воле оперировать их нейродинамическими кодами. Следовательно, как бы это странно ни звучало на первый взгляд, **я могу по своей воле оперировать некоторым классом своих мозговых нейродинамических систем, т.е. управлять ими** (хотя и совершенно не чувствуя этого; не ведая, что творю!). Более того, это означает, что я могу оперировать не только некоторым наличным множеством собственных мозговых нейродинамических систем, активировать и дезактивировать их определенную последовательность, но и формировать направленность кодовых преобразований (в тех или иных пределах) и, наконец, создавать новые кодовые паттерны типа X. Нельзя же отрицать, что человек своим творческим усилием продуцирует оригинальные мысли, уникальные художественные образы. Эти новообразования в сфере его СР имеют свое необходимое кодовое воплощение в его мозговой нейродинамике.

**П-д.** Поскольку способность создавать новообразования в сфере СР равнозначна способности порождать новообразования на определенном уровне мозговой нейродинамики (кодовой организации типа X), то это дает основание говорить о постоянной *возможности расширения диапазона возможностей саморегуляции, самосовершенствования, творчества*. И это относится, конечно, не только к управлению своими психическими процессами, но и к управлению телес-

ными процессами, к психосоматическим контурам саморегуляции. Когда человек, как иногда говорят, силой воли подавляет боль (или когда йог вызывает у себя замедление сердечного ритма), то это означает, что он формирует у себя такие паттерны мозговой нейродинамики, такую цепь кодовых преобразований, которые «пробивают» новый эффекторный путь и «захватывают» вегетативные и другие нижележащие уровни регуляции, обычно закрытые для произвольного управления.

**II-е.** Но способность управлять собственной мозговой нейродинамикой может быть истолкована только в том смысле, что нейродинамические системы типа **X**, взятые в их актуальной и диспозициональной взаимосвязи, являются **самоорганизующимися**, образуют в мозгу человеческого индивида личностный уровень мозговой самоорганизации (эго-систему). Следовательно, акт свободы воли (как в плане производимого выбора, так и в плане генерации внутреннего усилия для достижения цели) есть акт **самодетерминации**. Тем самым устраняется тезис о несовместимости понятий свободы воли и детерминизма, но последнее должно браться в смысле не только внешней, но и внутренней детерминации (задаваемой программами самоорганизующейся системы).

**III.** Изложенное выше дает ряд существенных оснований для сопоставления мозга и компьютера, прежде всего в плане осмысления различия информационных процессов, осуществляемых каждым из них, понимания той весьма большой дистанции, которая существует между ними. **У компьютера нет субъективной реальности. Информационный процесс, лишенный качества СР, отличается по своей организации, по своим структурным, оперативным и целевым характеристикам от того информационного процесса, который специфичен для эго-системы головного мозга.** Разумеется, многие информационные процессы и в головном мозгу и тем более в других подсистемах нашего организма идут «в темноте» (как выражаются некоторые западные философы), не дают о себе знать в виде субъективных проявлений. С чисто функциональной точки зрения добавка в виде явления СР кажется излишней. Но это именно кажимость. На самом деле возникновение в ходе эволюции СР ознаменовало новый этап и новый тип самоорганизации. Здесь возникает ряд вопросов, которые должны стать предметом тщательного анализа.

**III-а.** Пока нам известны **два вида СР** – животный и человеческий. Весьма вероятно, что в других звездных мирах есть существа, обладающие СР совершенно иного типа (предмет для фантазирования, но и для теоретических размышлений). Хотя некоторые выдающиеся ученые и философы (Дж. Экклз, К. Поппер и др.) отрицают возможность создания таких систем ИИ, которые способны обладать СР, подобная возможность теоретически обоснована (тем более в от-

ношении различных симбиозов ИИ с ЕИ). Это вытекает из принципов функционализма (которые, по моему убеждению, сохраняют рациональный смысл, могут служить не только редукционистским целям, но и нередукционистским объяснениям СР, что я пытался продемонстрировать выше).

Функциональное описание и объяснение логически независимо от физического описания и объяснения, что, как известно, убедительно было показано Тьюрингом, Патнэмом и др. Это обязывает принять тезис об изофункционализме систем, развитый А. Тьюрингом (один и тот же набор функций может быть воспроизведен системами, различными по своим субстратным, физическим свойствам). Для обоснования тезисов об изофункционализме систем и о возможности обретения искусственным интеллектом качества СР важное теоретическое значение имеет принцип инвариантности информации по отношению к физическим свойствам ее носителя (**ПИ**). Из него следует возможность возникновения различных вариантов кодовой самоорганизации. Тот вариант, который возник в ходе эволюции был не единственно возможным. Разумеется, **ПИ** не означает безразличия физических свойств носителя информации, но лишь то, что одна и та же информация может иметь носители с разными физическими свойствами. В ходе эволюции отбирались коды наиболее экономичные в энергетическом отношении, наиболее компактные по своей организации и т.п. (так сформировались фундаментальные коды земных самоорганизующихся систем – код ДНК, частотно-импульсный код в нервной системе, язык). Да, теоретически, были возможны иные варианты, но в доступном нам мире существует лишь тот вариант самоорганизующейся системы, наделенной СР, который был изобретен, создан биологической эволюцией. И он так или иначе указывает пути сближения ИИ с ЕИ.

**III-б.** Вопрос о возникновении СР – это прежде всего вопрос о *способе представленности информации* для сложной самоорганизующейся системы и *способе использования ее для управления* своим целостным функционированием (поведением). Этот новый способ представленности информации и оперирования ею в целях управления возник в связи с чрезвычайным усложнением живой системы (включающей множество самоорганизующихся подсистем) и потребностью нахождения оптимальных средств поддержания ее целостности, **централизации самоотображения и управления** (как условия реализации адекватного поведения и, в конечном итоге, выживания). В силу множества уровней и структур самоорганизации в развитии организме (клетки, органы, системы дыхания, кровообращения и др.) эволюция постоянно решала проблему соотношения **иерархических, кооперативных и конкурентных контуров управления в структуре целостного организма, соотношения централизации и относительной автономности в функционировании его подсистем.**

Возникновение психики, способности отображения и управления в форме СР явилось ответом на эти проблемы. Представленность информационных процессов в форме явлений СР – чрезвычайно удобный, экономичный, высоко оперативный способ получения, переработки и использования информации в целях эффективного управления многосложным организмом, централизации его действий, которая (централизация) интегрирует нижележащие уровни управления (в клетках, органах и т.п.), сохраняя их определенную автономию. Представленность информации в форме СР позволила резко расширить не только «содержательный» («когнитивный») диапазон информации, но в еще большей степени ее **ценностные** измерения – посредством различных эмоциональных состояний, таких мощных субъективных регуляторов и стимуляторов поведения как боль, оргазм, чувство голода. У высших животных СР достигает значительной степени индивидуализации и разнообразия психических модальностей.

**III-в.** В процессе антропогенеза произошло качественное развитие психического отображения и управления – возникло сознание, отличительная черта которого в том, что ***СР сама становится объектом отображения и управления в форме СР.*** Другими словами, создается возможность по существу *неограниченного производства информации об информации* и способность наряду с информационным управлением телесными изменениями так же и *управления информационными процессами на уровне СР.* Для этого формируется специальная кодовая система – язык. Развивается способность абстрагирования, возникает высокая степень свободы оперирования информацией в «чистом» виде – типа мысленных действий, предваряющих реальные действия, мысленного моделирования вероятных ситуаций, прогнозирования, проектирования, фантазирования, творческих решений, самополагания и волеизъявления. Все эти функции ЕИ заведомо отсутствуют у компьютера.

В отличие от высших животных (у которых тоже можно обнаружить нечто сходное с нашим Я, своего рода Самость, выражающую психические особенности данного индивида, его целостную форму СР, способность совершать выбор в сложных ситуациях) у человека СР обретает новую динамическую структуру. Элементарный феноменологический анализ показывает, что наша СР имеет своей базовой динамической структурой *единство противоположных модальностей Я и не-Я.* Это означает, что Я всегда полагает себя через свое не-Я, в форме которого способно выступать любое «содержание» (информация) – внешняя предметность, явления собственного тела, социальные общности с которыми Я себя идентифицирует или которым оно себя противопоставляет, наконец, само Я и его отдельные атрибуты. В двуедином динамическом контуре «Я – не-Я»

постоянно совершается как бы двойное отображение, двумерный информационный процесс (в котором одновременно отображается и некоторый объект и само его отображение); в нем происходят все текущие ценностно-смысловые преобразования, формируются и реализуются деятельно-волевые проекции (подробно эти вопросы анализируются в 4. См. раздел «Структура субъективной реальности»). Это прежде всего и знаменует качественное отличие сознания от психики высших животных. Платой за такой новый уровень организации информационных процессов у человека является сумасшествие, патологическое раздвоение личности. Животные не болеют шизофренией.

**III-г.** Феноменологические характеристики ЕИ (см. 7), посредством которых обычно указывают на качественное отличие ЕИ от ИИ, выражают существенные структурно-функциональные особенности информационных процессов в головном мозге. Как свидетельствуют данные нейроморфологии и нейрофизиологии, в головном мозге переработка информации совершается одновременно, параллельно во многих различных по своим функциям структурах, результаты которой анализируются и выборочно интегрируются в зависимости от актуализованной цели, от хода решения задачи. Переработка информации в головном мозгу, выражающая текущую мыслительную деятельность, совершается отнюдь не по жесткой двоичной логической схеме. Скорее эта логика представляет собой **многозначную логику, в которой число значений истинности есть величина переменная; при этом число значений истинности меняется в зависимости от характера решаемой задачи и, возможно, от разных этапов ее решения.** В этой многомерной динамической структуре двоичная логическая схема лишь один из существенных моментов процесса переработки информации. Головному мозгу присущи развитые функции вероятностного прогнозирования, весьма оригинальные, эффективные способы сжатия информации и выборки нужных элементов из памяти, эвристического синтеза и другие операции, которые вряд ли допустимо приписывать современным компьютерам.

**III-д.** Сказанное, конечно, не умаляет роли и возможностей ИИ. Выдающиеся достижения компьютерных наук и информационных технологий положили начало новому этапу цивилизации – информационному обществу. Проблемы дальнейшего развития ИИ будут в существенной мере определять судьбы человечества. Но это обязывает нас к тщательному анализу и реалистическим оценкам широко-вещательных проектов. Среди ряда специалистов в области ИИ бытует убеждение, что быстро нарастающая вычислительная мощь компьютеров скоро приведет к появлению у них сознания. Так, наш бывший соотечественник профессор А. Болонкин, живущий ныне в США, рассуждает следующим образом: уже создан компьютер в 8 терафлоп, лет через 20 будет построен «суперкомпьютер, мощность

которого превзойдет мощность мозгов всего человечества». Следовательно, такой компьютер не может не обладать качеством, присущим отдельному человеческому мозгу (см. 1).

К сожалению, однако, это вовсе не тот случай, когда огромное количественное накопление приводит к новому качеству. Здесь как раз явно выступает узость, ограниченность («вычислительной») трактовки сознания и мышления, которая подвергалась основательной критике. Примером такой критики могут служить контраргументы выдающегося физика и математика Р. Пенроуза. Он убедительно показывает, что даже собственно «математическое понимание» не может быть сведено к вычислительным операциям, подчеркивает ««невычислимый» характер математического восприятия» (см. 12, с. 117). Но это относится не только к математическому, но ко всякому пониманию. Оно составляет важнейшую особенность всякой познавательной деятельности, неотъемлемое свойство сознания. Понимание «является весьма общей чертой, присущей всем человеческим существам, и эта способность принципиально не является вычислительной по своей природе, вне всякой зависимости от математики» (там же, с. 118).

«Вычислительная» трактовка сознания и мышления представляет собой один из вариантов редукции сознания к «когнитивным операциям». При этом качество СР оказывается чем-то эфемерным, нереальным. Это позиция так называемого «**Сильного ИИ**». Весьма популярный ее представитель, американский философ Д. Деннет, иронизируя по поводу «тайного огня сознания», прямо заявляет, что главным признаком сознания является «функционирование когнитивно-информационных процессов» (см.: 14, с. 216 – 218). В таком случае сознание можно приписывать всем системам, способным совершать «разумные операции» и прежде всего, конечно, компьютеру. Подобная позиция крайне уязвима, она отвергается большинством философов, обсуждавших проблемы ИИ.

Существенное сближение ИИ с ЕИ предполагает более глубокое исследование ЕИ. На прошедшей недавно конференции по философии ИИ многие докладчики справедливо подчеркивали, что наиболее узким местом компьютерного моделирования является недостаточная исследованность ЕИ. Наиболее основательно и многопланово это было выражено в докладе А.П. Огурцова «Достижения и трудности моделирования интеллектуальных актов». В нем убедительно показано, что «перспективы компьютерного моделирования и нейрокомпьютеринга – это перспективы философии и психологии сознания. Именно ими будет создан новый язык анализа сознания, в том числе интеллекта, который найдет свою амплификацию в компьютерных моделях» (11, с. 59). С этим нельзя не согласиться. А.П. Огурцов справедливо утверждает: «Все компьютерные модели далеки от биологической основы работы интеллекта» (там же, с. 57).

Это принципиальное обстоятельство осознают и специалисты в области компьютерных наук, те, кто непосредственно занимаются моделированием интеллектуальных функций. Так, В. Г. Редько, говоря о том, что развитие теории «нейронных сетей» связано с внедрением в «нейроинформатику» высокоэффективных математических методов, вместе с тем отмечает: «Однако, несмотря на чрезвычайную активность исследований по нейронным сетям и нейрокомпьютерам, многое в этих исследованиях настораживает. Зачастую изучаемые алгоритмы выглядят как бы «вырванным куском» из общего осмысления работы нервной системы. Исследуются те алгоритмы, для которых удастся построить хорошие модели, а не наиболее важные для понимания свойств мышления, работы мозга и для создания систем искусственного интеллекта» (13, с.94). «Все это указывает на необходимость максимально полного понимания работы биологических систем обработки информации и свойств организмов, обеспечиваемых этими системами» (там же, с. 94 – 95).

В этой связи принципиальное значение имеют современные нейрофизиологические исследования психической деятельности, в особенности использующие методы позитронно-эмиссионной томографии, функционально-магнитного резонанса, многоканальной записи электрических и магнитных полей мозга. В последнее время достигнуты существенные результаты в изучении тех мозговых информационных процессов, которые лежат в основе субъективных переживаний, обуславливают возникновение ряда явлений СР, в том числе относящихся к процессу мышления. Здесь прежде всего должны быть отмечены исследования А. М. Иваницкого и ряда других авторов, получивших близкие результаты (Дж. Эделмен, В.Я. Сергин и др.).

В работах А. М. Иваницкого (см.: 9, 10 и др.), проводимых на протяжении вот уже более тридцати лет, убедительно показано, что качество субъективного переживания связано с определенной организацией мозговых процессов, возникает при сопоставлении в зонах коры мозга вновь поступившей информации с той, которая извлечена из памяти. «Такое сопоставление возникает в результате кольцевого движения возбуждения с его возвратом к местам первоначальной проекции после дополнительной обработки в других структурах мозга» (10, с. 717). Автором четко обозначены временные параметры перехода физиологического процесса на тот уровень его организации, который связан с возникновением ощущения. Это цикл, длящийся примерно 150 мс, названный автором «кругом ощущений», в котором осуществляется сравнение сенсорного сигнала со сведениями, извлеченными из памяти, включая данные о значимости сигнала. Ощущение, как простейшее субъективное переживание, есть результат «информационного синтеза», совершающегося в рамках указанного цикла (см. там же, с. 717 – 718).

А. М. Иваницким показано, что принцип возврата возбуждения и информационного синтеза оправдывает себя не только в случае ощущений, но и при анализе весьма сложных субъективных состояний, связанных с процессами мышления. Найденный эволюцией и используемый в ходе развития психики животных, он сохраняет свое значение для понимания тех мозговых процессов, которые лежат в основе высших форм психической деятельности, человеческого сознания (см. там же, с. 719 – 721).

Важно подчеркнуть, что проводимые в этом плане исследования преследуют стратегическую цель: выяснение существенных и необходимых свойств *того типа самоорганизации*, который создает *представленность для системы информации в форме СР и способность оперировать ею (в «чистом» виде)*, что имеет первостепенное значение для создания новых направлений разработки ИИ.

В заключение хотелось бы еще раз подчеркнуть следующее. Психика и сознание – уникальный эффект биологической самоорганизации, найденный в процессе эволюции. Современные искусственные информационные системы далеки от самоорганизации такого типа, более того, сами по себе они не могут быть названы самоорганизующимися системами в точном смысле этого слова, ибо на входе и на выходе у них стоит человек, задающий программу и использующий результат их деятельности.

Как уже отмечалось, теоретически мыслимо искусственное создание самоорганизующихся систем такого типа, которые обретут, наконец, главное отличительное свойство ЕИ – субъективную реальность (пусть хотя бы в ее элементарных формах). Но пока развитие компьютерной техники идет не в этом направлении. Возможно, к решению задачи такого рода нас в чем-то приблизят квантовые компьютеры. Но в любом случае ее решение должно быть связано с воспроизведением некоторых структурно-функциональных аналогов биологической самоорганизации. Иное трудно себе представить. Хотя, конечно, в общем плане, отметим это еще раз, нельзя исключать две возможности: 1) что где-то во Вселенной существует некий тип самоорганизации, качественно отличный от нашего, земного, но тоже обладающий СР (которая, впрочем, может быть столь же отличной от нашей, человеческой СР); 2) что человеческий разум изобретет некую самоорганизующуюся систему, качественно отличающуюся от самоорганизации биологического типа, у которой появится своя СР, пусть и отличная в чем-то или во многом от нашей человеческой СР. Однако способы реализации этих двух возможностей не поддаются конкретным представлениям (даже на уровне литературной фантастики анализ легко выявляет всё те же стандартные структуры нашей СР, неамысловатые интерпретации, аранжировки и перекомпоновки их смысловых, ценностных, оперативных фрагментов, не более того). Эти две отмеченные возможности



слишком маловероятны, чтобы служить какой-либо опорой при нынешнем проектировании дальнейших разработок проблемы ИИ.

Теоретически мыслима в общих чертах и та новая электронная цивилизация, которую с энтузиазмом обрисовывает проф. А. Болонкин (создание «электронного человека» путем переписи хранящейся в его мозгу информации в чипы и достижения таким способом его бессмертия, которое станет реальным, как обещает автор, всего через 20-25 лет). Эта цивилизация электронных трансгуманоидов якобы неизбежно придет на смену нынешней, имеющей биологическую основу. А постольку, как полагает проф. А. Болонкин, биологическая самоорганизация обречена!

Такого рода мыслимая возможность (как всякая мыслимая возможность) остается весьма проблематичной и, конечно же, она крайне далека от конкретного осуществления. Между тем наша земная цивилизация находится в сильном цейтноте (прежде всего из-за быстрого нарастания масштабов и последствий экологического кризиса). И мы должны сейчас быть озабочены в первую очередь не столько созданием «электронного человека» и его бессмертием, сколько сохранением жизни обычного человека и земной жизни в целом. Сама же стратегическая установка – я цитирую А. Болонкина – на «неизбежность гибели биологического человечества» (см.: **1**) не имеет достаточного основания, является крайне опасной, я бы сказал – самоубийственной.

Ей противостоит другая стратегическая установка, опирающаяся на традиционные ценности и здравый смысл. Она не спешит расторгнуть связь человека с его биологическим телом, не спешит заменять его трансгуманоидом, подчеркивает наличие больших ресурсов самоорганизации и самосовершенствования у биологических систем и человеческого общества, поддерживает веру в творческие и благие силы разума, научного познания, в продуктивное развитие информационных технологий и робототехники на благо людей. И она, конечно, уповает на преодоление экологического кризиса и сохранение земной жизни как фундаментальной, непреходящей ценности. Такая стратегическая установка является более реалистичной и ответственной, хотя она тоже в ряде отношений проблематична, требует дальнейшей основательной, творческой разработки. В этом плане высокую актуальность приобретает проблематика духовного и телесного, сознания и мозга, вопросы касающиеся развития связей и взаимодействий ИИ с ЕИ, различных форм их симбиоза в будущем, как средства решения насущных задач человечества.

## ЛИТЕРАТУРА

1. Болонкин А. (Человек – бессмертен! – Интервью с профессором Александром Болонкиным) // «Известия» от 8 сентября 1998 г.
2. Дубровский Д.И. Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М., «Наука», 1971.
3. Дубровский Д.И. Информация, сознание, мозг. М., «Высшая школа», 1980.
4. Дубровский Д.И. Проблема идеального. Субъективная реальность. Изд. 2-е. М., «Канон», 2002.
5. Дубровский Д.И. Проблема духа и тела: возможности решения // Вопросы философии, 2002, № 10.
6. Дубровский Д.И. Новое открытие сознания? (По поводу книги Дж. Серла «Открывая сознание заново») // Вопросы философии, 2003, № 7.
7. Дубровский Д.И. Гносеология субъективной реальности. К постановке проблемы // Эпистемология и философия науки, 2004, № 2.
8. Дубровский Д.И. В «Театре» Даниэля Деннета (По поводу одной популярной концепции сознания) // Философия сознания: история и современность. М., МГУ, 2003.
9. Иваницкий А.М. Главная загадка природы: как на основе работы мозга возникают субъективные переживания // Психологический журнал, 1999, № 3.
10. Иваницкий А.М. Естественные науки и проблема сознания // Вестник Российской Академии Наук, 2004, т. 74, № 8.
11. Огурцов А.П. Достижения и трудности в моделировании интеллектуальных актов // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции. М., 2005.
12. Пенроуз Р. Большое, малое и человеческий разум. М., 2004.
13. Редько В. Г. Эволюционная кибернетика. М., 2003.
14. Dennet D. Consciousness Explained. Boston, 1991.

---

# ФИЗИОЛОГИЧЕСКИЕ ОСНОВЫ СОЗНАНИЯ И ПРОБЛЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА\*

*А.М. Иваницкий*

## **Введение**

Сознание – важнейшая функция мозга. По существу, наше сознание – это и есть наша жизнь, состоящая из череды впечатлений, мыслей и чувств. Говоря об изучении физиологических механизмов сознания, можно поставить вопрос, насколько правомерно объяснить движением нервных импульсов то, что мы ощущаем как цвет, звук, мысль или эмоцию. Субъективный мир – это внутреннее чувство, «внутренний театр» каждого, и задача науки о мозге – понять, какие нервные процессы приводят к его возникновению. Несмотря на видимую сложность, загадка «психика и мозг» по своей методологической трудности не уникальна и стоит в ряду других тайн природы. По существу, возникновение нового качества происходит на каждом этапе принципиального усложнения природных процессов. Как пример такого качественного перехода Ф.Крик и К.Кох приводят возникновение жизни в результате некоторых особенностей цепочек ДНК и ферментных белков. Присущие живым объектам свойства не вытекают непосредственно из физической химии каждой из этих молекул. Этот пример кажется особенно убедительным в устах одного из первооткрывателей генетического кода.

В решении проблемы «сознание и мозг» можно выделить два основных подхода, которые не исключают, а взаимно дополняют друг друга. Первый из них – представление о том, что субъективный опыт возникает на основе поступательного распространения возбуждения от первичных зон коры к структурам более высокого уровня, к которым, в первую очередь, относится лобная кора. Она обладает тремя уникальными свойствами: способностью к оперированию абстрактными символами, запоминанию временной последовательности происходящих событий и наличием центров речи. Эти три качества могут быть непосредственно соотнесены с признаками сознания.

Второй подход основан на предположении, что субъективный опыт возникает в результате определенной организации процессов мозга и

---

\* Работа выполнена при поддержке Российского гуманитарного научного фонда (проект № 05-06-06-215а), программы Президиума РАН «Фундаментальные науки – медицине» и программы ОБН РАН «Интегративные механизмы регуляции функций и организмов».

сопоставления в зонах коры вновь поступившей информации с той, которая извлечена из памяти. Благодаря этому информация о внешних событиях как бы проецируется на индивидуальный опыт субъекта, встраиваясь в личностный контекст. Данная гипотеза в настоящее время разделяется большинством авторов. Впервые она была выдвинута нами в 1970-х годах в результате проведенных исследований мозговых механизмов ощущений.

### **Механизмы ощущений**

В работе была поставлена задача сравнить количественные показатели физиологии и психологии, описывающие ответ на поступивший сигнал. Исследуемый решал задачу по различению интенсивности двух раздражителей. На предъявляемые стимулы записывалась электрическая активность мозга в виде так называемых вызванных потенциалов (ВП), то есть ответа мозга на сигнал от органов чувств. Для получения количественных параметров ощущений были использованы методы психофизики, то есть «науки о количественном соотношении души и тела», как определил эту науку ее основатель К.Фехнер. В данном случае были применены методы теории обнаружения сигнала, которая рассматривает восприятие как результат взаимодействия сенсорных и мотивационных факторов.

После получения соответствующих количественных показателей (физиологических и психологических) вычислялись корреляции между ними. Наиболее существенной была корреляция промежуточных волн ВП с обоими факторами восприятия: показателем сенсорной чувствительности  $d'$  и критерием решения. Эта двойная корреляция отражала синтез информации о физических и сигнальных свойствах стимула на нейронах проекционной коры. Пиковая латентность соответствующих волн ВП составляла около 150 мс.

Принципиально важно, что это время достаточно точно совпало со временем возникновения ощущений, полученным ранее в психофизических экспериментах. В основном, эти измерения были проведены с использованием феномена «обратной маскировки». Суть ее заключается в том, что, если после первого слабого стимула через короткий интервал следует второй, более сильный, первый стимул не воспринимается. Постепенно увеличивая промежуток времени между первым и маскирующим сигналом, можно найти интервал, при котором маскирующий эффект исчезает, так как ощущение на первый сигнал уже сформировано. Таким путем было установлено, что ощущение появляется примерно через 150 мс после действия стимула. Наиболее точные данные были, однако, получены в начале 1990-х годов, когда в качестве маскирующего сигнала была использована прямая транскраниальная стимуляция коры коротким магнитным импульсом. Полученные здесь данные также показали цифры, близкие к приведенным выше. При этом существенно, что магнитный импульс вызывал маски-

рующий эффект только в том случае, когда он прикладывался к проекционной, в данном случае, зрительной коре, то есть только там, где наблюдалась описанная выше двойная корреляция волн ВП с показателями восприятия. Все эти данные свидетельствовали о том, что ощущение возникает значительно позднее прихода сенсорных импульсов в кору и, следовательно, является результатом сложной организации нервных процессов, которая и была изучена в нашей работе.

Основываясь на данных о физиологическом генезе волн вызванного потенциала, нами был описан механизм, обеспечивающий информационный синтез. Он включал кольцевое движение возбуждения из проекционной коры в ассоциативную кору (височную для зрительных стимулов), затем в область гиппокампа, мотивационные центры гипоталамуса с последующим возвратом по системе диффузных проекций в кору, включая и зоны первичной проекции. Условно этот цикл был назван нами «кругом ощущений». Суть его в том, что он обеспечивает сравнение сенсорного сигнала со сведениями, извлеченными из памяти, что, как мы предполагаем, и лежит в основе перехода физиологического процесса на уровень психических, субъективно переживаемых явлений. Эти представления были обозначены как гипотеза информационного синтеза.

За прошедшие 25 лет в литературе появилось много работ, подтверждающих эти мысли, включая как данные о топографии отделов мозга, входящих в «круг ощущений», так и использование самой идеи возврата возбуждения для объяснения механизмов сознания. Среди последних работ наиболее фундаментальны исследования Дж.Эдельмана, нобелевского лауреата за изучение структуры антител, который так же, как и Ф.Крик, а до него И.П.Павлов, после получения этой высшей научной награды стал заниматься проблемами сознания. Дж.Эдельман использовал термин «повторный вход (re-entering)». Созданная им нейробиологическая теория сознания и сам термин повторный вход получили достаточно широкое распространение. Повторный вход обеспечивает и интеграцию отдельных признаков стимула в единый образ. Исследованиями последних лет показано, что в такой интеграции важную роль играет гамма-ритм ЭЭГ с частотой около 40 Гц, который возникает, как правило, примерно через 100 мс после стимула.

Ощущение относится к достаточно простым психическим феноменом, которые некоторые авторы относят к так называемому «первичному сознанию». К этому же классу можно отнести и эмоции, в исследовании которых важный вклад внес П.В. Симонов. Он впервые предложил формулу эмоций, согласно которой сила эмоции пропорциональна потребности, умноженной на разность между сведениями, имеющимися у индивидуума, и информацией, необходимой для удовлетворения данной потребности:

$$\mathcal{E} = P^*(\text{Инф. налич.} - \text{Инф. необх.})$$

Из этой формулы следует, что эмоции, так же, как и ощущения, возникают в результате сравнения двух информационных потоков. Здесь действует, таким образом, некоторая универсальная закономерность.

### **Мышление**

Более сложные психические феномены, в первую очередь, связанные с появлением речи, относят к сознанию высшего порядка. По мнению П.В.Симонова, оно имеет коммуникативную природу и возникло в процессе общения между людьми. В исследовании более сложных функций мозга, таких как мышление, значительный вклад внесли также работы Н.П.Бехтеровой и ее школы.

Современные данные показывают, что кора высоко специализирована, и разные ее поля отвечают за различные когнитивные операции. Поэтому существенную роль в процессе мышления приобретают корковые связи. В исследовании связей велика роль русской нейрофизиологической школы. Ведущей идеей здесь является то, что возникновению связи способствует согласование ритмов работы нейронных ансамблей (А.А.Ухтомский, М.Н.Ливанов, В.С.Русинов). В наших работах по изучению мышления был применен новый метод картирования корковых связей, базирующийся на тех же идеях. В работе были использованы задачи на образное, пространственное и абстрактно-вербальное мышление. Задание предъявлялось на экране монитора. Проводилась многоканальная запись ЭЭГ в период между предъявлением задачи и нахождением решения. Главным результатом этих исследований было установление факта, что симметричный в покое рисунок связей при решении задачи изменялся: связи начинали сходиться к определенным полям коры. Возникающие в результате этого центры связей были обозначены как фокусы взаимодействия. Их топография была различна при разных видах мышления. Так, при образном мышлении (опознании эмоций на фотографиях лица) фокусы локализовались в теменно-височной коре. При абстрактно-вербальном мышлении (на примере решения анаграмм или категоризации слов) они были расположены в лобной коре. При пространственном мышлении, включающем элементы как образного, так и абстрактного мышления, образование фокусов происходило в теменной и лобной коре.

Было установлено также, что информация приходит к фокусам по связям, которые поддерживаются на разных частотах, при этом каждая из связей несет информацию из различных отделов коры, имеющих свою специализацию. В фокусе, нейронные группы которого соединены жесткими связями, происходит синтез этой информации, благодаря чему, как можно предполагать, и достигается решение. Тем самым идея информационного синтеза была распространена и на процессы мышления. Различие заключается в том, что вместо двух

информационных потоков при ощущении при мышлении их число значительно больше. Сюда относятся сигналы от органов чувств, из оперативной и долговременной памяти и, что очень важно, мотивация из мотивационных центров. Другое различие в том, что центры синтеза при мышлении находятся в ассоциативной, а не в проекционной, как при возникновении ощущений, коре. Интересно также, что при всех видах задач, даже в том случае, когда речевой ответ испытуемого не требовался, на последнем этапе решения фокусы возникали в речевой левой височной области, что свидетельствует о том, что вербализация является у человека важным компонентом мышления. В этом смысле можно говорить о конвергенции двух подходов к решению проблемы «сознание и мозг» – психическое возникает на основе определенной организации нервных процессов, которая захватывает и некоторые ключевые области лобной коры.

Несколько слов об анализе вербальных сигналов и понимании словосочетаний, образующих предложение. С помощью метода позитронно-эмиссионной томографии и вызванных потенциалов мозга была установлена роль лобной коры в расшифровке семантики слова. Эти данные существенно дополнили классические представления о том, что значение слова определяется в левой височной коре в зоне Вернике. В то же время установлено, что в объединении нескольких слов в единую смысловую конструкцию важную роль играют связи между лобной корой и зоной Вернике. Эти данные были получены в наших совместных исследованиях с лабораторией Университета штата Орегон в США, возглавляемой М.Познером.

### **Декларативная память и внимание**

Важным свойством сознания является способность к запоминанию и воспроизведению последовательности событий. Это свойство памяти французский философ А.Бергсон назвал «памятью души», в отличие от «памяти тела», ответственной за образование простых навыков. В исследовании механизмов человеческой памяти в последние годы был достигнут значительный прогресс. Работами ряда отечественных и американских авторов была установлена важная роль гиппокампальных структур, расположенных внутри височных отделов полушарий, в оперативной памяти. Эти образования имеют обширные связи как между собой, так и с сенсорными и ассоциативными отделами коры. Считается, что при запоминании они направляют сигнал в ассоциативную кору для длительного хранения, а при воспроизведении указывают адрес, где хранится связанная с поступившим сигналом информация. Гиппокампальный комплекс выполняет, таким образом, роль, сходную, в известной мере, с библиотечным каталогом, благодаря которому можно быстро найти нужную книгу. Различие между этими двумя основными структурами комплекса заключается в том, что энторинальная кора участвует в запоминании

информации вне ее связи с контекстом, а гиппокамп играет важную роль в декларативной памяти, которая представляет собой осознаваемую память о событиях, сведения о которых могут быть переданы другим лицам. Для пояснения различия между семантической и декларативной памятью можно привести простой пример. Предположим, что вы встречаете человека, лицо которого вам знакомо, но вы не помните, кто он. Это узнавание, связанное с семантической памятью. В случае декларативной памяти вы знаете, кто это, и при каких обстоятельствах вы с ним познакомились.

Роль гиппокампа была, в частности, установлена в результате наблюдения над больными с повреждением этой структуры мозга, что приводило к нарушению декларативной памяти (М.Мишкин). Такие больные могут достаточно хорошо усваивать новые сведения, они могут даже успешно учиться в школе и иметь высокий интеллектуальный коэффициент. В то же время эти больные беспомощны в повседневной жизни, так как не помнят последовательности событий, не ориентируются во времени, не могут составить плана на будущее. Интересно, что эта неполноценность памяти проявляется только с 5-6-летнего возраста, то есть с того времени, когда здоровый человек начинает себя помнить.

Наряду с гиппокампом, важную роль в запоминании последовательности событий играет и лобная кора. Там находятся группы нейронов, способные сохранять след от действовавшего сигнала до того момента, когда необходимо дать поведенческий ответ на него (П.Гольдман-Ракич).

Сознание также тесно связано со вниманием: осознается только то, на что обращается внимание. Нашими исследованиями было показано, что память играет важную роль и в механизмах избирательного вербального внимания, когда человек должен воспринимать и реагировать лишь на определенный класс сигналов, выделяя их из массы других. Способность к этому иногда называют «эффектом коктейль-парти», когда человек должен поддерживать диалог с собеседником, выделяя его речь из других разговоров и других звуков. В работе использовалась запись вызванных потенциалов мозга на слова, одновременно предъявлявшиеся на экране монитора и звучавшие из компьютерных колонок. Задача испытуемого состояла в необходимости запомнить как можно больше слов, поступавших по одному из каналов, игнорируя другие. Запоминание и извлечение вербальной информации имеет определенное электрофизиологическое выражение в виде «когнитивных» компонентов вызванного потенциала с латентностью от 400 до 700 мс. Было установлено, что ВП на игнорируемый сигнал характеризовался сдвигом потенциала, обратным по полярности тому, который имеет место при запоминании, что свидетельствовало об активном торможении процессов памяти. Это дает



возможность считать, что избирательность внимания обеспечивается тем, что ненужная информация воспринимается, так как компоненты ВП, ответственные за восприятие, сохранены, и человек слышит и видит слово, но затем имеет место блокада передачи сведений на структуры гиппокампа, вследствие чего ненужная информация не сохраняется в декларативной памяти и вытесняется из сознания.

### **Основные механизмы сознания**

Подведем итог представлениям о возможных механизмах сознания.

Возврат возбуждения и сравнение новой информации с памятью – фундаментальный принцип организации процессов мозга, лежащих в основе субъективного опыта. Его специфика определяется топографией центров синтеза.

Синтез информации в проекционной коре лежит в основе ощущений, в ассоциативной коре – в основе мышления и поиска решений.

Лобная кора ответственна за абстракцию и речь. Вместе с височной корой она осуществляет объединение существительного и глагола, что образует элементарную единицу вербального мышления.

Гиппокамп вместе с лобной корой обеспечивает сохранение в памяти последовательности событий и избирательность внимания.

Общим радикалом перечисленного является участие в механизмах сознания памяти. Постоянное сопоставление вновь поступившей информации с памятью определяет содержание сознания как постоянную корректировку личного опыта, того, что можно определить как чувство внутреннего «я».

Решение проблемы о происхождении сознания и его связи с мозгом – одна из важнейших проблем современной науки. Оно необходимо для создания непротиворечивой картины мироздания, которая бы объединяла естественнонаучное и гуманитарное знание. Не менее существенно оно и для практики: организации труда и отдыха, образования и воспитания. Особое значение имеет проблема для медицины.

Скажем и о некоторых следствиях из наших знаний о природе сознания, которые носят более общий характер. Сознание человека формируется в течение жизни индивидуума. В центре сознания находится ощущение своего «я» как хозяина и распорядителя своих действий. Чувство свободы поэтому изначально, его неоправданное ограничение деформирует сознание. Разумным ограничением свободы каждого является требование не нарушать свободу других людей. Это основная нравственная заповедь, которая, в отличие от первичного чувства свободы, должна быть предметом воспитания. Сознание не только отражает бытие, оно обладает и способностью к прогнозу, формируя бытие будущего. Опираясь на знания, мы можем и должны создавать общество будущего, в котором принципы общественного

устройства находились бы в гармонии с базисными свойствами сознания человека.

### **Физиология мозга и проблема искусственного интеллекта**

Встает вопрос, в какой мере знания о механизмах мышления и сознания человека могут быть использованы при создании искусственного интеллекта. Ведь, в принципе, создание искусственного разума возможно и без использования принципов работы мозга, как, например, конструкция автомобиля не имитирует естественные методы передвижения человека в виде ходьбы или бега. Есть, однако, причина, по которой использование принципов работы мозга может оказаться полезным. Дело в том, что мозг по многим важным параметрам превосходит искусственные вычислительные устройства: он более надежен, экономичен и легко обучаем. Главное же различие, по мнению известного британского физика и математика Р. Пенроуза, состоит в том, что мозг обладает «пониманием», которого лишен компьютер. Можно предполагать, что простое усовершенствование уже имеющихся искусственных систем не может, очевидно, достичь поставленных целей, так как принципы работы мозга во многом отличаются от тех, на которых построен современный компьютер. Мозг отличается от компьютера тем, что он не имеет центрального процессора, который в компьютере оперирует четкими сигналами и который обрабатывает эти сигналы по заданным программам. Напротив, мозг получает, как правило, недостаточно определенные сигналы, оценка которых во многом зависит от контекста, и сам создает программы в результате обучения. Наряду с жесткими связями, мозг использует и гибкие связи, которые образуются на основе синхронизации ритмов деятельности нейронных ансамблей. Благодаря этому, мозг осуществляет эффективный поиск хранящейся в памяти информации, актуальной для выполнения определенной функции. Этот поиск включает в себя и элементы эвристики, неожиданных, но полезных решений.

Надежность мозга при наличии таких, в общем, несовершенных систем анализа сигналов обеспечивается их многократным дублированием. Важной находкой эволюции является также принцип возврата информации к местам первоначальных проекций. Благодаря этому новая информация в мозге постоянно сопоставляется с той, которая хранится в памяти. При мышлении механизм информационного синтеза получает наибольшее развитие. Как уже говорилось, при этом в коре возникают центры связей – фокусы взаимодействия, принимающие информацию из других отделов мозга и осуществляющие ее синтез. Данное положение важно с той точки зрения, что оно указывает на существенный и не неиспользуемый в большинстве искусственных нейросетей прием: нейронные сети мозга не гомогенны, а построены по иерархическому принципу.

Высшие функции мозга возникают, таким образом, на основе высокой степени интеграции специализированных звеньев в единую систему. Эта система представляет собой более высокий уровень организации и обладает поэтому иной логикой и закономерностями саморазвития. Эти законы могут в порядке нисходящей детерминации влиять на протекание процессов, входящих в состав системы и образующих ее. Так логика и грамматика речи могут управлять движением нервных импульсов, определяющих артикуляцию. По мысли Д.И. Дубровского, отсутствие такой нисходящей детерминации является существенным отличием компьютера от живого мозга. По существу, видением общей задачи вычисления обладает лишь программист, но не центральный процессор.

Перейдем к самому, пожалуй, таинственному отличию мозга от компьютера. По мнению Р.Пенроуза, сознание не может быть сведено к вычислениям, так как живой мозг отличается от компьютера тем, что обладает способностью к пониманию. На вопрос, что такое понимание и каковы его мозговые механизмы, ответ должен дать физиолог. Почему же животное, например кошка, понимает, а компьютер – совершенное создание высокого интеллекта – нет. Представляется, что понимание возникает в результате подкрепления, то есть на основе такого фундаментального принципа работы мозга, как условный рефлекс, который связывает внешний стимул и ответные действия субъекта с удовлетворением определенной потребности. Понимание, таким образом, имеет глубокие эволюционные корни и высокий жизненный смысл. В естественных условиях животное научается производить определенные действия, чтобы удовлетворить ту или иную потребность, то есть начинает понимать связь внешних событий, своего поведения и достижение желаемого результата. На этом же основана и дрессировка: чтобы научить собаку выполнять, то есть понимать, определенные команды, дрессировщик использует подкрепление в виде пищи или наказания. По существу, все это изначально относится и к человеку. Так, хороший работник имеет более высокую зарплату, а нерадивого работника штрафуют, герой получает награду, а преступника сажают в тюрьму, чтобы он понял, что нельзя нарушать закон. Но это, конечно, крайние случаи. Практически же все поведение основано на тех же принципах.

Для обоснования своих взглядов Р.Пенроуз использует теорему Гёделя о том, что нельзя доказать вычислением правильность основных действий арифметики, например, что  $1+1=2$ . Но живое существо в этом убеждается, когда получает два банана или двух врагов, добавляя в результате тех или иных действий к первому объекту (или субъекту) второй (или второго). При этом понимание сущности удвоения (или сложения вообще) возникает в эволюции раньше, чем умение считать. Описан, например, случай, когда коренной житель

севера не знал, сколько у него оленей, но легко мог перечислить каждого из них по признакам. Ребенок также может перебрать в памяти всех окружающих его людей или свои игрушки, хотя еще не знает счета.

Подкрепление показывает, является ли поведение верным или ошибочным. В стохастической машине, какой является мозг, такое указание может быть решающим фактором. При этом собственно информационная составляющая в виде вычислений и показатель ее правильности, как правило, поступают по разным каналам и являются, в известной степени, ортогональными. Два компонента «понимания» могут быть представлены сенсорным сигналом и импульсами из мотивационных центров, которые, как говорилось выше, необходимо присутствуют при переходе чисто физиологического процесса в его психический эквивалент. Импульсы из центров мотивации первично непосредственно сигнализируют об удовлетворении потребности. Впоследствии они также могут быть включать систему вычислений, которая оценивает успешность продвижения к цели и выдает команду, когда достигается приемлемая в данной ситуации вероятность получения полезного результата. При этом мотивационная составляющая выступает как аксиома и завершает вычисления. Аксиоматичность подкрепления базируется на знаковой оценке полезности или вреда, которая воспринимается как не требующие доказательств. Мозг вынужден действовать по методу проб и ошибок и учиться на них, так как реальные жизненные условия настолько сложны, что их практически невозможно просчитать до конца.

Мозг живет, таким образом, «по понятиям». Не поэтому ли нам так трудно дается жизнь по закону, основанному преимущественно на вычислениях? Жизнь, по существу, представляет собой жесткую игру, в которой умение и случай определяют судьбу человека. Она совсем не похожа на существование компьютера

Сказанное выше, конечно, является далеко не полным. Мы еще многое не знаем о работе мозга, и особенно о том, что лежит в основе его высших функций и человеческого сознания. Тем не менее, прогресс в этой области в последние годы достаточно очевиден. Оптимистичны и те рекомендации, которые наука о мозге может дать создателям искусственного интеллекта. И мозг, и компьютер находятся в физическом мире и подчиняются его законам. Никакой из перечисленных выше принципов работы мозга не выходит за эти рамки и может быть воспроизведен и усовершенствован в устройствах, созданных человеком.

---

# ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ. НЕКОТОРЫЕ АСПЕКТЫ СХОДСТВА И РАЗЛИЧИЯ

*А.В. Мцхветадзе*

Мы живем в мире, где постоянно совершенствуются машины, называемые искусственным интеллектом. Однако несмотря на постоянный прогресс в этом направлении, по-прежнему налицо существенная разница между двумя видами интеллекта: естественным и искусственным. Что же их разделяет?

Как бы ни были сложны машины, собственного «я», т.е. психики в человеческом понимании мы пока в них не усматриваем, они остаются для нас машинами. Может ли машина по мере усложнения перейти рубеж, за которым она сможет приобрести свое «я»? По мнению ряда кибернетиков, это произойдет, когда у машины появится критическое отношение к самой себе, т.е. своей работе.

Существует мнение, что психика представляет единство отражения действительности и его отношения к чувствительности. Но, конкретно, чего к чему? Критическое отношение к себе в нашем понимании состоит из двух положений: 1) определение степени целесообразности того или иного акта как для сохранения гомеостаза внутри системы, так и в отношении к окружающей эту систему внешней среде; 2) сравнения системы с аналогичными другими и выявление разницы их работы. Постоянно складываются отношения организма с меняющейся информацией, поступающей в мозг, организм уже может проявлять свою реакцию или целиком, или какой-то своей частью. А потому его реакция в целом и есть то самое искомое ощущение, которое эволюционировало от низших к высшим. «Весь организм» реагирует, потому что только на организменном уровне решается вопрос, что более всего требуется для него для сохранения гомеостаза. Только весь организм может «отобрать» вариант реакции, который необходим более всего для жизнедеятельности организма. И тогда может возникнуть психика как отношение всего организма к определенным структурам мозга и различным внешним и внутренним раздражителям. Появляется практическое отношение, т.е. организм решает, насколько эти структуры из памяти соответствуют его требованиям на данный момент, а определяет это мотивация, которая постоянно присутствует.

Таким образом, с нашей точки зрения, «я» или само-сознание – результат постоянно складывающегося отношения всего организма к

своей деятельности, в первую очередь, к работе головного мозга. Но отношение одного к другому не может быть материальным, поэтому мысль – идеальный продукт головного мозга. Со времени пробуждения психики в головном мозгу складываются структуры, реагирующие на определенные раздражители или комплекс раздражителей. Это есть суммирование различных структур памяти, которые способны быстро и четко отреагировать на группу раздражителей. Чем сложнее организм, тем больше у него таких структур, тем четче и быстрее он может реагировать. Вот эта реакция структур и есть, с нашей точки зрения, то отношение к новому раздражителю, которое становится психическим продуктом.

Почему именно «весь организм»? Потому что только на организменном уровне решаются кардинальные вопросы, а именно, что более всего требуется для него для сохранения равновесия (гомеостаза) с окружающей средой. Поэтому чем сложнее организм, тем он более «разборчив» в выборе дальнейших действий, а это определяется психикой, т.е. его сознательной и подсознательной деятельностью. Только **весь организм целиком** может «отобрать» вариант, который необходим более всего для жизнедеятельности организма. Только он может «критически посмотреть» на деятельность мозга в соответствии с возникшей информацией, мотивированной определенной ситуацией. И только тогда может возникнуть психика как **отношение всего организма к определенным структурам мозга и различным внешним и внутренним раздражителям.**

Появляется отношение, но не просто отношение, а критическое, т.е. организм решает, насколько эти структуры из памяти соответствуют его требованиям на данный момент. Организм постоянно «критически» реализует только ту информацию из памяти, которая необходима на данный момент.

Но кто или что определяет, что нужно организму? Ответим: **мотивация**, которая постоянно присутствует и регулирует все процессы. Если организм голоден, задействованы те структуры из памяти, которые так или иначе связаны или с пищей или ее поиском. В ходе эволюции удовлетворение не только вегетативных потребностей, но и более сложных (например, духовных) может стать источником мотивации и модуляции этих процессов. Одним из главных мотиваторов поведения человека может стать его любознательность.

Но отношение одного к другому не может быть материальным, поэтому мысль, конечно – идеальный продукт головного мозга, хотя рождена вполне материальными процессами.

При переходе от примитивных существ к более сложным идет процесс расширения диапазона возможностей саморегулирующейся системы, что приводит к максимально большему подключению ней-

ронов. Чем больше будет их количество, тем больше вероятность сигнала отображения.

Что же такое Идеальное? Выше нами было изложено предположение, что наше «я» есть результат постоянно складывающегося отношения всего организма с нейродинамическими процессами, протекающими в первую очередь в головном мозге, причем это отношение всегда «критическое», определяемое степенью целесообразности той или иной деятельности с целью сохранения гомеостаза.

Со времени пробуждения психики в организме главным образом, в головном мозге, складываются структуры, реагирующие на определенные раздражители или комплекс раздражителей. Это есть суммирование различных структур памяти условных рефлексов и иных взаимосвязей (возможно, гормональных систем), которые способны быстро и четко отреагировать на группу раздражителей.

Память есть постоянное поддержание физиологическими структурами облегченного пути контакта с другими структурами. Чем легче этот путь, тем лучше память. Структуры памяти, объединяясь, создают комплексы, которые и реагируют на раздражители. Но и комплексы взаимодействуют между собой. Вот здесь-то и складывается «критическое» отношение между ними. Эти комплексы постоянно «проверяют» друг друга.

Но это отношение в разные моменты времени далеко неодинаково, оно весьма **избирательно**, так как прежде всего складывается отношение с теми структурами, которые наиболее «востребованы» в данный момент организмом. Но «отбор» этих структур идет не путем обмена потоков импульсов, а путем взаимодействия всех или большинства структур организма, связанных с тем или иным событием, могущих в какой-то мере нарушить гомеостаз в нем. В результате этого возможно такое взаимоотношение между информационными системами, где одна может «критиковать» другую, оппонировать ей, возникают «я» и «ты» или двойное «я». Эти системы постоянно конкурируют друг с другом, и какие-то из них в определенный момент «побеждают», причем «побеждает» та, которая в наибольшей степени отвечает основной задаче, стоящей перед субъектом.

Когда говорится об отношении всего организма к той или иной информационной системе, то, конечно, подразумевается не реакция каждой его клетки, речь идет об отношениях, складывающихся между теми структурами организма, которые так или иначе были когда-то связаны с информацией, пришедшей через афферентную систему. Если же память никак не связана с поступившей информацией, то в зависимости от силы раздражителя возникает соответствующей степени ориентировочная реакция организма.

Так что же это за отношения? Однако прежде вспомним о том, что есть два способа принятия решения: алгоритмический и эвристиче-

ский. Мозг работает по второму типу, машина – по первому. Принцип вероятности и «размытости» позволяет мозгу решать значительно большее количество задач, чем это может машина. Поведенческие реакции в соответствии с вероятным прогнозом помогают ему резко уменьшить число ошибочных решений и являются эффективным средством приспособления к окружающей среде.

Отношения, о которых упоминалось выше, – суть отношений между раздражителями, действующими в данный момент на организм, и теми структурами памяти, которые «ответственны» за реакцию организма на эти раздражители. **Проще, это структуры, которые ведут поиск следов этого раздражителя или его аналогов по эвристическому типу,** заложенных в памяти. Здесь напрашивается некая аналогия с компьютером, который способен исправлять оператора, находить **«мгновенно»** ошибку в его действиях.

Выше упоминалось также о неких комплексах структур, которые реагируют на внешние воздействия. Что это за комплексы? Компьютер, который является тоже «думающим», хотя это и машина, может вмешиваться в рабочий процесс. Почему? В компьютере уже сформированы такие структуры с участием «памяти», которые моментально реагируют на какие-либо ошибки в этом процессе. Нечто подобное тоже может иметь место в психике: здесь тоже есть сформированные структуры для быстрой реакции на различные воздействия. (Сравнение, конечно, грубое, но в какой-то мере отражает суть явления).

В чем может быть аналогия? В компьютере, как и в мозге, раздражители (или сигналы) вступают в контакт с определенными ранее сложившимися комплексами структур памяти, где любые раздражители, входя в анализирующие структуры, последовательно или одновременно, соприкасаются с ними. На каждый импульс (сигнал) мозг реагирует той структурой, которая в наибольшей степени соответствует этому сигналу, т.е. лучше других структур способна восстанавливать гомеостаз, т.е. лучше других структур «оценить» эти раздражители. «Отработка» таких структур требует времени. Если раздражитель новый, ранее не действовавший, в организме возникает замедленная реакция («что такое?» по И.П.Павлову), идет поиск в «багаже» памяти возможных «следов» этого неизвестного раздражителя (с наименьшей затратой энергии).

Ранее говорилось, что когда задействован комплекс афферентных систем, генерируется идеальный продукт. Почему так? Обратимся к тому, что было сказано выше, а именно, как возникают ощущения: чем больше афферентных систем участвуют в реакции организма, чем сложнее их взаимодействие, «взаимокритика», что в свою очередь является результатом эволюции, тем больше вероятность появления субъективного элемента реакции – ощущения. В работе головного мозга, где по подсчетам специалистов при наличии миллиардов



нейронов и  $10^n$  (причем это  $n$  – очень большая величина) взаимоотношений между ними, вполне вероятен процесс какого-то особого, молниеносного проявления ответа сложного организма на внешние и внутренние раздражители. Поэтому это «субъективное» проявление организма или психический акт присутствуют у организмов, находящихся на высоких ступенях развития, что и подтверждается многочисленными наблюдениями.

Между отдельными материальными фрагментами возникают отношения, которые постепенно переходят в иное качество – цепочку логических умозаключений. Получается, что психика – это цепь сплошных отношений одних структур к другим, а также всего организма к отдельным структурам. Поэтому при рассмотрении нематериальных процессов иногда забывают, что нельзя их представлять в состоянии некоего покоя, как что-то неподвижное, как сумму неподвижных рисунков. Нематериальное возможно представить себе только **в движении**, как динамический процесс. Здесь очень большое значение имеет **скорость** самого процесса. Действительно, если физиологическое состояние в какой-то момент и можно представить недвижимым, конечно в виртуальном смысле (т.к. и это трудно вообразить), то психика вне движения не существует, так как она может быть **только в динамике и притом ее процессы протекают с большой скоростью**.

Результатом перехода к быстрой, практически мгновенной реакции и был переход от обычных скоростей биологических процессов к иным скоростям. Эти скорости могли возникнуть только как результат взаимодействия уже сложившихся в процессе фило- и оттогенеза комплексов в головном мозге. Здесь уже нет долгого процесса биологического взаимодействия – результата биологических реакций, здесь есть только результат некоего «взаимодействия», осуществляемого практически мгновенно.

Может ли машина достигнуть такого уровня совершенствования, что о ней можно будет сказать, что это – уже субъект? Выше отмечалось, что это может произойти тогда, когда у нее появится **критическое отношение** к тем процессам, которые в ней представлены. А возможно ли это? Когда мы говорим: «Я знаю, что я знаю», т.е. когда констатируется два «я», это значит, что в мозге присутствуют как бы два «интеллекта», и каждый критически смотрит на другой. В машине в отличие от организма мало емкости, так как она состоит совершенно из другого материала. Организм в процессе эволюции достиг такого совершенства, что **на очень малом пространстве** (несоизмеримом с машинным!) могут протекать такие же и гораздо более сложные процессы, чем в машине. Машине требуется очень большое пространство, чтобы принять такое количество афферентных систем, как принимает мозг. К тому же машина, фактически, только **мозг**, но

без тела. У нее поэтому нет «заинтересованности» в стабильности гомеостаза, нет мотивации для каких-либо действий для сохранения себя, так как мозг сам по себе не нужен, он существует для **обслуживания** организма. Организм состоит из клеток – чрезвычайно совершенных «малых» организмов, дающих колоссальную экономию объема. В машине пока этого нет.

Далее, естественно, возникает вопрос: если мозг представляет собой машину или хотя бы сходен с ней, то какого рода эта машина? Последнее время часто можно слышать о проведении аналогии между мозгом и электронной **цифровой** вычислительной машиной. Однако даже неквалифицированному в этом вопросе человеку ясно, что механизмы достижения порой и сходного эффекта совершенно различны.

Биологические механизмы, лежащие в основе работы мозга, конечно, ничего общего не имеют с цифровой вычислительной машиной. И все-таки многие специалисты в области вычислительной техники продолжают утверждать, что мозг и машина часто выполняют **одну и ту же работу**, хотя и разными путями.

Почему же именно человек и высокоорганизованное животное могут выстроить отдельные фрагменты мысли в единую стройную логическую систему? Ведь генетически они получают **только потенцию к логическому мышлению**, которую надо **реализовать**. А как происходит реализация? В умении быстро увидеть сходство и различие в предметах, быстро обнаружить причинно-следственные связи и спрогнозировать события.

На самом начальном этапе жизни психика у ребенка почти отсутствует, прежде всего потому что для нее необходима сравнительно богатая память, которая может возникнуть только на основе взаимодействия мозга с внешней средой. Постепенно, под влиянием мотивационного фактора мозг начинает все шире реагировать на внешний мир, по мере поступления информации постепенно возникают структуры памяти. Вначале они слабо взаимодействуют между собой и у ребенка еще отсутствует логика в поведении. Однако со временем, путем огромного количества проб и ошибок начинают складываться некие комплексы памяти, отвечающие определенным образом на определенные раздражители, и вот уже начиная с этого уровня развития вырабатываются отношения между структурами памяти, которые отвечают требованиям выполнения определенного поведенческого акта, а для этого нужна логика.

По мере развития и совершенствования психики, главным образом путем накопления информации и идет ее «распределение» с целью осуществления какой-либо функции организма. На этой основе и складывается логика поведения, но при обязательном участии генетического материала, так как без него представитель определенного вида не может стать таковым.

За это время скорость взаимодействия структур памяти постепенно увеличивалась за счет увеличения количества афферентных систем (физиологический аспект этого процесса) и роста нейронной сети (морфологический аспект этого процесса). Действительно, размер мозга человека относительно его тела не на много превышает это же соотношение у животного, хотя степень развития его интеллекта в сравнении даже с таким развитым существом как, например, шимпанзе, намного превышает уровень последнего. Возможно, что наряду с бурной эволюцией умственных способностей человека вследствие значительного увеличения скорости взаимодействия структур его памяти в морфологическом плане его мозг претерпел процесс уменьшения за счет снижения массы и объема отдельных его структур. Это не более чем гипотеза, но мы будем все же ее придерживаться. Почему? Если проследить процесс эволюции, то заставляет обратить на себя внимание тот факт, что биологический субстрат (в том числе и масса его) в значительной мере уменьшается по сравнению с неживыми структурами, которые могут в какой-то мере выполнять схожую функцию. Возьмем живую клетку, ведь чтобы ее смоделировать из неживого вещества потребуется огромное пространство, совершенно несоизмеримое с размерами клетки, а чтобы смоделировать мозг человека из механизмов, которые лежат в основе электронной техники, то даже трудно представить, сколько места займет эта модель.

Известно, что человек и высшие в психическом отношении существа возникли не сразу. Они прошли длительный путь эволюции, пройдя многие стадии развития. Но все или почти все, что они «проходили» в той или иной мере «закреплено» в генетическом коде ДНК и в первую очередь то, что нужно человеку. Таким образом, естественный интеллект содержит неисчислимое количество информации, полученной от своих биологических предков. Поэтому возможностей у естественного интеллекта несравненно больше, чем у искусственного. Если машина запрограммирована совершать вычислительный процесс, она никогда не будет писать стихи или играть в шахматы, в то время как возможности естественного интеллекта огромны и совершенно непредсказуемы. Поэтому, с нашей точки зрения, генетическая программа, идущая от биологических предков, является одним из главных препятствий для достижения искусственным интеллектом возможностей естественного. Но это – на сегодняшний день, что будет потом, покажет будущее.

---

# ФУНКЦИОНАЛЬНАЯ СТРУКТУРА БЕССОЗНАТЕЛЬНОГО И ВОЗМОЖНОСТЬ ФОРМИРОВАНИЯ НОВЫХ ПРИНЦИПОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*О.В. Панов*

О.В. Панов

Для новых направлений разработки проблемы искусственного интеллекта важное значение имеет исследование особенностей информационных процессов на бессознательном уровне психики, поскольку они составляют необходимый, крайне существенный компонент всякой сознательной деятельности.

Вопрос о соотношении сознательного и бессознательного в человеческой психике является весьма сложным, включает многочисленные планы и аспекты, которые требуют теоретически корректного соотнесения друг с другом (см.: Дубровский Д.И. Феномен бессознательного и познавательные процессы // *Философские науки*, 1986, № 1).

Сознательные состояния обладают неотъемлемым качеством субъективной реальности (субъективного переживания). Это качество выражает непосредственную представленность информации для личности и способность оперирования ею. Конечно, эта информация воплощена в мозговых кодах, но личности она дана как бы в «чистом» виде (в том смысле, что ее мозговой нейродинамический носитель никак не отображается; мы ничего не знаем, не чувствуем, что происходит в нашем головном мозгу, когда переживаем образ предмета, осознаем движение собственной мысли и направляем ее развитие). Такой способ представленности информации как раз и составляет важнейшую черту психического отображения и управления, связан с личностным Я, которое реализует акты осознания. В подобном случае правомерно использовать понятие сознания (сознательной деятельности) в узком смысле, имея в виду длящееся в данном интервале осознаваемое «содержание», которое дано личности актуально и в рефлексивной форме (хотя за ним всегда присутствует его диспозициональное и арефлексивное основание).

В отличие от этого бессознательные информационные процессы протекают в «темноте», представляют арефлексивный и диспозициональный уровни психического отображения и управления. Общеизвестно, что эти процессы, несмотря на их скрытость от осознания, играют исключительно важную роль во всех формах и проявлениях сознательной деятельности, начиная от переживания чувственных образов и кончая высшими побуждениями нашего Я. Процессы, совершаю-

щиеся на бессознательном уровне, нередко оказывают решающее влияние на характер оценки наличной информации и выбора сознательного действия, играют определяющую роль в творческих актах.

Когда речь идет о моделировании особенностей бессознательных психических процессов то следует учитывать не только их когнитивные («содержательные») аспекты, но также мотивационные, оценочные, целеполагающие факторы, присущие им способы выборки информации из памяти, принятия решения и выхода на эффекторные пути для производства действия.

Разумеется, весьма существенны такие оперативные параметры как скорость переработки информации на бессознательном уровне и те структурно-динамические показатели, которые определяют «логику» («приемы») переработки информации на этом уровне. Последнее особенно важно, поскольку есть основания полагать, что на бессознательном уровне дело не ограничивается двузначной логикой и вероятностным принципом, что тут действуют иные, во многом пока еще нам непонятные, логические и внелогические способы переработки информации.

Всё это присуще естественному интеллекту, и указанные специфические особенности бессознательного должны быть выявлены и четко осмыслены, если мы хотим серьезно продвинуться в области искусственного интеллекта. Ведь искусственный интеллект создается по образу и подобию естественного интеллекта, более того, должен избегать недостатков последнего и в чем-то превосходить его.

Даже если мы ограничимся сознанием в узком смысле, то и здесь сохраняются значительные трудности его основательного исследования.

Однако в этой области достигнуты существенные результаты. Создан ряд удачно функционирующих итеративных моделей, которые воспроизводят последовательную логическую цепочку получения выводов из исходных данных. Есть также модели, копирующие обратный процесс.

В области эвристического моделирования (воспроизведения процессов принятия решений на основе интуиции, вероятностной логики) кроме самообучающихся нейронных сетей практически нет ничего более или менее близкого к естественному процессу мышления. Но даже эти модели недостаточно полно удовлетворяют практическим запросам.

Скорее всего, отсутствие удачных эвристических моделей обусловлено отсутствием должного количества объективных знаний о структуре бессознательного, процессах его влияния на сознание и о самом сознании. Ещё не так давно не было достаточно объективных инструментов исследования бессознательного и его связей с сознанием (психоанализ и другие методы психотерапии и психологии не могут строго считаться объективными инструментами исследования

бессознательного, так как они не располагают средствами непосредственного доступа в бессознательное, а сознание как исследуемого, так и исследователя способно вносить существенные искажения и в процесс отбора исходных данных, и в их интерпретацию).

Наличие двух «чёрных ящиков» с промежуточным влиянием сознания и эмоций на результаты знания о неосознаваемых процессах и их влиянии на создаваемую деятельность, полученные старыми способами, мало достоверными и не пригодными для создания адекватных математических моделей функционирования бессознательного, подсознательного и их взаимодействия с сознанием.

В последние 10 – 15 лет благодаря распространению быстродействующих компьютеров стало возможным «обойти» сознание, его психологические защитные механизмы, общаться с бессознательным напрямую и при этом обрабатывать полученные «сырые данные» не вручную, а с помощью тех же машин. Оптимизированные для сознательной обработки данные анализируются исследователем, выдвигаются гипотезы, которые можно проверить в модельном эксперименте.

При накоплении достаточного количества статистически достоверного материала появляется возможность создания адекватной модели функционирования бессознательного, его связи с сознанием и поведением. Гипотеза проверяется в «слепом» эксперименте. Данные компьютерного анализа определённых областей бессознательного, полученные от респондента о котором нет никаких предварительных объективных данных, проходят модельную обработку. Полученная информация сравнивается со сведениями из жизни исследуемого, полученными из объективных источников (трудовая книжка, история болезни и т.д.). Процент совпадения данных полученных обеими способами позволяет говорить о степени адекватности модели.

Конечно, этот процент никогда не может равняться 100 из-за следующих объективных причин: 1) объём информации представленный в бессознательном и скорость его переработки на этом уровне несоизмеримы с возможностями сознания; 2) не всё, что есть в бессознательном, проявляется объективно (поведение, определенные действия во внешней среде, речевые высказывания).

Что же можно сказать на основании работы с этими новыми инструментами о структуре бессознательного и процессах происходящих в нем?

1. Бессознательное нелинейно и имеет на несколько порядков больше «степеней свободы» по сравнению с сознанием.
2. Бессознательное не ограничено рамками индивида – в том смысле что оно имеет структуры постоянно подключенные к внешним источникам информации, поступление которой не ограничивается обычными пятью каналами восприятия (зрение, слух, обоняние, осязание, кинестетика).

3. Поступившая информация в бессознательном неуничтожима известными нам способами (забывание, локальные травмы, заболевания мозга), теряется только путь, доступ к информации.
4. Рождение человека не является отправной точкой начала записи индивидуальной информации. Матрица будущей личности закладывается раньше и формируется из генетической информации и отношения родителей к зачатому ребёнку или даже к возможности его зачатия.
5. Первичное наполнение матрицы бессознательного начинается до рождения, оно может противоречить вторичному ее пополнению и создавать противоречия в структуре субъективной реальности.
6. Скорее всего, почти всё что есть в бессознательном представлено в сознании, но в символической (зашифрованной) форме. Поэтому требуется узнать язык перевода неосознаваемых данных в осознаваемые и наоборот, чтобы правильно сформулировать условия для построения математических моделей искусственного интеллекта и формирования способов общения искусственного интеллекта с естественным.
7. Объектами для изучения «хороших» эвристических моделей мышления должны стать выдающиеся в различных областях деятельности люди, так как именно они изначально обладают пока неизвестной, но продуктивной структурой бессознательного или доступом к тем его областям, которые ответственны за выдающиеся творческие результаты. Все остальные могут являться группой сравнения.

Резюмируя всё изложенное выше можно сказать, что качественный прорыв в развитии искусственного интеллекта возможен только при качественно новом, углублённом информационном исследовании естественного интеллекта, человеческой психики – как в сфере сознания, так и особенно в ее бессознательной области. Эти качественно новые данные должны послужить основой для создания математических моделей искусственного интеллекта, более приближенных по своим свойствам к процессам функционирования естественного интеллекта. В этом отношении теоретической основой и инструментом для исследований может служить образно-лингвистический анализ, использующий в качестве инструмента анализаторы семантических систем.

---

# КВАНТОВЫЙ КОМПЬЮТЕР И КВАНТОВЫЙ МОЗГ

*Л.Г. Антипенко*

Тематика данной статьи посвящена теоретическому обоснованию гипотезы, согласно которой идеальная мыслительная деятельность человеческого мозга протекает по тем законам, которыми управляются вычислительные процессы в квантовых компьютерах. Вероятно, мы имеем дело уже с теоретическим фактом – фактом функционального тождества квантового компьютера и церебральной системы человека.

Проверку этого факта проще всего будет изложить при рассмотрении квантовых процессов, описываемых уравнением Шредингера, под тем углом зрения, когда они предстают как процессы вычислительные, приводящие, при их регистрации, к получению определённой (нужной экспериментатору) информации.

Что для достижения намеченной здесь цели предстоит сделать? Для этого нам придётся указать на существование различных каналов передачи информации, в зависимости от которых она подразделяется на классическую и квантовую. Нам придётся ознакомить читателя с квантовой спецификой пространства и времени, без чёткого понимания которой вообще невозможно получить правильное представление о предмете той научной дисциплины, что называется *физикой квантовой информации* [1]. Затем мы приступим непосредственно к описанию квантовых вычислений с тем, чтобы сравнить их с работой мозга. В обычной математической логике употребляются понятия объектного языка и метаязыка. В вопросах квантовых вычислений необходимо будет ещё апеллировать к понятиям наблюдателя и метанаблюдателя, которые служат в качестве аналогов объектного языка и метаязыка.

Начнём с разбора вопроса о пространстве и времени.

## **§1. Квантовая специфика пространства и времени в квазиклассическом приближении**

Первое её описание мы находим в трудах П.А. Флоренского. Наибольший интерес в этом плане представляют лекции автора по теории искусства, относящиеся к 20-м годам прошлого столетия [2]. В них Флоренский показал вообще несостоятельность классической концепции пространства и времени как с точки зрения современной физики, так и с точки зрения других – гуманитарных и исторических



– наук. Применительно к пространству сформулированные им принципиальные положения выглядят так.

1. «В действительности (как она предстаёт перед исследователем. – Л.А.) нет ни пространства, ни реальности, – нет, следовательно, также вещей и среды. Все эти образования суть только вспомогательные приёмы мышления ...» [2;83].

2. При рациональном познании свойства действительности описываются с помощью рациональной модели. Они «куда-то должны быть помещены в модели, т.е. в пространство, вещи или в среду. Но куда именно – это не определяется с необходимостью самим опытом, и зависит от *стиля* мышления и вообще от строения мышления, а не от строения опыта» [2;83].

3. В общем случае свойства действительности распределяются таким образом, что часть их относится к среде, а другая часть к пространству- времени. Граница между средой и пространственно- временным многообразием подвижна и в значительной мере условна, но *проведение* такой границы необходимо.

Можно поэтому смотреть на вещи как на «складки» или «морщины» пространства, места его особых искривлений (там же).

Сказанное о пространстве относится *mutatis mutandis* и ко времени, так как, согласно теории относительности, время образует вместе с пространством единый четырёхмерный пространственно-временной комплекс. Но у времени есть ещё и своя специфика, о которой речь ниже.

Указывая на подвижный и условный характер границы между пространством–временем и вещью или средой, Флоренский предупредил, однако, о недопустимости релятивистских ошибок. Выбрав однажды границу между тем и другим, мы должны считаться с ней в каждой конкретной области исследований и не менять по своему субъективному произволу. В квантовой физике вещь (частица) сначала растворяется в конфигурационном пространстве, а затем она выделяется из него в процессе измерения. Конкретнее об этом – в следующем параграфе.

## **§2. Квантовая спецификация пространства–времени с учётом свойств физического вакуума**

Каждый студент-физик прекрасно знает о том, что квантовая механика имеет дело с конфигурационным пространством, которое символизируется волновой функцией. Изменение во времени волновой функции есть не что иное, как временная эволюция конфигурационного пространства. Выше уже было сказано о том, что в процессе измерения частица перестаёт быть свойством («складкой») пространства, она разделяется с ним. Но свойства конфигурационного пространства срastaются ещё со свойствами физического ва-

куума, что вносит в него фактор нелокальности (дальнодействия), смысл которого понятен далеко не каждому физика. А без уяснения этого смысла трудно было бы понять и специфику работы квантового компьютера. Посему придётся дать несколько дополнительных разъяснений относительно понятия конфигурационного пространства в его полном объёме.

Если конфигурационное пространство строится исходя из наличия  $n$  частиц, то его размерность будет выражаться числом  $3n$ , поскольку каждая частица имеет три степени свободы передвижения (три координаты, три координатных оси в пространстве). К трём данным степеням свободы – их принято называть внешними – добавляются ещё внутренние степени свободы каждой частицы, обусловленные наличием у неё спина. Последние используются, как будет видно далее, при организации квантово-компьютерных вычислений. Но пока нас интересует другое, а именно то обстоятельство, что при описании конфигурационного пространства никак нельзя обойтись без мнимой единицы. Казалось бы, волновое пси-поле ( $\Psi$ -функция) не должно принципиально отличаться от, скажем, электромагнитного поля, но во втором случае описание (распространения) поля обходится без мнимых и комплексных чисел, а в первом случае добиться такого описания невозможно. В чём здесь дело? А дело в том, что в квантовой механике нельзя сохранить приверженность к четырёхмерной теории близкодействия, которую всё ещё и до сих пор многие ищут. На эту особенность квантовой теории в своё время обратил внимание П. Эренфест (1880–1933), хотя понимания от других в этом вопросе при своей жизни он не добился.

Прочитываем его высказывания. Он, в частности, писал: «Классические уравнения Максвелла представляют собой обычную теорию поля в четырёхмерном пространстве  $x, y, z, t$ . В первоначальной концепции де Бройля казалось естественным, что «волны материи» также должны подчиняться четырёхмерной теории поля, чётким подтверждением которой считались интерференционные эксперименты простейшего типа. Но мы лишились (быть может, не навсегда?!) веры в возможность такой теории поля, после того как Шредингер для описания взаимодействия  $n$  электронов должен был использовать обобщение  $\Psi$ -функции на  $3n$ -мерное «конфигурационное» пространство, причём все попытки сохранить четырёхмерный континуум потерпели неудачу» [3;173]. Далее Эренфест разъясняет, что наличие дальнодействующей связи, скажем, между двумя электронами в конфигурационном пространстве не зависит от расстояния между ними. Достаточно зафиксировать в каком-то месте пространства один электрон, как мы получим информацию, заключённую в волновой функции, о месте расположения другого электрона. Такая связь, говорим

мы теперь, реализуется посредством той среды, в которой нет поляризации на пространство (протяжённость) и время (длительность).

Итак, когда строится и используется в квантовой теории конфигурационное пространство, частицы вещества не только растворяются в этом пространстве, но ещё и «кооптируются» в физический вакуум. Через вакуум осуществляется их дальное действие, проявляющееся при регистрации. Мнимая единица служит формальным признаком участия физического вакуума в таких процессах. Но за мнимой единицей или, вообще говоря, за комплексным числом следует его комплексно-сопряжённый двойник, волновой функции соответствует комплексно-сопряжённая функция. В релятивистской квантовой теории комплексно-сопряжённые функции описывают два разных состояния (движения) частиц. О том, чем они принципиально отличаются друг от друга, мы расскажем далее. Сейчас только заметим, что с конфигурационным пространством связан *целостный* комплекс физических величин, так что его преобразование, отвечающее переходу от волновой функции к функции, с ней комплексно-сопряжённой, приводит к преобразованию таких попарно взятых величин, как энергия и время, импульс и координата. Двум комплексно-сопряжённым функциям соответствуют две разные компоненты времени. Одну из них называют энтропийной, другую – антиэнтропийной, или эктропийной. Результат квантово-вычислительных процессов зависит от того, в какой системе они протекают, с каким из двух противоположных типов конфигурационного пространства приходится иметь дело.

Знакомство со спецификой конфигурационного пространства позволяет уточнить статус подвижной границы между вещью и пространством-временем в рамках квантовой теории. Здесь эта граница перемещается в цепочке «субъект – вещь (квантовый объект) – конфигурационное пространство». Она поэтому приобретает чётко выраженный гносеологический характер. Физики называют её *подвижной гранью*, которая в процессе квантово-механического измерения отделяет либо объект (микросистему) от прибора, либо прибор от (сознания) наблюдателя. В общем, как неоднократно разъяснял В. Паули, само сознание наблюдателя «требует, чтобы между субъектом и объектом можно было провести грань, существование которой диктуется логической необходимостью, тогда как положение её остаётся до известной степени произвольным» [4;63]. (Правильнее было бы говорить не о логической, а о гносеологической необходимости).

### **§3. Об идейных истоках концепции квантово-компьютерных вычислений**

Концепция квантово-компьютерных вычислений является частью физики квантовой информации. Для её изложения требуется язык теории квант, который дополняется определённым арсеналом терминов и

понятий, необходимых для описания работы квантового вычислительного средства. Укажем на узловые моменты квантово-компьютерной идеологии.

Исходное понятие – квантовое состояние движения с его отличием от состояния движения (покоя) в классической физике. Как в классической физике, так и в квантовой механике, состояние движения частицы изменяется под воздействием силового фактора, или фактора силового поля. В классике фактор этот представлен в виде функции Гамильтона (или функции Лагранжа). В квантовой механике функция Гамильтона заменяется оператором Гамильтона (гамильтонианом), который действует на волновую функцию. Способ изменения волновой функции во времени описывается уравнением Шредингера.

Говорят, что частица, состояние которой находится под влиянием силового фактора, представленного гамильтонианом, взаимодействует с соответствующим источником силового поля. При этом может случиться так, что источником силового поля выступает частица или античастица того же сорта, что и частица, находящаяся под наблюдением. Тогда у двух таких частиц (их может быть больше двух) появляется общность, позволяющая описывать их единой волновой функцией, т.е. объединять в одной функции состояния обеих частиц. Такие квантовые состояния Э.Шредингер назвал в своё время скрещёнными (от нем. *Verschränkung* – скрещение) [5]. К сожалению, в русском языке стали использовать менее удобную терминологию, в которой фигурируют «перепутанные состояния». Эти издержки перевода на русский язык английского *Entanglement*, что буквально означает *запутанность, затруднительное положение*, внесли немалую путаницу в умы физиков. Более удачным мы считаем термин «сцеплённые состояния», который фигурирует в русском переводе книги Э. Стинга [6]. Им мы и будем пользоваться в дальнейшем.

«Сцеплённые состояния» – ключевой термин в физике квантовой информации. Такие состояния фигурируют в известном мысленном эксперименте, сформулированном Эйнштейном, Подольским, Розеном в 1935 году. Этот эксперимент рассматривается как выражение своеобразного парадокса, так как в нём обнаруживается, при строгом следовании законам квантовой теории, как раз наличие нелокального влияния одной части сцеплённого состояния на другую. Такое влияние проявляется в процессе измерения. Здесь, однако, несколько иначе высвечивается феномен нелокальности, о котором шла речь в предыдущем параграфе. Но понятием сцеплённых состояний удобно пользоваться, поскольку оно позволяет уже непосредственно подключиться к проблематике квантово-компьютерных вычислений.

Квантовый компьютер реализуется посредством фиксированного множества сцеплённых частиц, каждая из которых может находиться только в двух состояниях, символизируемых обычно посредством нуля

и единицы, т.е. тех цифровых элементов, которые используются для двоичного выражения чисел. Количество сцеплённых частиц называют обычно квантовым регистром. Каждая ячейка регистра отождествляется с частицей. Регистр квантового компьютера отличается от регистра компьютера классического тем, что в нём фигурирует одновременно суперпозиция всех возможных состояний, реализуемых квантовым регистром, и только при измерении фиксируется в качестве результата наблюдения одно из них. Такая суперпозиция может выполнять вычислительную задачу только при условии, что она подвергается унитарным преобразованиям, т.е. тем преобразованиям, что имеют место при унитарной эволюции системы. (Напомним, что условие унитарности состоит в равенстве единице суммы всех тех вероятностей, которые соотносятся посредством амплитуд вероятности с каждым членом, входящим в квантовую суперпозицию).

В Западной литературе констатируется как уже твёрдо установленный факт, что квантовое зацепление является информационным ресурсом, и в связи с этим ставится задача выявления ответов примерно на такие вопросы:

Может ли Природа, изучаемая физикой, рассматриваться по сути как информационный процессор? И можно ли из этого извлечь пользу?

Может ли классический или квантовый компьютер имитировать все процессы Природы? [6; 45].

При ответе на второй вопрос предпочтение отдаётся как раз квантовому компьютеру, ибо в квантовых алгоритмах логические операции подчиняются критерию унитарных преобразований, что ставит их в один ряд с физическими закономерностями, описываемыми уравнением Шредингера или его релятивистскими аналогами. Как отмечает Р.Джозса (R.Jozsa), один из авторов книги «Физика квантовой информации», описание квантовых вычислений удобно формализовать в терминах модели, выстраивающей параллель с формализмом классических вычислений. «По сути, в квантовом случае элементы памяти в компьютере – это кубиты, а не биты, а логические операции – это унитарные преобразования, а не булевы операции классического вычисления. Можно утверждать, что модели такого рода достаточно, чтобы описать любой квантовый процесс» [1;139]. (Поясним, что кубиты являются квантовыми аналогами битов, от последних они отличаются тем, что могут находиться в промежуточном состоянии между состоянием «да» и состоянием «нет» с разными степенями вероятности).

Для нас представляет особый интерес задача выяснения связи теории квантовых вычислений с философией и практикой математических доказательств. Есть такие математические доказательства, которые проводятся методом рекурсивных вычислений, и есть такие теоремы (к примеру, Гёделевы теоремы неполноты), которые выходят за

пределы рекурсивных вычислений. Может быть, квантовые вычисления смогут пролить свет на то, как совершается выход за такие пределы? Ниже будет указано направление поисков ответа на данный конкретный вопрос. Вообще же, как указывают авторы вышеупомянутой книги, проведение вычисления, приводящего к определённой результату, эквивалентно доказательству того, что наблюдаемый результат является одним из возможных результатов вычисления. Поскольку мы можем описывать операции компьютера математически, такое доказательство всегда может быть переведено в доказательство некоторой математической теоремы. В классическом случае, при отсутствии интерференционных эффектов, всегда можно проследить шаги вычисления и таким образом произвести доказательство, которое удовлетворяет классическому определению: последовательность предложений, каждое из которых есть либо аксиома, либо следует из предыдущих предложений в последовательности в соответствии со стандартными правилами логических умозаключений. В квантовом случае такое определение уже не работает. «В дальнейшем, – указывают В.Дойч и А.Экерт, – доказательство должно рассматриваться как процесс – само вычисление, а не его запись, – поскольку мы должны принять, что в будущем квантовые компьютеры будут доказывать теоремы методами, которые ни человеческий мозг, ни какой-либо другой арбитр не будет в состоянии проверить шаг за шагом, поскольку если бы «последовательность предложений», соответствующая такому доказательству, была бы распечатана, то бумага много раз заполнила бы наблюдаемую вселенную» [1;138].

С позиции результатов наших собственных исследований дело обстоит так, что человеческий мозг работает как квантовый компьютер, и поэтому проверить шаг за шагом его работу, проконтролировать его деятельность в принципе невозможно. Можно, конечно, нарушить его нормальную работу, но не более того. Для обоснования этого тезиса нам придётся ближе познакомиться с понятием информации, с её термодинамическими характеристиками, которые сказываются как на специфике работы квантовой вычислительной системы, так и мозга.

#### **§ 4. Информационно-термодинамические характеристики работы вычислительных систем двух разных типов**

Информационно-термодинамический подход к изучению работы вычислительных систем позволяет проделать сравнительный анализ функционирования трёх интересующих нас систем: церебральной системы человека, классического компьютера (представленного, скажем, в виде идеальной машины Тьюринга), квантового компьютера. Первый вопрос, который встаёт при изучении работы мозга, состоит в том, как ему удаётся преодолевать энтропийный фактор, с которым несовместима присущая мозгу идеальная деятельность, направленная на логи-

ческую обработку информации, на проведение, порой, сложнейших математических вычислений и умозаключений, и т.п.

В опубликованной в 1971 г. книге Н.И. Кобозева «Исследование термодинамики процессов информации и мышления» намечены те общие рамки, в которых обычно ищут определение информации. Рамки эти автор соотнёс с двумя законами: 1) с законом энтропии для физико-химических систем, т.е. молекулярных множеств любого уровня; 2) с законом тождества для мышления [7;4]. Информация, указывал он, лежит между двумя этими законами, а именно: как физический сигнал она примыкает к закону энтропии, т.е. к исчислению вероятностей; как точно кодируемое сообщение она подчиняется закону тождества. Поэтому главной проблемой в термодинамике мышления является выявление тех условий, при которых молекулярный аппарат мозга, подчиняющийся закону энтропии, способен продуцировать процессы мышления в соответствии с законом тождества (там же, с.7).

Под законом тождества здесь понимается не просто логический закон тождества  $A=A$ , а принцип тождественных преобразований, скажем, принцип получения тождественно-истинных формул в исчислении высказываний. С теми или иными оговорками его действие распространяется на всю логику математическую деятельность мышления.

Неожиданным оказалось открытие, согласно которому антиэнтропийная характеристика мыслительной деятельности, протекающая по закону (принципу) тождества, была установлена в самой его структуре. В качестве образца такой логико-математической формы мышления был выбран как раз тот тип вычислений, практикуемых в элементарной арифметике, о котором мы уже упоминали выше. Речь идёт о *рекурсивных* вычислениях. Они получили статус особой математической дисциплины вместе с тезисом Чёрча, согласно которому объём понятия эффективной вычислимости совпадает с объёмом понятия рекурсивной вычислимости: всё, что поддаётся процедуре эффективной вычислимости, укладывается в рамки вычислимости рекурсивной.

Понятие рекурсивных вычислений и послужило отправным пунктом для дальнейших логико-математических изысканий, приведших к открытию термодинамических аспектов логико-математического мышления. Первым шагом на этом пути стали, конечно, доказанные в 1931 г. австрийским математиком К. Гёделем теоремы о неполноте, разрушившие иллюзию о самодостаточности формальных умозаключений в математике. Они справедливы для любой формально структурируемой системы, включающей в себя, по крайней мере, элементарную арифметику. При тщательном изучении гёделевых теорем выяснилось, что в них неявно используется термодинамическое понятие обратимости/необратимости [8; 207]. Процесс обратимый отличается тем, что при всех, описывающих его течение преобразованиях, в нём остается неизменным его энтропийный параметр. Фактор

же необратимости означает, что этот параметр не остаётся неизменным, а отклоняется либо в сторону увеличения (рост энтропии), либо в сторону уменьшения. Акты уменьшения энтропии, фиксируемые в той или иной системе, означают, что в ней действует некоторый агент, противостоящий её энтропийной хаотизации и даже повышающий уровень её организации.

Сравнение идеальной деятельности человеческого мозга с работой классического компьютера, функционирующего в соответствии с тезисом Чёрча, показывает, что компьютерные вычисления удовлетворяют критерию обратимости тех действий, которые выполняет (классический) компьютер. Но он (компьютер) не может выдать результат, знаменующий собой антиэнтропийный (эктропийный) скачок. В то же время мозг человека на выдачу такого результата способен, и об этом свидетельствуют как раз гёделевы теоремы неполноты.

Сравнительный анализ работы классического компьютера и компьютера квантового был бы мало продуктивным без учёта открытий Гёделя. Но их открытие – результат работы человеческого мозга. Они-то и дают возможность отождествить работу мозга и работу квантового компьютера. Почему? – Да потому, что в них мы находим ключ к пониманию того, как в упорядоченной системе рассуждений возникает новая информация, выходящая за пределы такой системы. Поясним кратко, как это происходит.

Речь идёт о рассуждениях, излагаемых на языке логики, языке исчисления предикатов. Рабочим инструментом в руках Гёделя послужило узкое, или первостепенчатое, исчисление предикатов. В теории рекурсивных вычислений и в формальных системах типа формализованной элементарной арифметики предикаты относятся только к целым положительным числам. Одноместными предикатами представляются свойства чисел, двухместными и многоместными – отношения.

Из того, что проделано Гёделем, ясно видно, что динамическая структура формальной арифметической системы характеризуется наличием в ней как обратимых, так и необратимых процессов. Над обратимыми процессами возвышается специфическая гёделева формула – показатель неполноты системы, – которая и символизирует антиэнтропийный скачок в системе. Конкретно устанавливаются следующие положения:

1) рекурсивно перечислимое множество доказуемых, средствами формальной аксиоматической системы, формул является неразрешимым;

2) множество всех дедуктивных выводов (цепочек формул), приводящих к доказуемым формулам, разрешимо в отношении этих формул;

3) не существует в рассматриваемой системе дедуктивного вывода, который привёл бы к гёделева формуле-истине.



(Читатель, желающий разъяснить для себя более подробно содержание п.3, может обратиться к соответствующей литературе, в частности, к с. 153—170 вышеупомянутой книги [8]).

Понятие разрешимости означает, что по отношению к данной формальной системе существует алгоритм, позволяющий определить, принадлежит ли та или иная, правильно построенная, формула к множеству доказуемых формул или не принадлежит. Если доказано, что такого алгоритма не существует, тогда уже встаёт вопрос, как связан этот факт с понятием необратимости. Это – один из центральных вопросов нашей задачи. Поэтому очень важно заранее понять, как идеи термодинамики проникают в теорию формальных систем. Путь этот пролагается именно посредством теории информации.

Дело в том, что в физической теории информации понятие (количества) информации существенно связано с противоположным ему понятием энтропии. Связь такая обусловлена тем, что, согласно определению информации, информация исчезает всякий раз, когда две ранее различавшиеся ситуации становятся неразличимыми. В тех физических системах, где отсутствуют силы трения, информацию невозможно уничтожить. Ведь при уничтожении информации должно быть рассеяно и, следовательно, обесценено некоторое количество энергии за счёт её перехода в тепловую форму. Простейший пример, заимствованный нами из статьи Ш.Г. Бенье и Р. Ландауэра «Физические пределы вычислений» (см. ж. «В мире науки», 1985, №9, с.24-34), состоит в следующем. Рассматриваются две легко различающиеся ситуации. В одной из них резиновый мячик поддерживается на высоте 1 м от пола, другой – на высоте 2 м. Мячики отпускаются и падают, а затем отскакивают от пола вверх. При отсутствии трения и при условии, что мячи абсолютно упругие, наблюдатель всегда сумеет сказать, каким было исходное состояние мячика (в данном случае на какой высоте он находился в начальный момент времени), поскольку мячик, упавший с высоты 2 м, отскочит выше, чем в случае, когда он падает с высоты в 1 м.

Однако при наличии сил трения при каждом отскоке мячей от пола рассеивается некоторое количество энергии. В конце концов мячи перестают прыгать и остаются лежать на полу. В таком случае мы уже лишаемся возможности определить, каковы были исходные положения того и другого мяча: мяч, упавший с высоты 2 м, будет полностью идентичен мячу, упавшему с высоты 1 м. Значит, в результате диссипации энергии, произошла потеря информации.

Примерно то же самое мы постигаем в гёделевых теоремах неполноты. Из-за того, что в формальной арифметической системе нельзя установить, является ли правильно построенная формула доказуемой или нет, все доказуемые формулы-теоремы уравниваются с правильно построенными формулами и теряются в более широком

множестве правильно построенных формул. Поэтому когда мы рассматриваем рекурсивно перечислимое множество доказуемых формул, без тех последовательностей (цепочек) формул, которые к ним приводят, мы утрачиваем часть информации относительно всего дедуктивного процесса в системе и превращаем его в процесс необратимый. Мы можем сделать его обратимым, когда каждую доказуемую формулу соотнесём с соответствующим ей дедуктивным выводом, т.е. с той цепочкой дедуктивно связанных звеньев, в которой она занимает место последнего звена. Разумеется, мы должны знать заранее о том, что каждое звено дедуктивного вывода представляет собой либо аксиому, либо заранее выведенную формулу.

Гёдель ввёл систему кодирования посредством натуральных чисел как самих доказуемых формул, так и тех цепочек доказательств, которые к ним приводят. При этом выяснилось, что кодовые номера тех и других находятся во вполне определённом арифметическом отношении. Затем он построил формулу, кодовый номер которой выпадает из данного отношения, хотя рекурсивная проверка показывает, что вновь найденная формула истинна. Так был установлен способ пополнения информации о свойствах чисел в арифметической системе, который выходит за рамки дедуктивного процесса её построения. А прирост информации означает уменьшение энтропии в системе. В системе имеет место, таким образом, необратимый процесс, но уже не за счёт увеличения энтропии, а за счёт её уменьшения.

Гёделево кодирование превратило всю формальную систему в систему наличной информации, которую можно было бы передавать по информационным каналам. В то же время автор теорем неполноты показал, как можно добиваться прироста информации и уменьшения энтропии в системе, используя возможности, заложенные в самом человеческом мозге. Выяснилось таким образом, что церебральные функции человека представляют собой нечто большее, нежели то, что может выполнять компьютер на основе классических алгоритмов (рекурсивных вычислений). В этом как раз и кроется одна из причин, почему возник интерес к *квантовым* компьютерам, квантовым вычислениям.

Специфика возникновения в квантовых системах новой информации (в виде информационных скачков) становится доступной пониманию, когда уясняется более общий вопрос о том, как с квантовой теорией (с квантовой механикой) совмещаются элементы классической (макроскопической) картины мира, которые и позволяют снимать информацию, вырабатываемую квантовыми системами. Физики согласны в том, что проникновение классической картины мира в изображение квантовых явлений обусловлено феноменом *декогеренции*, описываемой как переход когерентной суперпозиции в смесь, из которой затем выбираются доступные прямому наблюдению фрагменты

реальности. Феномен декогеренции обязан акту необратимости, которая выглядит поначалу довольно странной на фоне теории, построенной на обратимом уравнении Шредингера. Однако эта странность отпадает, как только мы начинаем смотреть на сам обратимый процесс по-новому, нежели это принято, т.е. рассматривать его как процесс, в котором энтропийные отклонения в ту или иную сторону (увеличение энтропии или уменьшение её) взаимно уравниваются.

При таком подходе к изучению явлений микромира становится понятно, что в когерентных, или волновых, процессах электроны ведут себя как безэнтропийные, т.е. нейтральные в отношении энтропии и эктропии, объекты. Но электрон имеет двойственную природу в отношении этих параметров и проявляет себя либо как частица энтропийная, либо как эктропийная, в зависимости от системы, в которой он регистрируется. В эксперименте Эйнштейна, Подольского, Розена описание физической реальности в терминах квантовой механики действительно страдает неполнотой, ибо в нем (эксперименте) не указывается, в какой системе проводится измерение одного из двух сцепленных электронов. Второй электрон реагирует на акт измерения, проводимого над первым электроном, но в приобретенном им состоянии отсутствует указание на характер необратимого процесса, в результате которого он оказался в данном состоянии.

Квантово-информационный подход к изучению явлений микромира позволяет рассматривать электрон в качестве носителя информационного сигнала. А эктропийные свойства электронов помогают понять, как осуществляется мгновенная связь между двумя сцепленными частицами. Похоже, что оба элемента пары должны одновременно проявлять свои энтропийные или эктропийные характеристики. Существо работы квантового компьютера обычно демонстрируют на поведении такой пары. Если каждая из двух ее частиц может находиться в состоянии «0» и «1», то ее можно отождествить с двухкубитовым регистром, в котором реализуются комбинации 00, 01, 10, 11. Регистр называется кубитовым, поскольку его работа протекает в режиме унитарной эволюции, которую претерпевает квантовая суперпозиция.

До тех пор, пока выдерживается когерентность данной суперпозиции, сохраняется и связь сцепления между частицами. А результат работы квантового регистра выдается в виде одного члена суперпозиции. Элементы новизны, предлагаемые в нашем обзоре квантово-компьютерных технологий, заключаются в том, что необратимый процесс, который сопровождает акт выбора одного из членов вышеуказанной суперпозиции (измерение в квантовом компьютере), вовсе не обязан быть энтропийным. Обратимый процесс унитарной эволюции, описываемой уравнением Шредингера, и такой же обратимый процесс обработки информации в квантовом компьютере образуют ту равновесную линию движения, от которой становятся вполне ре-

альными отклонения от заданного уровня энтропии в сторону ее уменьшения. Такие антиэнтропийные отклонения сопровождаются *внутренним* приростом информации в системе. Видимо, так работает и мозг человека. Ведь прирост логико-математической информации, полученной в гёделевой формуле, нельзя рассматривать иначе, как информацию, полученную из внутренних источников. Но она не может быть выработана с помощью тех методов вычислений, что используются в классическом компьютере.

Обратный ход мысли, когда мы начинаем рассматривать процесс унитарных преобразований в квантовой теории как процесс вычислительный, приводит к тем же самым выводам. В самом деле, такой процесс поддаётся экспериментальной проверке только в том случае, если он заканчивается определённым результатом. Но полученный таким образом извне результат воспринимается субъектом и вписывается в его сознание. Самое существенное здесь состоит в том, что внешний результат усваивается сознанием путём встраивания его в ряд результатов внутренних квантово-информационных вычислений, т.е. вычислений, производимых самим мозгом. Поскольку ставятся в один ряд внешнее и внутреннее восприятия, они уподобляются друг другу. Это вполне естественно, но, к сожалению, их часто смешивают между собой.

Так И. [Дж.] фон Нейман сделал следующее заявление с целью объяснения феномена редукции волновой функции: «... неотъемлемо всецело верно, что измерение (квантовомеханическое. – Л.А.) или родственный процесс субъективного восприятия является новой сущностью по отношению к физическому окружению и не сводится к последнему. Действительно, субъективное восприятие заводит нас в интеллектуальную внутреннюю жизнь индивида, которая сверхчувственна (extra-observational) по самой своей природе....»

Тем не менее, в рамках фундаментальной научной точки зрения – так называемого принципа психофизического параллелизма – должно быть возможно описать экстрафизический процесс субъективного восприятия, как если бы это имело место в действительности в физическом мире...» [9; 407].

Как понимать «экстрафизический процесс субъективного восприятия»? Фон Нейман, конечно же, имеет в виду внешнее восприятие. Но он ошибочно отождествляет его с внутренним восприятием, поскольку в обоих случаях для получения информации необходимо располагать унитарно-вычислительным процессом и прерывать его в какой-то момент времени, чтобы фиксировать те результаты, к которым он приводит.

На факт наличия антиэнтропийных необратимых процессов в церебральной системе человека обращал внимание Н.Бор. Касаясь таких категорий сознания, как память, Бор писал: «Прежде всего самое

слово сознание относится к опыту, который может удержаться в памяти; это обстоятельство показывает нам сравнение между сознательным опытом и физическими наблюдениями». И далее: «С биологической точки зрения мы можем толковать признаки психических явлений, только считая, что всякий сознательный опыт соответствует остаточному следу в организме, сводящемуся к остающейся в нервной системе необратимой записи исхода процесса» [10; 108]. Бор констатирует наличие необратимых записей в мозге, подразумевая, конечно, необратимость антиэнтропийную. Иначе его высказывания были бы абсурдными, иначе мы должны были бы признать, что структурирование, упорядочивание, обогащение памяти ведёт к увеличению энтропийного хаоса.

Вместе с тем надо учитывать, что, в отличие от внутренних процессов, наблюдаемая декогеренция во внешнем физическом мире является в большинстве случаев энтропийно-необратимой. Исключения имеют место тогда, когда физик работает с термодинамическими системами, характеризующимися отрицательными, по абсолютной шкале, температурами. Но это тема отдельного разговора.

### **§5. Принцип двойственности в идеологии квантово-компьютерных технологий**

Неизбежное использование мнимой единицы в математическом формализме квантовой теории служит признаком двойственного характера квантовых явлений в более глубоком смысле по отношению к тому, что принято называть *корпускулярно-волновым дуализмом*. Каждому комплексному числу ставится в соответствие сопряжённое с ним число, каждой комплексной функции соответствует комплексно-сопряжённая функция. Этот фундаментальный дуализм в теории квант маскируется тем обстоятельством, что в нерелятивистской квантовой механике на равных правах фигурируют в гильбертовом пространстве как комплексный вектор, так и сопряжённый с ним со-вектор, как исходная волновая функция, так и функция с ней сопряжённая. П.А.М. Дирак в своё время пояснял: «Вследствие взаимно-однозначного соответствия между векторами и со-векторами любое состояние динамической системы в определенный момент времени может быть охарактеризовано как направлением вектора состояния, так и направлением соответствующего со-вектора» [11;39]. Однако в релятивистской квантовой механике всё выглядит иначе: вектором и со-вектором описываются разные состояния движения. При общем решении квантово-релятивистского уравнения Дирака выясняется, что с двумя комплексно-сопряжёнными волновыми функциями соотносятся два состояния движения одной и той же частицы, в конкретном случае, электрона: состояние «досветовое», энтропийное, и состояние «сверхсветовое», эктропийное [12]. Различие определяется системами, в которых электрон наблюдается. Эктропийное со-

стояние электрона характеризуется отрицательными, по абсолютной шкале Кельвина, температурами.

В рамках фундаментального дуализма квантовая информация разделяется на две дополнительные по отношению друг к другу части: информую (собственно) квантовую и информую классическую. Это разделение определяется двумя разными каналами, по которым передаётся та и другая. Чтобы понять, о чём здесь конкретно идёт речь, нам придётся коснуться известной, в физике квантовой информации, теоремы о запрете клонирования.

. Теорема обычно формулируется так: *неопределённое, в классическом смысле, квантовое состояние не может быть клонировано*. Неопределённость (неизвестность) в данном случае означает, что мы располагаем квантовым состоянием, относительно которого не имеется классической информации.. Из доказательства теоремы становится ясно, о каком именно состоянии идет речь. Из-за чисто технических трудностей, возникающих при вёрстке статьи, я не стану здесь вдаваться в детали доказательства теоремы. Придётся ограничиться небольшим разъяснением её сути. Если один лаборант располагает квантовым состоянием (движения) частицы желает передать его своему другу, находящемуся где-то в другой лаборатории, то он должен считаться с неизбежной потерей: он не сможет снять с него копию, чтобы оставить её у себя. Получается так, что с наличным квантовым состоянием можешь манипулировать только ты сам или твой партнёр. Акт передачи нельзя представить в виде некоторого процесса, имеющего начало и конец.

Получение копий по отношению к состояниям движения возможно лишь тогда, когда они суть элементы классического мира. Таков общий вывод, извлекаемый из данной теоремы. Всё это касается процедуры квантовой телепортации. Отметим здесь два существенных момента. Во-первых, как принято в литературе говорить, передача квантового состояния от Алисы к Бобу, происходит так, что Алиса теряет то, что она направляет Бобу, поскольку она не имеет возможности оставить у себя копию состояния. Во-вторых, телепортация осуществляется с помощью двух каналов. Полученная Бобом квантовая информация дополняется классической информацией, которую Алиса диктует Бобу обычным способом по радио или по телефону. Первый канал, как уже указывалось выше, есть канал внепространственный (квантовая информация переносится мгновенно), второй канал пространственно-временной – информация распространяется в пространстве с течением времени.

Хотя теорему о запрете клонирования пытаются как-то увязать с эйнштейновским принципом запрета сверхсветовых скоростей, но в таком увязывании концы с концами не сходятся. Теорема о запрете клонирования и методика передачи квантового состояния от Алисы к Бобу свидетельствуют о том, что преодоление светового барьера как

раз имеет место, но только не так, что существуют скорости движения каких-то физических объектов, превышающие скорость распространения света в вакууме. Теорема связана с фактом существования канала мгновенной передачи квантовой информации от одной системы к другой. Проявляющийся таким образом в квантовой телепортации принцип дополнительности (Боровский принцип) распространяется на отношение, имеющее место между физическим вакуумом и пространственно-временным многообразием.

И последнее замечание. В гёделевых теоремах неполноты используются как объектный язык, так и метаязык формальной системы. Сам субъект выступает при их доказательстве в качестве метанаблюдателя. В квантово-компьютерных вычислениях, осуществляемых церебральной системой, функцию метанаблюдателя выполняет, судя по данным специальных исследований (см., например, [13]), правое полушарие мозга.

#### ЛИТЕРАТУРА

1. Физика квантовой информации. М.: Постмаркет, 2002.
2. Священник Павел Флоренский. История и философия искусств. М.: Мысль, 2000.
3. П. Эрэнфест. Относительность, кванты, статистика. М.: Наука, 1972.
4. В. Паули. Физические очерки. М.: Наука, 1975.
5. E. Schrödinger.. Die gegenwärtige Situation in der Quantenmechanik // *Naturwissenschaften*, 23, 807; 823; 844 (1935).
6. Э. Стин. Квантовые вычисления. Москва–Ижевск, 2000.
7. Н.И. Кобозев. Исследование процессов информации и мышления М.: изд-во МГУ, 1971.
8. Л.Г. Антипенко. Проблема неполноты теории и её гносеологическое значение. М.: Наука, 1986.
9. И. [Дж.] фон Неман. Математические основы квантовой механики. М., 1964.
10. Нильс Бор. Атомная физика и человеческое познание. М.: ИЛ, 1961.
11. П.А.М. Дирак. Принципы квантовой механики. М.: Физматгиз, 1960.
12. Л.Г. Антипенко. К вопросу об общем и частном в решениях квантово-релятивистского уравнения Дирака // 100 лет квантовой теории. Труды международной конференции. М.: НИИ–Природа, 2002.
13. С. Спрингер, Г. Дейч. Левый мозг, правый мозг. М.: Мир, 1983.

---

## ЗЕРКАЛЬНЫЙ МОЗГ, КОНЦЕПТЫ И ЯЗЫК: ЦЕНА АНТРОПОГЕНЕЗА\*

Т. В. Черниговская

Попытки определить и понять в рамках научного знания, в чём кардинальное отличие человека от других биологических видов и какова его природа, имеют не такую уж долгую историю — менее 150 лет: в 1859 году Дарвин издал «Происхождение видов», а в 1871 — «Происхождение человека» [63, 64]. С тех пор наши представления о своей биологической истории, особенно с введением в эту область науки генетических исследований, неизмеримо выросли, и мы можем построить генеалогическое древо до времени формирования современного человека на территории Африки. По независимым оценкам разных групп исследователей, что показал анализ митохондриальной ДНК, временем появления *Homo sapiens* как биологического вида следует считать период около 185 тысяч лет назад. Около 60–70 тысяч лет назад, ещё до выхода из Африки эта популяция разделилась по крайней мере на три подгруппы, давшие начало африканской, монголоидной и европеоидной расам [48, 130]. Сопоставимые результаты дают и исследования Y-хромосомы, хотя и указывают на более поздние сроки — 140–175 тысяч лет назад [136]. Мы знаем также, что младенец, рождённый сейчас, генетически очень мало отличается от рождённого в начале нашей биологической истории; известно, какие линии оказались тупиковыми, а какие привели к возникновению человека современного типа и разных расовых и этнических групп [6, 16, 9]. Следует, однако, обратить внимание на только что (сентябрь 2005г) опубликованные результаты, которые показывают, что микроцефалин (ген, регулирующий объём мозга) продолжает адаптивно эволюционировать; генетическая разновидность *ASPM*, детерминирующая объём мозга у *Homo sapiens*, возникла примерно 5800 лет назад и с тех пор нарастает под давлением положительного отбора. Это свидетельствует о том, что человеческий мозг всё ещё находится под воздействием адаптивных эволюционных процессов [73, 110].

Генетическая близость человека и нашего ближайшего родственника шимпанзе даёт основание считать их подродами одного рода: *Homo Pan* и *Homo Homo* [85]. В Эфиопии, Кении и Чаде обнаружены

---

\* Работа выполнена при поддержке РФФИ (грант № 03-06-80068 и № 03-06-00206) и РГНФ (грант № 04-04-00083а).



миоценовые гоминиды древностью 7–5 млн лет [40], что указывает на отрезок 9–7 млн лет назад как на время дивергенции. Считается, что человек и шимпанзе — два триба одного подсемейства, что ставит вопрос о правомочности отнесения представителей человеческой линии к семейству гоминид, однако этот вопрос далеко не ясен. Генетически человек не противостоит своим ближайшим родственникам — шимпанзе и горилле, но родство его с шимпанзе больше, чем каждого из них — с гориллой, и ясно, что человек — эволюционно поздняя ветвь на родословном древе приматов, которая резко отделилась и стала развиваться по абсолютно особой траектории.

Несомненно, что основные эволюционные приобретения человека следует искать в структуре и функциях головного мозга. Несмотря на растущий объём знаний о психике человека — его языке, семиотических возможностях и способности к формированию концептов и на данные о сопоставлении этих функций с высшими проявлениями психических способностей других биологических видов, мы тем не менее очень плохо представляем себе, что такое Сознание — главная наша характеристика как вида (наряду с языком) — и как оно обеспечивается мозговой активностью. В этой связи стоит вспомнить дискуссию “*The Self and Its Brain*”, происходившую почти 30 лет назад между крупнейшим нейрофизиологом Джоном Экклзом и крупнейшим философом науки Карлом Поппером [122] и признать, что всё нарастающая лавина надёжных данных функционального картирования мозга и некоторый прогресс в теоретических знаниях тем не менее не привели за это время к значимому прорыву в осмыслении проблемы. Вероятно, следует возлагать надежды не на ещё большее усложнение разрешающей способности техники, а на методологический и даже философский прорыв, который должен привести к возникновению новой мульти-дисциплинарной научной парадигмы.

В физиологической науке такой шаг может отталкиваться, например, от недостаточно освоенных и оцененных и, к сожалению, практически неизвестных вне России трудов А. А. Ухтомского, не только выдающегося физиолога, но крупного и высоко нравственного мыслителя, опередившего своё время почти на век. Достаточно отметить помимо общеизвестного принципа доминанты такие важнейшие для продумывания, постановки и интерпретации экспериментов положения как невозможность игнорировать роль наблюдателя (давно признанная квантовой физикой), идея хронотопа и роли адаптивного поведения, введение в биологическое знание таких понятий как этика и нравственность — роль чего ещё не осознана естественными науками, в основном базирующимися на бихевиористских схемах. Идеи Ухтомского о построении интегрального знания о человеке, согласно которым *разобцение функций — абстракция*, вполне могут определить научное и философское пространство XXI века.

Язык является дифференцирующим признаком, характеристикой человека как вида. Это вполне ясно формулировал уже Дарвин, подчёркивая, что дело не в артикуляции как таковой, что доступно, например, некоторым видам птиц при совершенно иной анатомии звукопроизводящих органов, а в способности связать определённые звуки с определёнными идеями (*power of connecting definite sounds with definite ideas*). Однако далее Дарвин говорит о том, что такая способность хоть и характеризует именно человека, но не является автономной, а базируется на развитии ментальных способностей вообще. Это очень важное замечание, т. к. с тех пор и до сего времени основные споры именно и ведутся вокруг двух диаметральных позиций: «особости», отдельности, в том числе и анатомической, языковой способности человека или включения её в число других высших психических функций, считая одним из видов присущих мозгу вычислительных операций.

Это положение вызывает множество вопросов. Орудия труда, изготовленные первобытным человеком *Homo erectus* и требовавшие, как очевидно, развитых мыслительных возможностей, серийной организации деятельности, планирования и способов передачи этих знаний другим членам сообщества и следующим поколениям, датируются в разных местах возрастом 500 000–800 000 лет, что иногда интерпретировалось как указание на наличие языка с его символическими и концептуальными возможностями. Однако, вопрос этот крайне противоречив, и данных для такой хронологии возникновения языка (а стало быть и человека современного типа) явно недостаточно. Общеизвестно, что объём мозга в процессе антропогенеза увеличивался, в основном за счёт неокортекса и фронтальных его отделов. Тем не менее, несмотря на уже сопоставимый с современной популяционной нормой объём, это почему-то не обеспечивало никакого видимого материального прогресса в течении сотен тысяч лет (что видно, например, по орудиям труда). Это вызывает естественные вопросы, на которые нет удовлетворительных ответов: что «мешало» мозгу такого объёма обеспечить необходимые процедуры для усложняющейся деятельности, гарантируя успешную конкуренцию? что позволило мозгу, который уже сотни тысяч лет был достаточно для возникновения сложного поведения и языка объёма, внезапно стать несравнимо более эффективным?

Археологами и антропологами фиксируется «внезапный» взрыв креативных способностей древних людей, произошедший примерно 75 000–50 000 лет назад. Это ассоциируется с ростом интеллекта и сознания; вполне вероятно, что именно в это время формируются высшие психические функции, необходимые не только для языка как такового (в частности, для синтаксиса), но и шире: многоэтапное планирование, цепочки логических операций, изобретение игр на основе конвенцио-

нальных правил, поиск закономерностей в наблюдаемых явлениях, и музыка [88, 74, 103].

В этой связи необходимо остановиться на очень важных работах М. Дональда [70, 71, 72], где обсуждается роль разных видов памяти и обучения в эволюционных процессах, формировавших человека, и одним из важнейших называется мимезис — способность копирования, подражания, имитации. Долгое время (сотни тысяч лет) наши биологические предки могли обходиться без вербального языка, развивая при этом весьма сложные навыки, а значит и мозг. Это время, вероятно, было заполнено и формированием концептов-примитивов, позволяющих создавать некие гипотезы о характере и свойствах внешнего мира. С другой стороны, ясно, что формирование любых, даже и самых первичных концептов требует языка для их дифференциации и номинации. Как и когда возник язык в собственном смысле слова — вопрос открытый [27]. Весьма вероятно, что это произошло много позже и, скорее всего, по одному из двух возможных сценариев: «грамматический взрыв» как результат макромутации или как результат отбора мелких мутаций, т. е. гораздо более постепенного процесса (см., например, [119, 123, 117, 120, 55]).

Чрезвычайно важными для обсуждения этого вопроса являются работы Джеккендоффа [97]. Основная идея их сводится к спору со сторонниками генеративной грамматики, для которых центром языка, его комбинаторных возможностей является синтаксис. Джеккендофф считает, что более обоснована предлагаемая им и вызывающая горячие споры представителей самых разных наук концепция параллельной архитектуры, где фонология, синтаксис, лексикон и семантика являются независимыми генеративными системами, связанными друг с другом интерфейсами [98]. Эта концепция гораздо более совместима как с данными нейронаук и менталистской теорией семантики, так и с более правдоподобными, чем идея мутации, гипотезами эволюции языковой способности человека. Даже в недавних работах главного адепта идеи макромутации Хомского с соавторами [91] и дискуссии вокруг неё [118] ясно показано, что большая часть «вычислительных» и сенсорных способностей разделяется нами с другими млекопитающими, и научение, в том числе и языковое, несомненно включает в себя семантический компонент. По Джеккендоффу, именно *значение* (а не синтаксические структуры) должно было быть первым генеративным компонентом, вызвавшим возникновение и дальнейшее развитие языка. Первая стадия была, скорее всего, выражена символическим использованием простейших вокализаций (или жестов), без какой-либо грамматической организации. На этой стадии, конечно, нет синтаксиса, так как это «однословные» сигналы, но это уже палео-лексикон, отражающий концепты-примитивы. Потом начинает появляться первичный синтаксис, дающий возможность дифференцировать, например,

объект и субъект, маркируя это очерёдность следования компонентов сообщения. И только потом, по мере усложнения выражаемой семантики и конвенциональных правил соотнесения её с фонологией, возникают синтаксические структуры в современном понимании. Такой подход, конечно, в гораздо большей мере, чем предшествующие, открывает путь к интеграции различных областей знаний для построения непротиворечивой теории.

Книга Джеккендоффа вызвала резкую критику сторонников основной генеративистской парадигмы, помещающей синтаксис на привилегированное место и настаивающих на внезапном, а не эволюционном возникновении языка. Так, Биккертон [37] не видит объяснений тому факту, что постепенно развивающийся, по Джеккендоффу, язык почему-то не вызывал никаких изменений в других видах когнитивной эволюции, как будто застывшей на сотни тысяч лет. Он также не видит причин дополнять ещё в 1990 году [35] сформулированные им две стадии возникновения языка: асинтаксический прото-язык и основанный на синтаксисе язык современных людей. Необихевиористы, напротив, сопоставляют ряд положений предлагаемой теории с подходами Скиннера и подчёркивают важность учёта поведения, а не только языковых структур [47].

Арбиб [26] выводит несколько свойств, которые должны были возникнуть, чтобы мозг стал готовым (*language-ready*) для появления языка: способность к имитации комбинаций сложных движений; способность ассоциировать определённый символ с классом объектов, действий и событий; способность «соучаствовать», понимая, что слушающий и говорящий разделяют общее знание о ситуации; интенциональность коммуникации (понимание того, что должен быть результат); понимание иерархической структуры объектов и действий и временной организации; возможность вспоминать и предвидеть; долгий период детства с зависимостью от взрослых и жизнь в социуме, обеспечивающие возможности сложного научения. Нужно, однако, добавить, что этого недостаточно и появление фонологической структуры, организованной цифровым образом для базисного кодирования языка, является крупнейшим когнитивным шагом, выходящим за рамки биологической необходимости нечто выразить [98]. И, конечно, есть огромная разница между закрытыми списками врождённых коммуникационных сигналов других биологических видов и использованием открытого и ничем не лимитированного репертуара знаков, организация бесконечного множества которых только и возможна с помощью фонологического кодирования и далее — правил сложного синтаксиса.

Открытие Ризоллати и Арбибом [29, 127, 129, 128] зеркальных нейронов и вообще т. н. зеркальных систем дают совершенно новые подтверждения принципиальной важности имитации и даже самого

факта фиксации действий другого в нервной системе для когнитивного развития в фило- и онтогенезе и даже для возникновения языка и рефлексии как основ сознания человека [25, 26]. Зеркальные нейроны были открыты в префронтальной моторной коре макака, в частности в зоне F5. Было обнаружено, что эти системы картируют внешнюю информацию — действия (не просто движения), совершаемые другим существом, необязательно того же вида, но с понятной системой координат и интерпретируемым поведением, т. е. с относительно разделяемой семиотической системой (семиосферой) [139]. Зеркальные нейроны реагируют только на *определённое* действие (не любое) и вне модальности стимула: когда субъект делает что-то сам, когда видит это действие или слышит о нём. Риззолатти говорит и о *зеркальных системах*, которые есть практически во всех отделах мозга человека, и активируются, в том числе, при *предвидении* действия, при сопереживании эмоций или воспоминании о них и т. д. Гомологичная исследованной на макаках в связи с открытием зеркальных нейронов зона мозга человека — 44 поле по Бродману, частично являющееся зоной Брока и обеспечивающей речь (см. также [95]). Оказалось, что и у человека эта зона отвечает как за сами хватательные движения, так и за наблюдения за ними [28], что показывает, на основе чего развился мозг, готовый для функционирования языка и построения моделей сознания других людей.

Вполне вероятно, что первые гоминиды уже имели некий прото-язык на основе достаточно примитивной системы знаков и вполне возможно — жестовый [33], что и подготовило мозг к организации семиотической системы, оснащённой сложным синтаксисом с его продуктивностью (принципиальной возможностью создавать бесконечное количество новых сообщений на основе единых алгоритмов). Следует иметь в виду, что общие принципы эволюционных процессов могут быть прослежены применительно к таким различным системам как гомеостатические и системы естественного и искусственных языков языковые [12, 112, 33, 34].

Риззолатти и Арби рассматривают язык (продукцию и восприятие) как способ соединения когнитивной, семантической и фонологической форм, релевантный как для звукового, так и для жестового языка. Активность зеркальных нейронов в зоне F5 интерпретируется как часть кода, которая должна соединиться с нейронной активностью в какой-то другой зоне мозга и завершить тем самым формирование целого кода указанием на объект и/или субъект. Эта гипотеза имеет первостепенное значение как для объяснения организации языковых функций, в частности для лингвистической дифференциации субъекта и объекта, так и для научения вообще, так как позволяет связать в оперативной памяти *агенса* (деятель), *пациенса* (объект действия) и *инструмент* (способ или орудие). Чрезвычайно важным яв-

ляется и формирование с помощью этих систем надёжных механизмов самоидентификации, что нарушается при психической патологии — шизофрении — и также оказывается связанным с функционированием зеркальных систем [28].

Таким образом, впервые показано, как происходит формирование ментальных репрезентаций и механизм, посредством которого это оказывается возможным. По всей видимости, существует некий «словарь» действий как таковых, независимо от того, чем (рукой, ногой, ртом...) и кем они совершаются, сопоставимых с концептами-примитивами, (хватание, доставание, кусание и т. д.), и именно на это реагируют зеркальные системы.

Способность высших (в отличие от низших) обезьян к имитации общеизвестна. Естественно в свете вышесказанного, что такая способность была залогом развития новых моторных и когнитивных возможностей за счёт обучения через мимезис [72], механизмы чего после открытия Ризолатти объяснены нейрофизиологически, и это имеет первостепенное значение для исследований происхождения языка и сознания.

Человеческий язык — не просто одна из высших психических функций, а совершенно особая, видоспецифичная вычислительная способность мозга, не только дающая возможность строить и организовывать чрезвычайно сложные коммуникационные сигналы, но обеспечивать мышление — формирование концептов и гипотез о характере, структуре и законах мира, способность, обеспечивающая функционирование знаковой системы высокого ранга и символическое поведение.

Языковая способность (*language competence*) — система базисных универсальных правил, врождённое (как это считает Хомский [55]), свойство человеческого мозга, представляющее собой основу речевой деятельности человека (*language performance*). Можно говорить о «модулях», составляющих язык: это — лексикон, представляющий собой сложно и по разным принципам организованные списки лексем, словоформ и т.д; вычислительные процедуры, обеспечивающие грамматику (морфологию, синтаксис, семантику и фонологию) и механизмы членения речевого континуума, поступающего из-вне; и прагматическая система. Эти модули взаимодействуют.

Сторонники классического модулярного подхода считают, что правила Универсальной Грамматики, по которым построены все человеческие языки, описывают организацию языковых процедур как: (1) символические универсальные правила, действующие в режиме реального времени и базирующиеся на врожденных механизмах, запускаемых в оперативной памяти, и (2) лексические и другие гештальтно представленные единицы, извлекаемые из долговременной ассоциативной памяти [120, 123, 99, 38, 137].

Сторонники противоположного взгляда считают, что все процессы основываются на работе ассоциативной памяти, и мы имеем дело с постоянной сложной перестройкой всей нейронной сети, также происходящей по правилам, но иным, и гораздо более трудно формализуемым [132, 121]. Возможны и не совпадающие ни с одним из этих подходов гипотезы [87, 17, 18, 19, 52]. Языковые процессы пытаются картировать и, по возможности, локализовать в нейролингвистических исследованиях, проверяющих непротиворечивость подтверждения гипотез (см., например, [5, 10, 11, 42, 14, 69]).

Механизмы, обеспечивающие язык и другие высшие функции, рассматриваются на протяжении всей истории изучения то в рамках локализационистской, то — холистической моделей. В настоящее время, несмотря на огромный накопленный за эти годы надежный фактический материал, ситуация мало прояснилась и вышеупомянутые парадигмы продолжают сосуществовать или чередоваться. И всё же, благодаря клиническим данным и функциональному картированию мозга, осуществляемому с помощью различных методов, можно с достаточной степенью уверенности говорить о зонах мозга, обеспечивающих различные аспекты языковой деятельности человека. Например, показано, что существительные и глаголы обрабатываются разными отделами мозга, и вообще разные грамматические категории имеют разные нейрональные представительства [133]. Нужно однако заметить, что эти данные очень неоднозначны и требуют обсуждения в специальной работе: на обработку синтаксиса и морфологии влияет много факторов — от модальности предъявления и типа задания до роли семантики и более широкого контекста: например, Фредеричи с соавторами [80], показали, что в синтаксические процедуры вовлекаются билатеральные механизмы передне-височные отделы коры и зона Вернике. Изучение восприятия эмоциональной просодики при помощи ПЭТ и фМРТ выявило вовлечение в этот процесс правой префронтальной и правой нижней фронтальной коры [96, 41], распределение функций между полушариями в зависимости от типа просодики [21, 94]

П.К. Анохин (опубликовано в 1978г) [1] и Д. Хебб [93] предложили модели, примиряющие локализационистский и холистический взгляды на мозговое обеспечение высших когнитивных функций, хорошо подходящие для описания распределённости по мозгу языковых процедур: клеточные ансамбли вполне определенной топографии могут организовываться в нейробиологические объединения для формирования когнитивных единиц типа слов или гештальтов иного рода, например зрительных образов. Такой взгляд кардинально отличается от локализационистского подхода, так как подразумевает, что нейроны из разных областей коры могут быть одновременно объединены в единый функциональный блок. Он отличается и от холистического подхода, так как отрицает распределение всех функций по

всему мозгу, но подчеркивает принципиальную динамичность механизма, постоянную переорганизацию всего паттерна в зависимости от когнитивной задачи. Это значит, что мы имеем дело с тонко настраиваемым оркестром, местоположение дирижера которого неизвестно и нестабильно, а возможно и не заполнено вообще, так как оркестр самоорганизуется с учетом множества факторов [125, 126] и настраивается на доминанту [15].

Об этом косвенно говорят и данные о распределении энграмм в памяти: один и тот же когнитивный объект оказывается компонентом сразу нескольких ассоциативных множеств — и по оси сенсорных модальностей, и по осям разного рода парадигматических и синтагматических связей. Речь идёт также о волне возбуждения, циркулирующей и ревербирующей по разным петлям нейронного ансамбля, которая в нейрофизиологических терминах может быть описана как пространственно-временной паттерн активности, охватывающий многие нейроны, и не только неокортекса. Необходимо заметить, что и сами функционально возникающие и когнитивно обусловленные ансамбли имеют иерархическую организацию, т. е. могут быть подмножествами других. Допущение такой организации необходимо, в частности для объяснения структуры соответствующих семантических репрезентаций, в частности языковых.

Человек обладает такой важной чертой как способность к аналогии, поиску сходства, а значит к объединению индивидуальных черт и феноменов в классы, что даёт возможность построения гипотез об устройстве мира. На этом пути чрезвычайную роль играют т. н. концепты-примитивы, которые по мнению целого ряда крупных представителей когнитивной науки, являются врождёнными и проявляющимися у детей очень рано, а не приобретёнными в результате раннего научения. Роль языка — не только в назывании, «констатации» объектов или явлений, но и в исполнении неких интенций, влиянии, в том что принято называть иллюкаторной силой и что выражается перформативами. Перформативы должны как минимум в глубинной синтаксической структуре иметь субъект первого лица и прямое или косвенное дополнение (объект действия), они должны быть утвердительными, иметь основной глагол в форме настоящего времени и включать в себя глаголы утверждения, просьбы, говорения, приказа, объявления и т. д.

Базисные врождённые концепты-примитивы сводятся, насколько сейчас известно, к списку примерно из 30 единиц: связанные с пространством и движением в нём — начало «пути», конец «пути»; внутрь «контейнера», из «контейнера»; на поверхность, с поверхности; вверх, вниз; соединение; контакт; ритмическое/прерывистое движение, прямое движение; живые объекты, начинающие двигаться без внешних воздействий (связей и контактов) и ритмично; неоду-



шивлённые объекты, для движения которых нужны внешние воздействия и т. д. Считается, что концепты организованы иерархически и, следовательно, представляют собой систему. Эта система генетически заложена в мозгу человека, где есть также механизм генератора новых концептов, обеспечивающий возможность формулирования гипотез. [79].

Говоря об антропогенезе и развитии высших когнитивных функций и языка, нельзя обойти дискуссию, уже не первый год публикующуюся в био-психологической, нейролингвистической и медицинской литературе — поиски так называемого «языкового гена», или «гена грамматики» [113, 83, 86].

Исследования двух семей с языковыми нарушениями, отмеченными в разных поколениях, дали основания считать, что аномалии локализуются в определённом участке 7 хромосомы, содержащем около 70 генов. Зона поиска постепенно сужалась и в итоге привела к идентификации гена *FOXP2*, входящего в группу *FOX*, контролирующей активность других генов, продуцирующих белок, связывающий ДНК [77, 83, 116, 23].

Несмотря на чрезвычайный резонанс, вызванный публикацией этих данных, нужно подчеркнуть, что делать вывод об открытии «языкового гена» явно преждевременно: нарушения языковых процедур, ассоциирующиеся с этой аномалией, ещё не есть доказательство кодирования такой сложнейшей функции именно этим геном (как и вообще одним геном), тем более, что он был идентифицирован и у других биологических видов. Нельзя не вспомнить, что гены *FOX* известны как ключевые регуляторы эмбриогенеза, и их мутации приводят к видоспецифическим заболеваниям человека; аномалия *FOXP2* влечёт за собой нарушение развития в критический период, не позволяя адекватно сформироваться структурам, которые должны будут впоследствии обеспечивать усвоение языка. Нужно подчеркнуть, что идентификация *FOXP2* — не более, чем первый шаг к декодированию системы максимальной сложности.

Т. Дикон [67] высказывает точку зрения, согласно которой — язык «окупирует» мозг и адаптировался к нему в гораздо большей мере, нежели мозг эволюционировал в сторону языка. Мозг и язык ко-эволюционируют, но главную адаптационную работу, по Дикону, делает язык. Дети, таким образом, уже рождаются с мозгом, готовым к синтаксическим процедурам именно из-за развития языка в сторону наиболее вероятностных характеристик, что и фиксируется генетически. Мозг необходим для мышления, но недостаточен, нужен опыт. Нерофизиологический субстрат, необходимый для интеллектуальной деятельности, развивается: роль коры у новорожденных детей крайне мала. Общеизвестно, что общая масса мозга менее важна, чем его внутренняя организация и богатство связей. Долгое созревание мозга

и позднее начало нейрогенеза обеспечивает больший объём и сложность структуры [68, 34]; чем позднее рождаются нейроны, тем дальше они мигрируют, тем более высокое иерархически положение они занимают и тем большей сложности формируются связи [108], что является основой для высших функций, тем более языка.

Потенциальная возможность говорить зависит от генетических факторов, а реальная речевая продукция — от опыта. Важнейшими характеристиками человеческого языка являются его продуктивность (возможность создания и понимания абсолютно новых сообщений) и иерархическая, и даже «цифровая», структура, т. е. наличие уровней — фонологического, морфологического, синтаксического и уровня дискурса. Такая структурная специфичность общепризнана как уникальная особенность данной системы. Поэтому поиски как правил, описывающих собственно лингвистические феномены, так и генетических основ языковой компетенции базируются прежде всего на анализе этих характеристик.

Согласно одному из взглядов эволюция сделала рывок, приведший к обретению мозгом способности к вычислению, использованию рекурсивных правил и ментальных репрезентаций, создав тем самым основу для мышления и языка в человеческом смысле. Новая «грамматическая машина», как это называет Джэкендофф [97], позволила усложнять и наращивать языковые структуры для организации (мышление) и передачи (коммуникация) все усложняющихся концептов.

Было неоднократно показано, что ни у горилл [24, 44], ни у шимпанзе [76, 107], не обнаружена доминантность конечностей, что, несмотря на неоднозначность трактовок, всё же является одним из главных показателей мозговой латерализации. Известно также, что абсолютное большинство людей — праворуки и более того, имеют обеспечиваемый преимущественно левым полушарием мозга латеральный профиль, и только 6–12% населения в зависимости от критерия оценки и даже от разброса по разным регионам — левши; показаны также гендерные различия — мужчин-левшей больше [109, 114, 7, 13, 134, 57, 20]

Вопрос о роли церебральной асимметрии в развитии человека ставился многократно и в разных аспектах — влияние генетических факторов и среды (например, типа обучения или культуры), половой диморфизм, разная скорость созревания гемисферных структур, разная скорость протекания нервных процессов (что могло, например, повлиять на особую роль левого полушария в анализе требующих большой скорости обработки фонематических процедур со всеми вытекающими из этого для языковой доминантности последствиями) (См. обзоры: [3, 4, 2, 39, 49, 50, 51, 53]).

В 1997 году из печати выходит статья Т. Кроу [59], где впервые ставится вопрос о том, что язык и психоз имеют общие эволюцион-

ные истоки — генетические изменения или даже генетическое «событие», отделившее *Homo sapiens* от предшествующих биологических видов. Это связывается им с изменениями половых хромосом, произошедшими на территории Восточной Африки в период от 100 до 250 тысяч лет назад и вызвавшими асимметричный сдвиг в развитии полушарий, особенно в ассоциативной коре, так что левое полушарие стало регулировать наиболее сложные и одновременно ключевые компоненты языка — анализ и синтез фонологических цепочек, морфологию и синтаксис, в то время как к правому полушарию отошла функция регулирования процессов смыслообразования и прагматические аспекты речи.

Идея Кроу сводится к тому, что критические генетические изменения, позволившие полушариям развиваться в известной степени независимо, являются одновременно и теми, которые вызывают психоз, т. е. имеют общие корни [59, 60, 61]. Действительно, нарушение полушарного баланса при шизофрении является установленным фактом [8, 101, 23, 102, 42, 21], однако есть ли основания для такой жёсткой постановки вопроса о едином источнике языка и психоза? Последние психолингвистические исследования языка шизофрении в эволюционном контексте дискуссий вокруг общей генетической истории свидетельствуют о том, что мнения по этому вопросу неоднозначны [131, 66, 115, 65, 61, 58, 56].

Наиболее резким оппонентом Кроу является К. Бёрнс с эволюционной теорией шизофрении<sup>1</sup> и связей её с языком и происхождением человека. Шизорения является заболеванием многофакторной, но главным образом генетической этиологии, и есть мнение, что не затронутые этой болезнью родственники имеют эволюционные преимущества за счёт повышенной креативности, что, по его мнению, человечество вынуждено компенсировать наличием психозов в популяции. Бёрнс предлагает альтернативную Т. Кроу теорию, утверждая, что шизофрения является платой за эволюцию метакогнитивных способностей и социального мозга, а не языка.

Палеоантропологические и приматологические данные свидетельствуют, что у гоминид развились сложные кортикальные связи, особенно в лобно-височных областях, обеспечившие регуляцию социального поведения и интеллектуальные потребности, обусловленные социумом, что привело к уязвимости мозга для генетических или иных нарушений: такова плата за сложную организацию нейронной сети. Согласно этой теории, «генетические события», произошедшие с *Homo sapiens* до выхода из Африки дали толчок к появлению спектра развития психики: гомозиготная форма давала шизофренический фенотип, а гетерозиготная — шизотипический тип личности с нетри-

---

<sup>1</sup> <http://www.bbsonline.org/Preprints/Burns/Referees/Burns.pdf>

виальными когнитивными способностями высокого уровня, что делало таких индивидуумов относительно адаптированными и, более того, вносящими серьёзный вклад в культурную и научную историю человечества.

Однако, возможен сценарий, по которому гены, ответственные за шизофрению, существовали и до появления *Homo sapiens* и были вполне нейтральны или даже адаптивны; с развитием новых мозговых структур и связей, обеспечивающих новые сложные функции, эти гены могли измениться за счёт мутации или через иную экспрессию или за счёт интеграции, и с тех пор отвечать за развитие шизофренических нарушений.

Остаётся неясным, к чему приводят генетические изменения, вызывающие шизофрению: к нарушению нейрогенеза, миграции нейронов, арборизации, синаптогенеза или апоптоза. Скорее всего, все эти процессы нарушены, т. к. шизофрения — аномалия развития нервной системы. Возможно, это вызывает не столько патологию языка, сколько патологию *социального мозга*, функции которого включают в себя не только адаптивное социальное поведение, но и такие более конкретные свойства как понимание выражения лица, восприятие и продукцию эмоций, в том числе контроль над собственным настроением и способность оценивать и учитывать представления и точки зрения других людей, **метарепрезентацию**, *theory of mind* — термин, предложенный для описания возможностей шимпанзе [124], и ставший в последние годы чрезвычайно широко применяемым для исследования этой способности высших приматов, у маленьких детей и психически больных, в том числе в кросс-культурном аспекте [135].

Функциональное картирование показало, что при выполнении заданий, связанных с этой способностью, активированными оказываются левая медиальная префронтальная кора, орбито-фронтальная кора и левая височная кора [31, 84, 78, 100, 138]. Медиальная префронтальная кора (и особенно передняя часть поясной извилины) указываются наиболее часто [82, 45, 46, 40, 90]; из других структур называют также поясную борозду, заднюю часть поясной извилины, височно-теменную область и префронтальную кору [78, 82, 106].

В качестве нейрональной основы аутизма, при котором также грубо нарушены социальные навыки, были описаны амигдала или миндалевидное тело, орбито-фронтальная кора и верхняя височная борозда, которые обеспечивают адекватное поведение и способность к метарепрезентации и оказываются нарушенными при связанной с этим патологии [32, 30]. Искажение мышления при аутизме, возможно, вызывается дисфункцией медиальных префронтально-париетальных нейрональных систем, выражающейся в невозможности эффективно

модулировать нейронные связи в экстрастриарной зрительной коре и височных долях [81].

Доказано также, что амигдала и орбито-фронтальный кортекс играют жизненно важную адаптивную роль в обеспечении поведения за счёт правильной интерпретации эмоций и социальной ситуации и вовлечения эмоциональных компонентов в процесс научения [75, 22, 92]. Данные, полученные на грызунах, свидетельствуют о том, что окситоцин и вазопресин — два главных пептида, задействованных в таких процессах, осуществляют свои функции через амигдалу и некоторые части стриатума, стало быть, это — очень древняя функция [140]. У людей повреждение орбито-фронтальный кортекса приводит к нарушениям, встречающимся и при шизофрении — амбивалентности, импульсивности, отсутствию интереса к действиям других людей и возможности учёта этого в выстраивании собственных поступков, стереотипному и неадекватному поведению; всё это может протекать на фоне интактных интеллектуальных возможностей другого рода, в том числе и высокого уровня [62, 75]. Неудивительно, что исследования функций мозга у больных шизофренией методом функционального картирования показало значительное снижение активности амигдалы слева и гиппокампа билатерально [89].

Все эти факты дают основания согласиться с тем, что нарушения, приведшие к шизофрении, связаны с гораздо более древними структурами и механизмами, чем те, которые ассоциируются с возникновением языка: скорее, это — слом обеспечения социальных функций. Это не снимает очень сложного вопроса о соотношении языковой способности с одной стороны и развития сознания — с другой, механизмов формирования концептов и гипотез, рефлексии и адекватном поведении в сложно организованном и быстро меняющемся социуме. Всё это — главные характеристики нашего биологического вида, являющиеся высшим достижением эволюции, независимо от цены, которую мы за это заплатили.

## СПИСОК ЛИТЕРАТУРЫ

1. *Анохин П. К.* Избранные труды: Философские аспекты теории функциональной системы. М. Наука. 1978.
2. *Ариавский В. В.* Различия, которые нас объединяют (Этюды о популяционных механизмах межполушарной асимметрии). Рига. 2001.
3. *Балонов Л. Я., Деглин В. Л.* Слух и речь доминантного и недоминантного полушарий. Л. Наука. 1976.
4. *Балонов Л. Я., Деглин В. Л., Черниговская Т. В.* Функциональная асимметрия мозга в организации речевой деятельности. Сенсорные системы. Сенсорные процессы в асимметрии полушарий. Отв. ред Гершуни Г. В. Л. Наука. 99—114. 1985.
5. *Бехтерева Н. П.* О мозге человека. СПб. Нотабене. 1999.
6. *Бунак В. В.* Род Ното, его возникновение и последующая эволюция. М. Наука. 1980.
7. Доброхотова Т. А., Брагина Н. Н. Левши. М. Книга. 1994.
8. *Кауфман Д. А.* Экспериментальное изучение особенностей функциональной специализации полушарий мозга при шизофрении. Физиология человека. 5 (6) : 1007—1019. 1979.
9. *Козинцев А. Г.* Происхождение языка: новые факты и теории. Теоретические проблемы языкознания. К 140-летию кафедры общего языкознания Санкт-Петербургского государственного университета. СПб. СПбГУ. 35—50. 2004.
10. *Медведев С. В., Бехтерева Н. П., Воробьев В. А., Пахомов С. В., Рудас М. С.* Исследование обработки мозгом человека различных характеристик зрительно предъявляемых слов методом позитронно-эмиссионной томографии. Сообщение 2: Мозговая система обеспечения чтения слов. Физиология человека. 22 (3) : 5—10. 1996.
11. *Медведев С. В., Бехтерева Н. П., Воробьев В. А., Пахомов С. В., Рудас М. С.* Исследование обработки мозгом человека различных характеристик зрительно предъявляемых слов методом позитронно-эмиссионной томографии. Сообщение 3: Мозговая система обработки грамматического рода слов. Физиология человека. 22 (4) : 5—11. 1996.
12. *Наточин Ю. В., Мениуткин В. В., Черниговская Т. В.* Общие черты эволюции в гомеостатических и информационных системах. Журн. эволюционной биохимии и физиологии. 28 (5) : 623—637. 1992.
13. *Нейропсихология и психофизиология индивидуальных различий.* Ред Хомская Е. Д., Москвина В. А. Москва, Оренбург. 2000.
14. *Стрельников К. Н., Воробьев В. А., Рудас М. С., Черниговская Т. В., Медведев С. В.* ПЭТ-исследование мозгового обеспечения восприятия фраз с синтагматическим членением. Физиология человека. 30 (2) : 30—37. 2004.
15. *Ухтомский А. А.* Доминанта. СПб., 2002.
16. *Хуснутдинова Э. К.* Этногеномика и генетическая история народов Восточной Европы. Вестник Российской Академии наук. 73 (7) : 614—621. 2003.
17. *Черниговская Т. В.* Ното Loquens: эволюция церебральных функций и языка. Журн. эволюционной биохимии и физиологии. 40 (5) : 400—406. 2004.
18. *Черниговская Т. В.* Мозг и язык: полтора века исследований. В: Теоретические проблемы языкознания. К 140-летию кафедры общего языкознания

- Санкт-Петербургского государственного университета. СПб. СПбГУ. 16—35. 2004.
19. Черниговская Т. В. Язык, мышление, мозг: основные проблемы нейролингвистики. Труды отделения историко-филологических наук РАН. М. 2 : 40—50. 2004.
  20. Черниговская Т. В., Гаврилова Т. А., Воинов А. В., Стрельников К. Н. Сенсомоторный латеральный профиль: тестирование и интерпретация. Физиология человека. 31 (2) : 35—44. 2005.
  21. Черниговская Т. В., Давтян С. Э., Петрова Н. Н., Стрельников К. Н. Специфика полушарной асимметрии восприятия интонаций в норме и при шизофрении. Физиология человека. 30 (4) : 32—39. 2004.
  22. Amaral D. G. The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. Biological Psychiatry. 51 : 11—17. 2002.
  23. Andrew S. Communicating a New Gene Vital for Speech and Language. Clinical Genetics. 61 : 97—100. 2002.
  24. Annett M., Annett J. Handedness for eating in gorillas. Cortex. 27 : 269—275. 1991.
  25. Arbib M. A. Co-evolution of human consciousness and language. Cajal and consciousness: Scientific approaches to the centennial of Ramon y Cajal's Textura. Ed. Marijuan P. C. Annals of the New York Academy of Sciences. 929 : 195—220. 2001.
  26. Arbib M. A. The mirror system, imitation, and the evolution of language. Imitation in animals and artifacts. Eds Nahaniv C., Dautenhahn K. Cambridge (MA). The MIT press. 229—280. 2002.
  27. Arbib M. A. Protosign and protospeech: An expanding spiral. Behavioral and Brain Sciences. 26 (2) : 209—210. 2003.
  28. Arbib M. A., Mundhenk T. N. Schizophrenia and the mirror system: an essay. Neuropsychologia. 43 : 268—280. 2005.
  29. Arbib M. A., Rizzolatti G. Neural expectations: A possible evolutionary path from manual skills to language. Communication and Cognition. 29 : 393—424. 1997.
  30. Baron-Cohen S. Mindblindness: An Essay on Autism and Theory of Mind. Cambridge (MA). The MIT Press. 1995.
  31. Baron-Cohen S., Ring H., Moriarty J., Schmitz B., Costa D., Ell P. Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults. British J. of Psychiatry. 165 : 640—649. 1994.
  32. Baron-Cohen S., Ring H.A., Bullmore E.T., Wheelwright S., Ashwin C., Williams S. C. The amygdala theory of autism. Neuroscience and Biobehavioral Rev. 24 : 355—364. 2000.
  33. Becoming Loquens. (ser. Bochum Publications in Evolutionary Cultural Semiotics). Eds Bichakjian B., Chernigovskaya T., Kendon A., Moeller A. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien. Peter Lang. 1. 2000.
  34. Bichakjian B. H. Language in a Darwinian Perspective. Frankfurt am Mein, 2002.
  35. Bickerton D. Language and Species. Chicago. University of Chicago Press. 1990.
  36. Bickerton, D. Foraging Versus Social Intelligence in the Evolution of Protolanguage. The Transition to Language. Ed Alison Wray. Oxford. Oxford University Press. 2002.

37. *Bickerton, D.* Symbol and structure: a comprehensive framework for language evolution. *Language Evolution: The States of the Art.* Eds Christiansen M. H., Kirby S. Oxford. Oxford University Press. 2003.
38. *Bloom P.* *How Children Learn the Meanings of Words.* Cambridge, 2002.
39. *Brain Asymmetry.* Eds Davidson R., Hugdahl K. Cambridge (MA). The MIT Press. 1995.
40. *Brunet M., Guy F., Pilbeam D. et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature.* 418 : 145—151. 2002.
41. *Buchanan T. W., Lutz K., Mirzazade S., Specht K., Shah N. J., Zilles K., Jancke L.* Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Res.* 9 : 227—238. 2000.
42. *Burns J. K., Job D. E., Bastin M. E., Whalley H. C., McGillivray T., Johnstone E. C., Lawrie S. M.* Structural dysconnectivity in schizophrenia: a diffusion tensor MRI study. *British J. of Psychiatry.* 182 : 439—443. 2003.
43. *Byrne R.W.* Social and technical forms of primate intelligence. *Tree of Origin: What Primate Behavior Can Tell us about Human Social Evolution.* Ed De Waal F.B.M. Harvard University Press. 145—172. 2001.
44. *Byrne R. W., Byrne J. M.* Hand preference in the skilled gathering tasks of mountain gorillas (*Gorilla g.berengei*). *Cortex.* 27 : 521—546. 1991.
45. *Calder A. J., Lawrence A. D., Keane J., Scott S. K., Owen A. M., Christoffels I., Young A.W.* Reading the mind from eye gaze. *Neuropsychologia.* 40 : 1129—1138. 2002.
46. *Castelli F., Happé F., Frith U., Frith C.* Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage.* 12 : 314—325. 2000.
47. *Catania Ch. A.* *Learning.* Prentice Hall, Upper Saddle River, New Jersey. 2003.
48. *Cavalli-Sforza L. L.* *Genes, Peoples, and Languages.* New York. North Point Press. 2000.
49. *Chernigovskaya T.* Cerebral Lateralization for Cognitive and Linguistic Abilities: Neuropsychological and Cultural Aspects. *Studies in Language Origins.* Eds Wind J., Jonker A. Amsterdam, Philadelphia. John Benjamins Publishing Company. III : 56—76. 1994.
50. *Chernigovskaya T. V.* Cerebral Asymmetry — a Neuropsychological Parallel to Semiogenesis. *Language in the Wurm Glaciation: Acta Colloquii.* Ed Figge U. (Ser. Bochum Publications in Evolutionary Cultural Semiotics. Ed Koch W.). 27 : 53—75. 1996.
51. *Chernigovskaya T.* Neurosemiotic Approach to Cognitive Functions. *Semiotica: J. of the International Association for Semiotic Studies.* 127 (1/4) : 227—237. 1999.
52. *Chernigovskaya T.* Any Words in the Brain's Language? Does Mind Really Work That Way? XXVII Annual Conference of the Cognitive Society. Stresa. Italy. 430—434. 2005.6
53. *Chernigovskaya T., Gor K.* The Complexity of Paradigm and Input Frequencies in Native and Second Language Verbal Processing: Evidence from Russian. *Language and Language Behavior.* Eds Wande E., Chernigovskaya T. 20—37. 2000.
54. *Chernigovskaya T., Natochin Yu., Menshutkin V.* Principles of Evolution of Natural and Computer Languages and Physiological Systems. *Becoming Loquens* (ser. Bochum Publications in Evolutionary Cultural Semiotics). Eds Bichakjian



- B., Chernigovskaya T., Kendon A., Moeller A. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien. Peter Lang. 1 : 211—236. 2000.
55. *Chomsky N.* New Horizons in the Study of Language and Mind. Cambridge. Cambridge University Press. 2002.
56. *Condray R.* Language disorder in schizophrenis as a developmental learning disorder. *Schizophrenia Res.* 72 : 5—20. 2005.
57. *Corballis, M. C.* From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences.* 26 (2) : 199—208. 2003.
58. *Covington M. A., Congzhou H., Brown C., Naci L., McClain J. T., Fjordbak B. S., Semple J., Brown J.* Schizophrenia and the structure of language: The lingusit's view. *Schizophrenia Res.* 77 : 85—98. 2005.
59. *Crow T. J.* Is schizophrenia the price that Homo Sapiens pays for language? *Schizophrenia Res.* 28 : 127—141. 1997.
60. *Crow T. J.* Schizophrenia as the Price that Homo Sapiens Pays for Language: a Resolution of the Central Paradox in the Origin of the Species. *Brain Res. Rev.* 31 : 118—129. 2000.
61. *Crow T. J.* Auditory hallucinations as primary disorders of syntax: An evolutionary theory of the origins of language. *Cognitive neuropsychology.* 9 (1/2) : 125—145. 2004.
62. *Damasio A. R.* Descartes' Error: Emotion, Reason and the Human Brain. New York. Grosset/Putnam. 1994.
63. *Darwin C.* On the Origin of Species by Means of Natural Selection. London. John Murray. 1859.
64. *Darwin C.* The Descent of Man and Selection in Relation to Sex. London. John Murray. 1871.
65. *Davtian S., Chernigovskaya T.* Psychiatry in free fall: In pursuit of a semiotic foothold. *Sign Systems Studies.* 31 (2) : 533—545. 2003.
66. *de Lisi L.E.* Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. *Schizophrenia Bulletin.* 27 : 481—496. 2001.
67. *Deacon T. W.* The Symbolic Species: The Co-Evolution of Language and the Brain. New York. Norton. 1997.
68. *Deacon T. W.* Heterochrony in brain evolution: cellular versus morphological analyses. *Biology, Brains and Behavior.* Eds Taylor Parker S., Langer J., McKinney M. L. Santa Fe. SAR Press. 41—88. 2000.
69. *Démonet J. F., Thierry G., Cardebat D.* Renewal of the neurophysiology of language: functional neuroimaging. *Physiological Rev.* 85 : 49—95. 2005.
70. *Donald M.* Origins of the modern mind. Three stages in the evolution of culture and cognition. Cambridge (MA), London. Harvard University Press. 1991.
71. *Donald M.* Precs of Origins of the modern mind: Three stages in the evolution of culture and cognition. *Behavioral and Brain Sciences.* 16 (4) : 737—791. 1997.
72. *Donald M.* Mimesis and the executive suite: Missing links in language evolution. *Approaches to the Evolution of Language: Social and Cognitive Bases.* Eds Hurford J. R., Studdert-Kennedy M., Knight C. Cambridge. Cambridge University Press. 44—67. 1998.
73. *Evans P. D., Gilbert S. L., Mekel-Bobrov N., Vallender E. J., Anderson J. R., Vaez-Azizi L. M., Tishkoff S. A., Hudson R. R., Lahn B. T.* Microcephalin, a gene regulat-

- ing brain size, continues to evolve adaptively in humans. *Science*. 309 (5741) : 1717—1720. 2005.
74. *Falk D.* Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*. 27 (4) : 491—503. 2004.
  75. *Farrow T. F., Zheng Y., Wilkinson I. D., Spence S. A., Deakin J. F., Tarrier N., Griffiths P. D., Woodruff P. W.* Investigating the functional anatomy of empathy and forgiveness. *Neuroreport*. 12 : 2433—2438. 2001.
  76. *Finch G.* Chimpanzee handedness. *Science*. 94 : 117—118. 1941.
  77. *Fisher S. E., Vargha-Khadem F., Watkins K. E., Monaco A. P., Pembey M. E.* Localisation of a Gene Implicated in a Severe Speech and Language Disorder. *Nature Genetics*. 18 : 168—170. 1998.
  78. *Fletcher P. C., Happé F., Frith U., Baker S. C., Dolan R. J., Frackowiak R. S., Frith C. D.* Other minds in the brain: a functional imaging study of «theory of mind» in story comprehension. *Cognition*. 57 : 109—128. 1995.
  79. *Fodor J.* *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge. The MIT Press. 2001.
  80. *Friederici A. D., Meyer M., von Cramon D. Y.* Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language*. 74 (2) : 289—300. 2000.
  81. *Frith C. D.* Attention to action and awareness of others minds. *Consciousness and Cognition*. 11 : 481—487. 2002.
  82. *Gallagher H. L., Happé F., Brunswick N., Fletcher P. C., Frith U., Frith C. D.* Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*. 38 : 11—21. 2000.
  83. *Ganger J., Stromswold K.* Innateness, Evolution, and Genetics of Language. *Human Biology*. 70 : 199—213. 1998.
  84. *Goel V., Grafman J., Sadato N., Hallett M.* Modeling other minds. *Neuroreport*. 6 : 1741—1746. 1995.
  85. *Goodman M., Czelusniak J., Page S., Meiereles C.* Where DNA sequences place Homo sapiens in a phylogenetic classification of primates. *Humanity from African Naissance to Coming Millennium*. Eds Tobias P. V., Rath M. A., Moggi-Cecchi J., Doyle G. A. Firenze. 2001.
  86. *Gopnik M.* Some Evidence for Impaired Grammars. *Language, Logic, and Concepts*. Eds Jackendoff R., Bloom P., Wynn K. Cambridge. The MIT Press. 263—283. 1999.
  87. *Gor K., Chernigovskaya T.* Rules in the Processing of Russian Verbal Morphology. *Current Issues in Formal Slavic Linguistics*. Eds Zybatow G., Junghanns U., Mehlhorn G., Szucsich L. Frankfurt am Main etc. Lang. 528—536. 2001.
  88. *Gould S. J.* *The Panda's Thumb*. Penguin. 1980.
  89. *Gur R. E., McGrath C., Chan R. M., Schroeder L., Turner T., Turetsky B. I., Kohler, C., Alsop D., Maldjian J., Ragland J. D., Gur R. C.* An fMRI study of facial emotion processing in patients with schizophrenia. *Am. J. of Psychiatry*. 159 : 1992—1999. 2002.
  90. *Gusnard D. A., Akbudak E., Shulman G. L., Raichle M. E.* Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the U.S.A.* 98 : 4259—4264. 2001.

91. *Hauser M. D., Chomsky N., Fitch W. T.* The Faculty of Language: What is it, Who has it, and How Did It Evolve? *Science*. 298 : 1569—1579. 2002.
92. *Haxby J. V., Hoffman E. A., Gobbini M. I.* Human neural systems for face recognition and social communication. *Biological Psychiatry*. 51 : 59—67. 2002.
93. *Hebb D. O.* The Organization of Behavior. A Neurophysiological Theory. New York. Wiley. 1949.
94. *Hesling I., Clément S., Bordessoules M., Allard M.* Cerebral mechanisms of prosodic integration: evidence from connected speech. *NeuroImage*. 24 : 937—947. 2005.
95. *Hopkins W. D., Cantalupo C.* Brodmann's area 44, gestural communication, and the emergence of right handedness in chimpanzees. *Behavioral and Brain Sciences*. 26 (2) : 224—225. 2003.
96. *Imaizumi S., Mori K., Kiritani S., Kawashima R., Sugiura M., Fukuda H., Itoh K., Kato T., Nakamura A., Hatano K., Kojima S., Nakamura K.* Vocal identification of speaker and emotion activates different brain regions. *NeuroReport*. 8 : 2809—2812. 1997.
97. *Jackendoff R.* Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford. Oxford University Press. 2002.
98. *Jackendoff R.* Précis of Foundations of Language: Brain, Meaning, Grammar, Evolution. *Behavioral and Brain Sciences*. 26 : 651—707. 2003.
99. *Jaeger J. J., Lockwood A. H., Kemmerer D. L., Van Valin Jr. R. D., Murphy B. W., Khalak H. G.* A positron emission tomographic study of regular and irregular verb morphology in English. *Language*. 72 : 451—497. 1996.
100. *Levine B., Freedman M., Dawson D., Black S., Stuss D. T.* Ventral frontal contribution to self-regulation: convergence of episodic memory and inhibition. *Neurocase*. 5 : 263—275. 1999.
101. *Loberg E.-M., Hugdahl K., Green M. F.* Hemispheric Asymmetry in Schizophrenia: A “Dual Deficits” Model. *Biological Psychiatry*. 45 : 76—81. 1999.
102. *Loberg E.-M., Jørgensen H. A., Hugdahl K.* Functional Brain Asymmetry and Attentional Modulation in Young and Stabilised Schizophrenic Patients: A Dichotic Listening Study. *Psychiatry Res*. 109 : 281—287. 2002.
103. *Longhi E., Karmiloff-Smith A.* In the beginning was the song: The complex multimodal timing of mother-infant musical interaction. *Behavioral and Brain Sciences*. 27 (4) : 516—517. 2004.
104. *Loritz D.* How the Brain Evolved Language. Oxford. Oxford University Press. 2002.
105. *Marantz A., Miyashita Y., O'Neil W.* Image, Language, Brain. Cambridge. The MIT Press. 2000.
106. *McCabe K., Houser D., Ryan L., Smith V., Trouard T.* A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the U.S.A.* 98 : 11832—11835. 2001.
107. *McCrew W. C., Marchant L. F.* On the other hand: Current issues in and meta-analysis of the behavioral laterality of hand function in nonhuman primates. *Yearbook of Physical Anthropology*. 40 : 201—232. 1997.
108. *McKinney M. L.* Evolving behavioral complexity by extending development. *Biology, Brains and Behavior*. Eds Taylor Parker S., Langer J., McKinney M. L. Santa Fe. SAR Press. 25—40. 2000.

109. *McManus I. C.* The inheritance of left-handedness. *Biological Asymmetry and Handedness*. CIBA Foundation Symposium 162. 251—281. 1991.
110. *Mekel-Bobrov N., Gilbert S. L., Evans P. D., Vallender E. J., Anderson J. R., Hudson R. R., Tishkoff S. A., Lahn B. T.* Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science*. 309 (5741) : 1720—1722. 2005.
111. *Miller R.* Axonal Conduction Times and Human Cerebral Laterality: Psychobiological theory. Harwood. Harwood Academic Press. 1996.
112. *Natochin Yu., Chernigovskaya T.* Evolutionary Physiology: History, Principles. *J. of Comparative Biochemistry and Physiology*. 118 (A, 1) : 63—79. 1997.
113. *Newmeyer F. J.* Genetic Dysphasia and Linguistic Theory. *J. of Neurolinguistics*. 10 (2/3) : 47—73. 1997.
114. *Perelle I. B., Ehrman L.* An international study of human handedness: the data. *Behavior Genetics*. 24 : 217—227. 1994.
115. *Phillips W. A., Silverstein S. M.* Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. *Behavioral and Brain Sciences*. 26 (1) : 65—82. 2003.
116. *Pinker S.* Rules of Language. *Science*. 253 : 530—535. 1991.
117. *Pinker S.* The Language Instinct: How the Mind Creates Language. New York. Harper. 1994.
118. *Pinker S., Jackendoff R.* The faculty of language: what's special about it? *Cognition*. 95 (2) : 201—236. 2005.
119. *Pinker S., Bloom P.* Natural Language and Natural Selection. *Behavioral and Brain Sciences*. 13: 707—784. 1990.
120. *Pinker S., Prince A.* On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Cognition*. 28 : 73—193. 1988.
121. *Plunkett K., Marchman V.* From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*. 48 : 21—69. 1993.
122. *Popper K. C., Eccles J. C.* The Self and Its Brain: an argument for interactionism. Springer International. 28 : 73—193. 1977.
123. *Prasada S., Pinker S.* Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*. 8 : 1—56. 1993.
124. *Premack D., Woodruff G.* Does the chimpanzee have a 'theory of mind'? *Behavioral and Brain Sciences*. 4 : 515—526. 1978.
125. *Pulvermüller F.* Words in the Brain's Language. *Behavioral and Brain Sciences*. 22 : 253—279. 1999.
126. *Pulvermüller F., Mohr B.* The Concept of Transcortical Cell Assemblies: A Key to the Understanding of Cortical Lateralization and Interhemispheric Interaction. *Neuroscience and Biobehavioral Rev.* 20 : 557—566. 1996.
127. *Rizzolatti G., Arbib M. A.* Language within our grasp. *Trends in Neurosciences*. 21 : 188—194. 1998.
128. *Rizzolatti G., Craighero L.* The Mirror-Neuron System. *Annual Review of Neuroscience*. 27 : 169—192. 2004.
129. *Rizzolatti G., Fadiga L., Fogassi L., Gallese V.* From mirror neurons to imitation: Facts and speculations. The imitative mind *Development, Evolution, and Brain Bases*. Eds Meltzoff A., Prinz W. Cambridge. 247—266. 2002.

130. *Rosser Z. N., Zerial T., Hurles M. E. et al.* Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *The Am. J. of Human Genetics.* 67 : 1526—1543. 2000.
131. *Rotenberg V. S.* An Integrative Psychophysiological Approach to Brain Hemisphere Functions in Schizophrenia. *Neuroscience and Biobehavioral Rev.* 18 (4): 487. 1994.
132. *Rumelhart D. E., McClelland J. L.* On learning the past tenses of English verbs. Parallel distributed processing: Explorations in the microstructures of cognition. Eds McClelland J. L., Rumelhart D. E. Cambridge, MA. 2 : 216—271. 1986.
133. *Shapiro K., Caramazza A.* The representation of grammatical categories in the brain. *Trends in Cognitive Sciences.* 7 (5) : 201—206. 2003.
134. *Sommer I., Ramsey N., Kahn R., Aleman A., Bouma A.* Handedness, language lateralisation and anatomical asymmetry in schizophrenia: meta-analysis. *British J. of Psychiatry.* 178 : 344—353. 2001.
135. *Suddendorf T., Whiten A.* Mental evolution and development: evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin.* 127 : 629—650. 2001.
136. *Thomson R., Pritchard J., Shen P., Oefner P., Feldman W.* Recent common ancestry of human Y-chromosomes: Evidence from DNA sequence data. *Proceedings of National Academy of Sciences.* 97 (13) : 7360—7365. 2000.
137. *Ullman M. T.* Contributions of memory circuits to language: the declarative/procedural model. *Cognition.* 92(1-2) : 231—270. 2004.
138. *Vogeley K., Bussfeld P., Newen A., Herrmann S., Happé F., Falkai P., Maier W., Shah N.J., Fink G.R., Zilles K.* Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage.* 14 : 170—181. 2001.
139. *von Uexküll J.* *Theoretische Biologie.* Berlin. Springer. 1928.
140. *Young L. J.* The neurobiology of social recognition, approach, and avoidance. *Biological Psychiatry.* 51 : 18—26. 2002

---

# III

## ЕСТЕСТВЕННЫЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

---

---

### СТРУКТУРА И ФУНКЦИИ ЕСТЕСТВЕННОГО ИНТЕЛЛЕКТА В КОНТЕКСТЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА\*

*М.А. Холодная*

#### **Соотношение проблематики искусственного интеллекта (ИИ) и естественного интеллекта (ЕИ)**

При разработке проблемы ИИ обращение к исследованиям ЕИ может принимать разные формы. В одних случаях постулируется необязательность использования психологических знаний об особенностях устройства ЕИ при конструировании системы ИИ, в других – предметом моделирования выступают некоторые частные свойства человеческого интеллекта, в третьих – утверждается невозможность создания ИИ как аналога ЕИ в силу их принципиальной несопоставимости.

В рамках первого подхода признается, что программы и технические устройства, разработанные без привлечения знаний из области психологии интеллекта, могут быть вполне эффективны (например, шахматные программы).

Следование второму подходу приводит к тому, что большинство исследований в области ИИ ведется в русле функционализма, поскольку под интеллектом понимается способность выполнять определенные когнитивные функции (либо способность решать задачи). Соответственно формулируется задача имитации разных проявлений интеллектуальной деятельности человека на уровне операций.

В рамках этого подхода разработана концепция «сильного ИИ», согласно которой свойства разума могут быть присущи логическим действиям любого вычислительного устройства. Под интеллектом в

---

\* Работа выполнена при финансовой поддержке РГНФ (грант № 04 – 06 – 00139 а)

данном случае понимается выполнение некоторой хорошо определенной последовательности операций, или алгоритма. Интеллект – это набор программ, то есть «софт» («железо» рассматривается как несущественный фактор) или универсальный алгоритм в виде алгоритма поиска алгоритмов решения новых задач (А.В. Ворожков).

Не менее типичной для данного подхода является использование определения ЕИ как способности к адаптивному поведению, которое «перебрасывается» на разработки в области ИИ.

Наконец, представители третьего подхода склонны подчеркивать принципиальные различия между ИИ и ЕИ. Так, отмечается, что для ИИ существует только то, что заранее предусмотрено. Иными словами, функционирование ИИ жестко обусловлено набором констант (напротив, правила интерпретации в ЕИ произвольны, индивидуализированы, непредсказуемы).

Отмечается, что результатом интеллектуальной деятельности являются интеллектуальные новации, однако их появление не может быть описано в каких-либо приемлемых формально-логических формах, ибо «естественный человеческий интеллект принципиально случаен» (А.В. Болдачев). Поэтому изучение ЕИ не может ничего дать для создания ИИ.

Д. Сёрл считает, что, хотя разум человека и схож с ИИ по системе организации, однако, каким бы усовершенствованным не был компьютер, он не может обладать ментальностью. Следовательно, способности разума не могут быть воспроизведены в ИИ.

Несмотря на столь различные точки зрения, необходимость конструктивного взаимодействия между этими областями знаний – искусственным интеллектом и психологией интеллекта – является данью времени. Впервые в российской науке о необходимости встраивания психологических знаний в проблематику ИИ заявил О.К. Тихомиров. В 70-е гг. им была создана лаборатория «Психологические проблемы автоматизации умственного труда», сотрудниками которой были развернуты исследования по психологическим аспектам ИИ («Искусственный интеллект» и психология», 1976). И хотя в последние годы разработки в области ИИ вышли на качественно новые уровни, тем не менее понятие ИИ, введенного в научный оборот Дж. Маккарти (1956), на наш взгляд, не может быть содержательно возвращено без определения ЕИ.

### **Основные подходы к определению интеллекта в психологической науке**

В психологии существуют две основные парадигмы, в рамках которых используются разные методы исследования интеллекта и предлагаются разные его определения:

1. Тестологическая парадигма (интеллект – это то, что измеряется с помощью тестов интеллекта; в последнее время это определение уточнено за счет введения понятия «психометрический интеллект»).

2. Экспериментально-психологическая парадигма (интеллект – это свойство интеллектуальной деятельности, обусловленное действием определенных факторов). В свою очередь, среди экспериментально-психологических исследований можно выделить несколько основных подходов:

- 1) Феноменологический подход (интеллект как особая форма организации содержания сознания).
- 2) Генетический подход (интеллект как форма адаптации к требованиям естественной и искусственной среды).
- 3) Социокультурный подход (интеллект как результат социализации, а также влияния культуры в целом).
- 4) Процессуально-деятельностный подход (интеллект как особая форма человеческой деятельности).
- 5) Образовательный подход (интеллект как продукт целенаправленного обучения).
- 6) Информационный подход (интеллект как совокупность элементарных процессов переработки информации).
- 7) Функционально-уровневый подход (интеллект как система разноуровневых познавательных процессов).
- 8) Регуляционный подход (интеллект как фактор саморегуляции психической активности).

Таким образом, в психологических исследованиях интеллекта сложилась традиция определять интеллект как совокупность его свойств: в тестологии – как совокупность результативных свойств (через показатели успешности и скорости решения тестовых задач), в экспериментальной психологии – как совокупность функциональных свойств (тех или иных проявлений интеллектуальной деятельности с учетом факторов, влияющих на выраженность этих свойств – социокультурного, образовательного, нейрофизиологического и т.д.).

В итоге возникла парадоксальная ситуация: зная о свойствах интеллекта и о факторах, влияющих на эти свойства, мы ничего не знаем о том, что собой представляет интеллект как таковой (эффект «черного ящика»). Более того, такой способ анализа породил иллюзию «исчезновения» интеллекта как психической реальности.

### **Онтологическая теория интеллекта: интеллект как форма организации ментального опыта**

В рамках онтологической теории интеллекта обосновывается положение о том, что психическим носителем свойств интеллекта выступает индивидуальный ментальный (умственный) опыт (Холодная, 2002). Наблюдаемые проявления интеллектуальной деятельности



*производны* по отношению к особенностям структурной организации ментального опыта субъекта. Иными словами, вопрос «Что такое интеллект?» (с последующим перечислением его свойств) должен быть переформулирован следующим образом: «Что представляет собой интеллект как психический носитель своих свойств?»

Итак, по своему онтологическому статусу интеллект – это особая форма организации индивидуального ментального (умственного) опыта, особенности состава и строения которого предопределяют свойства интеллектуальной деятельности и, как следствие, продуктивность интеллектуального поведения.

По своему назначению интеллект – это общая познавательная способность, которая проявляется, во-первых, в том, как человек воспринимает, понимает, объясняет и прогнозирует происходящее, во-вторых, в том, какие решения он принимает и, в-третьих, в том, насколько эффективно он действует в тех или иных конкретных ситуациях (прежде всего новых, сложных, необычных).

В онтологической теории интеллекта устройство интеллекта описывается с позиций трех базовых категорий: «ментальные структуры», «ментальное пространство», «ментальная репрезентация».

*Ментальные структуры* – это устойчивая форма ментального опыта в виде системы психических образований, которые в условиях познавательного контакта субъекта с действительностью обеспечивают возможность поступления информации о происходящих событиях и ее преобразование, а также управление процессами переработки информации и избирательность интеллектуальной деятельности.

*Ментальное пространство* – это динамическая форма ментального опыта, поскольку ментальное пространство, во-первых, развертывается наличными ментальными структурами в условиях интеллектуальной деятельности, во-вторых, обладает способностью к одномоментному изменению своей топологии, метрики и содержания (таких своих характеристик, как развернутость границ, проницаемость, динамичность, размерность, иерархичность и т.д.) под влиянием субъективных и объективных факторов (аффективного состояния человека, появления дополнительной информации и т.п.), в-третьих, в ментальном пространстве возможны разного рода мысленные перемещения, при этом характер «движения мысли» зависит от меры его развернутости, структурированности и семантической сложности; в-четвертых, ментальные пространства могут быть «вложены» друг в друга, что создает возможность появления контрфактических, «невозможных» идей.

*Ментальная репрезентация* – это оперативная форма ментального опыта в виде актуального умственного образа того или иного конкретного события, то есть субъективная форма видения, понимания и интерпретации происходящего. Ментальная репрезентация за-

висит от обстоятельств и строится в конкретных условиях для специфических целей (Ришар, 1998).

Понимание особой роли ментальных репрезентаций в раскрытии механизмов интеллектуальных возможностей человека вынудило некоторых авторов полностью пересмотреть традиционный взгляд на проблему интеллекта. Так, Дж. Браун и Е. Лангер ввели понятие «осмысленности» (*mindfulness*), противопоставляя его традиционному понятию интеллекта. Осмысленность – это состояние, в котором субъект способен строить ментальные репрезентации специфического типа, а именно он открыт восприятию привычной информации в новом свете, чувствителен к контексту, способен создавать новые категории, имеет представление о множестве возможных перспектив какой-либо ситуации, отличается вариативностью способов мышления и т.д. (Brown, Langer, 1990).

### **Состав и строение ментального опыта**

В рамках анализа ментальных структур можно выделить три уровня (или слоя) опыта, каждый из которых имеет свое назначение.

1) *Когнитивный опыт* – это ментальные структуры, которые обеспечивают хранение, упорядочивание и преобразование наличной и поступающей информации, способствуя тем самым воспроизведению в психике познающего субъекта устойчивых, закономерных аспектов его окружения (в том числе *архитектурные структуры, способы кодирования информации, когнитивные схемы, семантические структуры, понятийные структуры*). Их основное назначение – оперативная переработка текущей информации об актуальном воздействии на разных уровнях познавательного отражения.

2) *Метакогнитивный опыт* – это ментальные структуры, позволяющие осуществлять произвольную и произвольную регуляцию процесса переработки информации, а также сознательно управлять работой собственного интеллекта (в том числе *произвольный и произвольный интеллектуальный контроль, метакогнитивная осведомленность, открытая познавательная позиция*). Их основное назначение – контроль за ходом интеллектуальной деятельности и состоянием индивидуальных интеллектуальных ресурсов.

3) *Интенциональный (эмоционально-оценочный) опыт* – это ментальные структуры, которые лежат в основе индивидуальных интеллектуальных склонностей (в том числе *предпочтения, убеждения, умонастроения*). Их основное назначение заключается в том, что они предопределяют субъективные критерии выбора определенной предметной области, направления поиска решения, источников информации, средств ее обработки и т.д.

В свою очередь, особенности организации когнитивного, метакогнитивного и интенционального опыта определяют свойства индиви-

дуального интеллекта на уровне частных интеллектуальных способностей (*конвергентные* и *дивергентные способности, обучаемость, познавательные стили*) и интегральных интеллектуальных способностей (*компетентность, талант и мудрость* как проявления интеллектуальной одаренности), а также на уровне индивидуального своеобразия склада ума (*персональный познавательный стиль*) (Холодная, 2002; 2004).

### **Качества (атрибуты) интеллекта как психической реальности (формы организации ментального опыта)**

Можно выделить три теоретических источника онтологической теории интеллекта: 1) структурно-интегративный подход (Веккер, 1976; Чуприкова, 1997 и др.; 2) субъектный подход (С.Л. Рубинштейн, 1997; Брушлинский, 1994); 3) ресурсный подход (Дружинин, 2001). Каждый из этих подходов позволяет сформулировать для психологии интеллекта определенные теоретические следствия и соответственно выделить в интеллекте – как особой психической реальности – некоторые новые его качества.

Структурно-интегративная методология выводит на первый план следующую важную позицию: природа любого психического объекта (в том числе интеллекта) должна быть описана одновременно в терминах субстрата (психического материала), структуры (составных компонентов и их связей) и функций (свойств объекта на уровне его поведения).

Теоретические следствия структурно-интегративного подхода относительно онтологической теории интеллекта:

1) особенности функционирования интеллекта (его свойства) производны по отношению к особенностям состава и строения ментального опыта; возможность анализа природы интеллекта в терминах онтологических категорий, таких как «ментальная структура», «ментальное пространство», «ментальное время», «ментальная ткань» (соответственно, при изучении интеллекта на первый план выходит такое его качество, как **субстанциональность**);

2) трактовка интеллекта как системного образования (**многомерность, самоорганизация, нелинейная динамика**).

Субъектный подход в психологии интеллекта предполагает исследование «внутренних условий» интеллектуальной деятельности, в том числе субъектного опыта, который «изнутри» инициирует и регулирует интеллектуальное поведение. С другой стороны, понятие «субъект» акцентирует индивидуально-своеобразный характер предпосылок и проявлений интеллектуальной активности.

Теоретические следствия субъектного подхода относительно онтологической теории интеллекта:

1) в качестве детерминант индивидуальных различий в интеллектуальных способностях на первый план выходят не генетические и средовые факторы, а те процессы, которые происходят внутри ментального опыта человека (**субъектность**);

2) роль содержательного контекста, который порождается и оформляется в условиях работы интеллекта с учетом ситуационных факторов (**контекстуальность**);

3) изучение индивидуально-своеобразных форм познавательного отношения к происходящему (**уникальность**).

Ресурсный подход предполагает, что в условиях индивидуальной интеллектуальной деятельности может активизироваться только часть ресурсов в определенной последовательности в зависимости от требований ситуации.

*Теоретические следствия ресурсного подхода относительно онтологической теории интеллекта:*

1) возможна одновременная актуализация ресурсов разного уровня и разного типа (**гетерогенность**);

2) интеллект рассматривается не как локальное свойство (стабильная черта) в виде одной позиции на некотором измерении, но как континуум, на оси которого данное свойство может занимать некоторое множество позиций (**континуальность**);

3) отдельная интеллектуальная способность может быть представлена как подвижная точка в некотором n-мерном пространстве, каждый вектор которого характеризует определенный тип ресурса (**мобильность**).

Список вышеперечисленных качеств (атрибутов) интеллекта, таких как многомерность, нелинейная динамика, контекстуальность, мобильность и др. меняет наш взгляд на природу интеллекта. При этом возникает принципиальный вопрос: если ЕИ как психическая реальность обладает набором таких качеств, то можно ли с помощью технических и программных средств смоделировать ЕИ? По-видимому, на сегодняшний день этот вопрос следует оставить открытым.

### **Особенности интеллектуальных ресурсов «экспертов»**

Разработчиков ИИ, конечно же, интересует не ЕИ вообще, а ЕИ в его наиболее продуктивных формах. Такая форма ЕИ присуща, в частности, «экспертам». Эксперт – это человек с высокими реальными достижениями в определенной предметной области; или, в терминах психологического подхода, интеллектуально компетентная личность.

*Признаки интеллектуально компетентной личности («эксперта»):* опытный; обученный; обладающий разными уровнями знаний (декларативным, процедурным, рефлексивным, неявным); адекватно воспринимающий ситуацию с учетом ее объективно значимых аспектов, в том числе в экстремальных условиях; способный к экспертному заключению

(«решающему совету»); систематически успешный в разрешении релевантных сфере его компетентности проблем.

Именно при изучении экспертов, то есть при изучении эффектов реальной интеллектуальной успешности взрослых людей отчетливо обнаружил свою недостаточность тестологический подход, в рамках которого уровень развития интеллекта оценивался в терминах величины IQ. В частности, было показано, что «эксперты» демонстрируют эффект «порога интеллекта», поскольку их IQ, как правило, находится в пределах выше средних значений, не превышая 120 единиц. Кроме того, «эксперты» – это, как правило, люди среднего и пожилого возраста, когда происходит снижение показателей психометрического интеллекта (за счет флюидных способностей), но тем не менее на этом фоне у них наблюдается рост реальной интеллектуальной продуктивности.

На наш взгляд, изучение интеллектуальных ресурсов «экспертов» (или шире – феномена интеллектуальной компетентности) может осуществляться в рамках онтологической теории интеллекта, которая позволяет описать эти ресурсы в терминах особенностей организации индивидуального ментального опыта.

Отличительные особенности интеллектуального ресурса «эксперта» на уровне когнитивного опыта.

1) Одновременное использование разных способов кодирования информации (словесно-логического, визуального, предметно-практического, сенсорно-эмоционального).

2) Высокая гибкость фреймов и семантических структур, за счет чего в интеллектуальной деятельности эксперта возможны «эффекты калейдоскопа».

3) Наличие в когнитивных схемах «ключевых точек», или «точек роста», которые, будучи чувствительными к релевантной информации, заставляют эти схемы работать в режиме интенсивной ассимиляции новой, критически важной информации;

4) Единство декларативных знаний (знания о том, *что*) и процедурных знаний (знания о том, *как*) в структуре понятийного опыта эксперта, высокий уровень их интеграции за счет действия системы операторов, которые способствуют превращению («переводу») декларативных знаний в процедурные.

5) Понятийные знания представлены в развернутом семантическом контексте (именно качество этого индивидуализированного контекста часто определяет инновационный потенциал идей эксперта).

6) Высокая степень артикуляции понятийных знаний (специалист-физик в задаче сразу же выделяет данные, важные для ее решения; врач-рентгенолог среди множества признаков видит решающий признак и безошибочно соотносит его с определенным типом заболевания и т.д.).

7) Свернутый характер понятийных структур (отсюда трудность их осознания и вербализации). Среди фаз образования понятий, таких как мотивация, категоризация, обогащение, перенос, свертывание, – в системе знаний эксперта ключевую роль играет фаза свертывания. Поэтому «эксперты – скорее блестящие опознаватели, нежели глубокие мыслители» (Р. Глезер).

*Отличительные особенности интеллектуального ресурса «эксперта» на уровне метакогнитивного опыта.*

1) Сформированность произвольного интеллектуального контроля, который обеспечивает оперативную избирательную саморегуляцию процесса переработки информации на субсознательном уровне). Его проявления: эффективное структурирование ментальных репрезентаций, торможение импульсивности в процессе принятия решений, привлечение понятий разной степени обобщенности при построении суждений, имплицитная обучаемость в ходе освоения нового вида деятельности и т.д.

2) Способность эксперта произвольно управлять своей интеллектуальной деятельностью (опознавать наличие проблемы, отбирать наиболее эффективные стратегии для ее решения, отслеживать ход интеллектуального поиска, планировать, оценивать, предвосхищать, контролировать свои действия с учетом обратной связи и т. д.).

3) Высокий удельный вес рефлексивных знаний и одновременно их свернутый характер. Согласно «парадоксу инженерии знаний» Уотермана, чем выше способности эксперта в проблемной области, тем менее он способен к описанию знаний, используемых им для решения задачи.

4) Особую роль в структуре интеллектуальных ресурсов эксперта играет метакогнитивная осведомленность (знание своих интеллектуальных особенностей, а также способность анализировать и обобщать свой личный профессиональный опыт, «извлекать из него уроки»).

5) Эксперты – в отличие от новичков – занимаются *глобальным*, или стратегическим планированием, что требует значительных затрат времени.

6) Наличие открытой познавательной позиции: высокий интерес к необычным, парадоксальным ситуациям и готовность осваивать противоречивый, альтернативный, «невозможный» опыт.

*Отличительные особенности интеллектуального ресурса «эксперта» на уровне интенционального опыта.*

1) Особую роль в организации интеллектуальной деятельности эксперта играют индивидуальные предпочтения, убеждения, унаследованные.

2) Высокий удельный вес «неявных знаний» в виде «профессиональной интуиции», «профессионального чутья». Отличительные черты «неявного знания»: приобретаются в собственном практиче-

ском повседневном опыте; не представлены в вербальной форме; являются в основном процедурными; имеют индивидуализированный характер; возникают в режиме «инкубации» и «инсайта» и т.д.

3) Ярко выраженная избирательность в процессе восприятия и понимания происходящего (в том числе содержания своей предметной области).

4) Гетерогенность стратегий переработки информации, которые часто основываются на личном опыте.

Нами было проведено три серии исследований интеллектуальной компетентности в рамках эмпирической модели «эксперт – новичок». В качестве «экспертов» выступили лица, имеющие реальные достижения в разных областях научной деятельности (научно-технической, научно-педагогической и научно-исследовательской).

В *первой серии исследований* в качестве «экспертов» выступили студенты старших курсов и аспиранты технического университета, имеющие реальные достижения в виде патентов, грантов, лицензированных программных продуктов и т.п. (сравнительно с хорошо успевающими студентами старших курсов того же технического университета) (Холодная, Кострикина, Берестнева, 2005). Полученные результаты позволили сделать следующие выводы. Во-первых, существует пороговое значение когнитивных способностей (116 единиц IQ по шкале Амтхауэра), ниже которого реальные достижения отсутствуют и выше которого их прогноз не представляется возможным (иными словами, для экспертов не обязательны высокие и сверхвысокие значениями уровня психометрического интеллекта). Во-вторых, экспертов отличает наличие рефлексивного, мобильного, независимого и гибкого когнитивных стилей (у мужчин), которые являются референтами сформированности способности к произвольному интеллектуальному контролю процессов переработки информации.

Во *второй серии исследований* изучались особенности понятийного и метакогнитивного опыта экспертов – университетских преподавателей физики в статусе кандидатов и докторов наук (сравнительно со студентами старших курсов физико-математического факультета) (Савин, 2002). Полученные результаты позволяют говорить о ряде значимых отличий в особенностях организации понятийного опыта экспертов: они формулируют более сложные проблемы как для профессионально-нейтрального (например, «болезнь»), так и профессионально-значимого («ядерная энергия») понятий; предлагают большее количество разнообразных дополнительных условий при предъявлении незавершенной физической задачи. Что касается метакогнитивного опыта, то экспертов отличает рефлексивный стиль переработки информации (по методике «Сравнение похожих рисунков» Дж. Кагана); более выраженная метакогнитивная осведомленность относительно своих про-

фессионально-важных качеств, препятствующих и способствующих профессиональной деятельности; более высокая толерантность к «невозможной ситуации», в частности, при формировании образа «вымышленного мира» (водная среда обитания людей) они строят более проработанные ментальные модели этой ситуации с использованием более сложных форм экстраполяции.

Кроме того, среди экспертов была обнаружена подгруппа «супер-рефлексивных» (это лица, которые отличаются чрезмерно замедленным временем принятия решения в сочетании с высокой точностью решений).

*Третья серия исследований* была посвящена изучению особенностей метакогнитивного и интенционального опыта лиц, занимающихся научно-исследовательской работой в разных предметных областях. В качестве «экспертов» выступали кандидаты и доктора наук, которые разрабатывают самостоятельное научное направление, возглавляют научно-исследовательскую лабораторию, имеют большое количество публикаций, обладают репутацией в научном сообществе (сравнительно с учеными, которые по характеру их научной деятельности были отнесены к категориям «новички» и «профессионалы») (Полякова, Холодная, 2005).

Полученные в этой серии исследований факты свидетельствуют о том, что ученые-эксперты – сравнительно с «новичками» и «профессионалами» – отличались, во-первых, наиболее медленным временем принятия решений при минимальном количестве ошибок (по методике «Сравнение похожих рисунков» Дж. Кагана), то есть «супер-рефлексивностью»; во-вторых, более выраженной полезависимостью (по методике «Включенные фигуры» Г. Уиткина); в-третьих, гетерогенностью проявлений стилевого поведения. В свою очередь, в подгруппе экспертов были отмечены различные форма проявления неявных знаний (на основе контент-анализа данных интервью и самоотчетов).

По этим трем сериям исследований можно сделать некоторые общие выводы.

1) Когнитивный ресурс в виде IQ, а также метакогнитивный ресурс в виде сформированности произвольного интеллектуального контроля в виде стилевых свойств интеллекта являются важными, но недостаточными предпосылками реальных сверхординарных интеллектуальных достижений в области научной деятельности.

2) Можно высказать предположение об особой роли в организации интеллектуальной деятельности успешных ученых-исследователей, во-первых, понятийного опыта, во-вторых, рефлексивного опыта (в виде произвольного интеллектуального контроля, высокого уровня метакогнитивной осведомленности и открытой познавательной позиции) и, в-третьих, интенционального опыта (в виде предпочтений,



убеждений, умонастроений). Именно эти ментальные структуры обуславливают возможность экстраординарных интеллектуальных достижений, блокируя негативные возрастные тенденции интеллектуального развития (напомним, что в наших исследованиях все эксперты – более старшая возрастная категория).

Полученные данные согласуются с результатами некоторых других психологических исследований. В частности, отмечается, что достижение экспертного уровня предполагает формирование системы практических и ментальных навыков, которыми человек овладевает в ходе длительной, «осознанной практики» (Ericsson, Krampe, Tesch-Romer, 1993). Подчеркивается тот факт, что эксперты приобретают интеллектуальную компетентность только после огромных затрат времени на изучение соответствующей предметной области, накопления опыта принятия решений и т.д. Отличительными признаками «осознанной практики» являются: высокий уровень мотивации к обучению; постоянная обратная связь на оценку правильности своих действий либо их ошибочность; глубина и тщательность обработки материала; инициативность и самостоятельность.

В последние годы было продемонстрировано ключевое значение мотивации и ценностных ориентаций в развитии интеллектуальной компетентности (Равен, 2002). Исследования Дж. Равена показали, что высококвалифицированные специалисты (компетентные лица) обладают рядом специфических личностных особенностей: они чрезвычайно целенаправленны в решении проблемы, их отличает высокий уровень инициативы и энтузиазма, они характеризуются готовностью преодолевать трудности и препятствия на пути своего профессионального роста, им присущ высокий уровень ответственности за свое дело и т.п.

### **Психологические аспекты разработки экспертных систем**

Экспертные системы (ЭС) – это воплощение в техническом устройстве (и программном обеспечении) опыта эксперта; ЭС использует знания и процедуры вывода для решения задач, являющихся трудными для человека. Выделяются разные типы ЭС в зависимости от характера задач: интерпретация данных; диагностика; мониторинг; проектирование; прогнозирование; обучение; поддержка принятия решений.

Необходимость учета психологических аспектов разработки ЭС в значительной мере связана со спецификой интеллектуальных ресурсов эксперта. Прежде всего, это проблемы, касающиеся методов извлечения экспертных знаний. Извлечение знаний из эксперта предполагает описание базы знаний (объектов, понятий, признаков, связей, правил вывода); способов обработки информации; средств деятельности, используемых экспертом. Как известно, часто эксперты опи-

сывают правдоподобные способы рассуждений, мало похожие на те, которыми они действительно пользуются, при этом не всегда помогает метод рассуждения вслух при решении конкретных задач.

Выше уже отмечалось, что в структуре знаний эксперта важное место занимают «неявные знания», извлечение которых требует специальных приемов, в том числе таких, как метод критических ситуаций (интервью о примерах эффективного и неэффективного поведения в профессиональных ситуациях, когда требуется принять решение или разрешить конфликт); имитационные тесты (наблюдение за поведением человека в профессионально значимых условиях); метод оценки ситуаций (беседа, сфокусированная на личном профессиональном опыте человека); использование метафорического подхода.

Актуальным является вопрос о критериях, на основе которых разрабатывается та или иная ЭС. Считается, что мощность ЭС зависит от количества и качества знаний, которые в ней хранятся. Однако ценность эксперта определяется не его многознанием, а уникальной способностью выявлять релевантные знания в потоке информации. Неудивительно, что с конца 90-х гг. одна из центральных функций ЭС определяется как «обнаружение знаний в области баз данных» на основе современных процедур анализа данных (например, Data Mining). По сути, разработка современных ЭС – это метод построения теории экспертных способностей и теории принятия решений.

Наконец, особую актуальность приобретают методики, позволяющие адекватно идентифицировать экспертов, ибо не каждый специалист – это эксперт (в том числе с учетом психологического своеобразия интеллектуальных ресурсов лиц, достигших в своем развитии экспертного уровня).

Таким образом, психологический взгляд на экспертов позволяет обсудить некоторые новые варианты решения вопроса о соотношении ИИ и ЕИ. В частности, наряду с классическим ИИ и синергетическим ИИ можно говорить о *симбиотическом ИИ* (как человеко-машинной системы, обладающей экспертными знаниями; как средства организации отношений сотрудничества в решении проблем и т.д.). В целом, целесообразность интеграции наших представлений об ИИ и ЕИ на основе междисциплинарных исследований представляется очевидной.

В заключение хотелось бы сослаться на мнение Ж.-Л. Лорьера о том, что ИИ – это экспериментальная дисциплина, ориентированная на то, чтобы с помощью математических моделей (программ) лучше понять функционирование человеческого разума.

ЛИТЕРАТУРА

1. *Брушлинский А.В.* (1996). Субъект: Мышление, учение, воображение. –М.– Воронеж: НПО «Модэк».
2. *Веккер Л.М.* (1976). Психические процессы. Мышление и интеллект. Т.2. – Л.: Изд-во Ленингр. ун-та.
3. *Дружинин В.Н.* (2001). Когнитивные способности: Структура, диагностика, развитие. – М.: ПЕР СЭ; СПб.: ИМАТОН-М.
4. «Искусственный интеллект» и психология (1976) / Под ред. О.К. Тихомирова. М.: Наука.
5. *Полякова Е.Г., Холодная М.А.* Неявные знания как один из факторов интеллектуальной компетентности в научно-исследовательской деятельности // Ярославский психологический вестник. Выпуск 13 / Москва – Ярославль.: Изд-во Рос. психологическое общество, 2004. С.159-164.
6. *Равен Дж.* (2002). Компетентность в современном обществе: Выявление, развитие и реализация. – М.: «Когито-Центр».
7. *Ришар Ж.Ф.* (1998). Менгальная активность. Понимание, рассуждение, нахождение решений. – М.: Изд-во «Ин-т психологии РАН».
8. *Рубинштейн С.Л.* (1997). Избр. философско-психологические труды. – М.: Наука.
9. *Савин Е.Ю.* (2002). Понятийный и метакогнитивный опыт как основа интеллектуальной компетентности. Дис. на соиск. уч. степ. канд. психол. н. – М.: Ин-т психологии РАН.
10. *Холодная М.А.* (2002) Психология интеллекта: Парадоксы исследования. – 2-е изд., перераб. и доп. – СПб.: Питер.
11. *Холодная М.А.* (2004). Когнитивные стили: О природе индивидуального ума. – 2 изд., перераб. и доп. – СПб.: Питер.
12. *Холодная М.А., Берестнева О.Г., Кострикина И.С.* (2005). Когнитивные и метакогнитивные предпосылки интеллектуальной компетентности в области научно-технической деятельности // Психол. журнал. Т. 26. № 1. С. 51-59.
13. *Чуприкова Н.И.* (1997). Психология умственного развития: Принцип дифференциации. – М.: Столетие.
14. *Brown G., Langer E.* (1990). Mindfulness and intelligence: A comparison // *Educ. Psychologist.* – V. 25 (3–4). – P. 305-335.
15. *Ericsson K.A., Krampe R.T., Tesch-Romer C.* (1993). The role of deliberate practice in the acquisition of expert performance // *Psychological Review.* – V. 100. – P. 363-406.

---

# ТРАНСЦЕНДЕНТАЛЬНАЯ ИНТЕРСУБЪЕКТИВНОСТЬ, ПРОБЛЕМА «ЧУЖИХ СОЗНАНИЙ», ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ\*

*Н.М. Смирнова*

В рамках современной философии искусственного интеллекта компьютерные технологии рассматриваются как средства целенаправленного изменения конститутивных механизмов сознания. Фантастически расширяя возможности «естественного» интеллекта, они по-новому высвечивают проблемы «предельных оснований» познания. И хотя проблема конституирования (интенциональной) предметности имеет более чем вековую историю, конститутивные механизмы сознания и сегодня изучены крайне фрагментарно.

Неявной когнитивной презумпцией классической философии сознания является идея абсолютной «прозрачности» собственного сознания для мыслящего субъекта – полнота саморефлексии как в отношении содержания мышления, так и его предпосылок. Неклассическая философия сознания исходит из фундаментального допущения недостижимости полноты саморефлексии что, по справедливому замечанию П. Рикера, «сначала феноменология, а затем герменевтика непрестанно относили ко все более отдаленному горизонту». Постулат «многослойности» сознания и «непрозрачности» его глубинных слоев для обыденного мышления лежит в основе теории и практики психоанализа – как классического (фрейдовского), так и философски более изощренных позднейших версий (Бинсвангер, Босс и др.).

Другим важнейшим допущением неклассической теории познания является представление об изначальной нагруженности «объективного» знания операционально-инструментальными характеристиками деятельности, с одной стороны, а также ценностями и жизненными смыслами, с другой. Неклассическая философия сознания отказалась от упрощенных просветительских представлений о сознании как «зеркале души», разработав философски изощренные концепции сознания как многомерного культурного феномена. Крупнейшую брешь в классической философии сознания пробил установка *трансцендентализма* немецкой классической философии. «Расщепив» сознание на эмпирическую (т.е. чувственную, психическую) и трансцендентальную (родовую, обусловленную априорными формами опыта и категориями рассудка) составляющие, И. Кант сосредоточил внимание на исследовании конститутивных механизмов «родового» сознания, т.е. сознания

---

\* Исследование выполнено по гранту РГНФ 04-02-00274а

«чистого» (трансцендентального) субъекта. Трансцендентализм Канта сыграл решающую роль в формировании предпосылок «феноменологического поворота» в философии XX столетия.

Наиболее развитые концептуальные формы исследования конститутивных механизмов работы сознания обрело в феноменологии Э. Гуссерля. Следуя кантовским традициям, Э. Гуссерль полагал, что теория познания имеет смысл лишь как исследование конститутивных механизмов человеческого сознания. Сегодня именно феноменология дает самое глубокое теоретическое обоснование форм опосредованности предмета познания характеристиками когнитивной деятельности человека.

В рамках феноменологического проекта изучение конститутивных механизмов сознания предстает как систематическое описание интенциональной активности трансцендентально чистого сознания по конструированию предметов опыта (интенциональной предметности). В ноэтико-ноэматическом единстве Э. Гуссерля привычные для классической рациональности «субъект» и «объект» познания выступают как моменты органической целостности, корреляты «сознающее – познаваемое», взаимно трансформирующие друг друга в процессе обретения нового опыта. Феноменологическое конституирование – это всеобщий процесс «возведения» смысла трансцендентальных предметностей как центров мировых отношений. Какое же отношение имеют теоретические изыскания феноменологии к проблеме компьютерного мышления?

Развитая Э. Гуссерлем в «Картезианских размышлениях» конститутивная методология является одной из философски наиболее изощренных форм исследования креативных способностей человеческого мышления. Механизмы феноменологического конституирования представляют собой универсальные принципы или идеальные схемы предметности, с помощью которых сознание образует и удерживает устойчивые и протяженные во времени смысловые единства – идеальные предметы. Предпосылкой смыслового синтеза, т.е. образования синтетических смысловых единств, является осознание временной протяженности внутреннего восприятия. Такая протяженность означает, что предмет конституируется сознанием в синтезе различных перспектив, в которых он воспринимается и помнится. Протяженность внутреннего восприятия позволяет к ранее образованным синтетическим единствам присоединять все новые и новые свойства, в результате чего ранее конституированный предмет обогащается все новыми свойствами по мере обретения нового опыта. Синтетическая деятельность сознания как фундаментальная форма конститутивной активности трансцендентального субъекта, т.о., постоянно расширяет область его горизонта. Горизонт же трансцендентального сознания по определению шире горизонта опыта эмпириче-

ского субъекта, поскольку охватывает все многообразие потенциальных предметностей, «возможных миров» последующего освоения. *Феноменологическое конституирование, т.о., сродни конструированию виртуальных объектов – как постоянное обращение к горизонту «возможного», «не запрещенного» – без каких-либо референций к его онтологическому статусу.* Феноменологическое конституирование и компьютерное моделирование смыслов – это создание продуктов сознания особого рода, экзистенция которых вполне исчерпывается формулой: *«существовать – значит быть возможным».*

Наиболее сложным объектом феноменологического конституирования является коммуникативный партнер, – Другой, обладающий собственным сознанием. Это «фундаментальное затруднение» коренится в самой природе феноменологического конституирования. Устойчивые смысловые образования – трансцендентальные предметности – формируются в сфере чистого сознания трансцендентального Эго, которое открывается в результате трансцендентально-феноменологической редукции. Понятно, что подобные образования неотделимы от сознания (трансцендентального), в рамках которого конституированы. Но как же в подобном случае обстоит дело с другими людьми, бытие которых не сводимо лишь к «бытию в возможности» в структурах сознания трансцендентального субъекта? На каком основании можно утверждать, что Другой не только «порождается» процедурами феноменологического конституирования (в пределах самого сознания), но и «осознается» им? Э. Гуссерль, как известно, обращается к рассмотрению этой амбивалентной по своей логической природе задачи в Пятом «Картезианском размышлении».

Феноменологическая теория intersубъективности как аналогизирующей апперцепции чужой телесности (с «первого творения» – собственного тела) не раз подвергалась критике последователями Э. Гуссерля (М. Шелером, А. Шюцем, М. Мерло-Понти, Х. Ортегой и др.). Но проблема intersубъективности имеет более широкий культурный контекст, нежели гуссерлева защита трансцендентальной феноменологии от обвинений в солипсизме. Вспомним роман С. Лема «Солярис». Странные фантомные создания являются землянам на этой загадочной планете. Они до боли знакомы: оставленные дети, утраченные возлюбленные... Каков их онтологический статус, существуют ли они «на самом деле» или рождаются в сознании энергетикой нравственного чувства (раскаяния, желания как-то изменить результаты прошлых поступков)? Компьютерные технологии продуцируют собственные фантомные создания, культурно-антропологический статус опыта общения с которыми проблематичен. Но что означает это «на самом деле» применительно к виртуальному пространству, продуцируемому современными компьютерными технологиями? Каков статус собственной телесности в виртуальном общении? Так сакрамен-

тальный вопрос бытия («Быть или не быть»?) обретает новые культурные очертания.

В виртуальном общении Другой (партнер по «чату») существует лишь как «симулякр» (Ж. Бодрийяр), сообщение, система знаков. Он мне лично («телесно») не знаком, да, вероятно, и не будет. Если он(а) пишет на русском, его гендерную принадлежность можно определить по глагольным окончаниям, на «компьютерном» английском – нет. А может быть, он и вовсе не существует, и мне отвечает усовершенствованный mail-demon? Оставим специалистам вопрос о совершенствовании теста Тьюринга. Философу же интересен вопрос, каков культурно-антропологический статус моего опыта с виртуальным коммуникативным партнером. Но это лишь фрагмент более общей проблемы того, каким образом конституируется смысл Другого в виртуальном общении. Ведь аналогизирующей проекции моей телесности здесь быть не может! И если я приписываю смысл сообщению как «послание Другого», то, быть может, аналогия не кинэстетических движений, а аналогия «наделения смыслом» является для виртуального общения конститутивной?

Здесь представляется целесообразным более подробно изучить концепцию интересубъективности и возможности её применения в проблемном поле дискурса философии искусственного интеллекта.

Концепция интересубъективности – философский провозвестник эволюции ценностных приоритетов классического Модерна – выражает недовольство философского разума крайними формами субъективистской онтологии, социологическим атомизмом и гносеологической робинзонадой, глубоко укорененных в философском самосознании индустриального общества. Их социально-онтологическими предпосылками стали процессы модернизации в Европе, разорвавшие свойственную традиционным обществам родовую пуповину личной зависимости и кланово-родовой идентификации. Ф. Бродель и основанная им школа «Анналов» усматривают социально-экономическую основу суверенизации личности в развитии торговли на дальние расстояния<sup>1</sup>. Подобная «профессиональная миграция», часто и подолгу отрывавшая человека от насиженных социокультурных гнезд и побуждавшая его обживать иные социокультурные миры, инспирировала углубленную культурную рефлексию в отношении своих собственных. Эпоха ранних буржуазных революций знаменует собой завершение процессов утверждения ценности индивидуальности, нередко в жесткой борьбе с кланово-родовыми и конфессионально-корпоративными интересами.

---

<sup>1</sup> Бродель Ф. Время мира. Материальная цивилизация. Экономика, капитализм. М., 1992.

Для высокотехнологичного постиндустриального общества проблема intersубъективности встает не только в контексте полемики с социологическим атомизмом и персонализмом, но и в связи с новыми формами коммуникации человека с фантомами техногенной цивилизации. Отдаленным провозвестником современного литературно-философского осмысления виртуальных коммуникаций является научно-фантастический роман С. Лема «Солярис», где объект коммуникации – вихревые возмущения в мыслящем океане загадочной планеты – способен визуализировать вполне реальные, «земные» объекты человеческих переживаний. В информационном же, компьютеризированном обществе, где виртуальное общение отнюдь не требует перемещения в космическом пространстве, проблема intersубъективности обретает еще большую остроту. Виртуальные животные («Тамагочи»), виртуальные друзья – завсегдатаи чат-сайтов, даже виртуальные возлюбленные... Каковы культурно-антропологические принципы конституирования смысла подобных фантомных созданий и каков культурно-эпистемологический статус нашего опыта общения с ними? Вопрос не праздный, если мы хотим оставаться людьми, а не расчеловечиться в «приставки к компьютерам».

Intersубъективность как проблема философского рационализма питается и осознанием социокультурной ограниченности классической рациональности с присущим ей картезианским противопоставлением материи и сознания, субъекта и объекта. Наиболее развернутое исследование проблема intersубъективности обрела в феноменологии Э. Гуссерля и его последователей. И для некоторых из них, например, М. Хайдеггера и Ж. П. Сартра, концепция intersубъективности как «феноменологическая прививка» (П. Рикер) против наивного объективизма классической философии, куда более значима, чем традиционная тема субъект-объектного отношения. Фундаментальный для самосознания неклассической науки факт изначальной зависимости «объективного» знания от операционально-инструментальных характеристик человеческой деятельности, с одной стороны, а также осознание недостижимости полноты саморефлексии, что, по справедливому замечанию П. Рикера, «сначала феноменология, а затем герменевтика непрерывно относили ко все более отдаленному горизонту»<sup>2</sup>, с другой, до основания потрясли устои классического (картезианского) рационализма. Но размывание наивного объективизма классической философии инспирировало поиск концептуальных средств выражения имманентной связи, взаимного «перетекания» характеристик субъективного и объективного в процессе духовно-практической деятельности человека. Таков пафос всей постклассической философии, обретший зрелые концептуальные формы в феноменологии Э. Гуссерля. Се-

---

<sup>2</sup> Рикер П. Герменевтика. Этика. Политика. М., 1995. С. 79.



годня именно феноменология дает самое глубокое теоретическое обоснование «неклассичности», т.е. форм опосредованности предмета познания характеристиками когнитивной человеческой деятельности.

Основоположник феноменологии отчетливо осознавал все трудности и подводные камни на пути решения проблемы intersубъективности. Тем не менее, он полагал, что она пугает лишь новичков в философии, в страхе шарахающихся от таких философских ярлыков, как солипсизм, психологизм и релятивизм. Э. Гуссерль утверждал, что в отличие от «онтологического солипсизма», эгологическая установка феноменологии является «трансцендентальным солипсизмом», т.е. исключительно методологическим принципом решения наиболее фундаментальных, «претендующих на высшее достоинство» теоретико-познавательных проблем. Однако сознавая недостаточную теоретическую проработку подобного «трансцендентального аргумента», Э. Гуссерль выстраивает свои доказательства таким образом, чтобы показать, что бытие Другого в трансцендентально-феноменологической сфере чистого сознания трансцендентального Эго не только возможно, но и «трансцендентально необходимо». Но значение изложенной им в 5 «Картезианском размышлении»<sup>3</sup> теории intersубъективности оказались куда большим, чем изначально задуманное опровержение обвинений в солипсизме. По завершении работы над ним он усматривал в развитой им трансцендентальной теории intersубъективности рациональный фундамент всех социальных наук – *обоснование социальности как intersубъективности высшего порядка*.

Самая существенная трудность философского анализа проблемы intersубъективности состоит в том, что духовное начало человека, его дух, сознание- объект особого рода, который никогда не дан другому непосредственно. Даже самый гениальный актер, вжившись в роль своего персонажа, лишь обыгрывает в собственном сознании ситуации чужого бытия, из которого, как из экскурсии, возвращается в свое Иное. В свою очередь, опосредованные подходы к проблеме «чужих сознаний» могут носить как рациональный, так и иррациональный характер. Наибольшую популярность в начале XX века приобрела *проективная теория эмпатии*, предлагавшая интуитивистски-герменевтический подход к решению проблемы «чужих сознаний». Так, У. Дильтей полагал, что хотя понимание и носит сугубо личностный характер («понять – значит пережить лично»), но «вчувствования» в смыслы других эпох и культур становится возможным во все возрастающей степени благодаря «слиянию культурных перспектив», т.е. унифицирующей роли модернизации. Однако ни «вживание», ни

---

<sup>3</sup> Husserl E. Cartesianischerr Meditationen und Pariser Vortrage. Haag, 1950. V Meditation. S. 121-177.

«наведение», ни «настройка», строго говоря, не являются рациональными, т. е. не следуют конвенциональным правилам общезначимых процедур. Трансцендентально-феноменологический подход к проблеме intersубъективности тем и отличается от неприемлемой для Э. Гуссерля проективной теории эмпатии, что сфокусирован на выявлении всеобщих и необходимых процедур феноменологического конституирования Другого, а потому в феноменологической постановке проблема intersубъективности, предстает скорее как методологическая, чем собственно эпистемологическая.

В рамках феноменологического проекта решение проблемы intersубъективности предстает как развернутое описание интенциональной активности Эго по конституированию Альтер-Эго, т.е. *способов конституирования смысла* Другого как «трансцендентально необходимого». Самый общий принцип феноменологического подхода состоит в том, что доступ к сознанию Другого, закрытого для непосредственного постижения, способна открыть лишь аналогизирующая апперцепция смысла собственного сознания, выступающего как «первое творение» и его последующая аппрезентация («осовременивание»). Иными словами, ее главная задача состоит в том, чтобы продемонстрировать аналоговую проекцию смысла собственного Я на чужую телесность.

Осуществление операции трансцендентально-феноменологической редукции, т.е. «заклочки в скобки» смыслов бытийных суждений и психологических предикатов («естественной установки сознания»), в результате чего открывается доступ к чистому сознанию трансцендентального Эго, является универсальной для конституирования любых идеальных предметностей как центров мировых отношений, будь то числа, геометрические фигуры, отношения последовательности и подобия и т. д. Для феноменологического же конституирования смыслов Другого как Alter Ego необходима дополнительная абстрагирующая операция, не используемая для конституирования смыслов иных интенциональных предметностей – редукция к «сфере принадлежности», или *примордиальной сфере*. Подобная редукция – уже в пределах поля чистого опыта трансцендентального Эго- элиминирует все проявления intersубъективности Другого, не-моего, т. е. того, к чему я не имею непосредственного доступа. Иными словами, если трансцендентально-феноменологическая редукция «выводит из игры» естественную установку сознания, тем самым открывая доступ к чистому сознанию трансцендентального Эго, то редукция к примордиальной сфере – это дополнительная абстрагирующая операция, проведенная в пределах трансцендентально-феноменологической сферы, для решения определенной познавательной задачи – выявления смысла «лишь мне прису-

щего», т.е. собственного Я<sup>4</sup>. В пределах этой примордиальной сферы, по Э. Гуссерлю, обнаруживается объект, который дан мне непосредственно, дорефлексивно, как поле собственного сенсорного опыта. Он называет его моим собственным телом и приписывает ему все атрибуты субъективности, начиная от простейшей чувственности и кончая мышлением. Чувственный центр подобных телесных переживаний и есть мое персональное Я. Оно контролирует телесные движения, чувствует, переживает и мыслит. И если же в сферу принадлежности попадает другой объект, телесные движения которого интерпретируемы по аналогии с моими собственными, происходит смысловое «спаривание» двух интенциональных объектов путем «пассивного синтеза» с последующим переходом ко множественным конфигурациям. При этом Э. Гуссерль особо подчеркивает, что *Paarung* – не логический вывод, а спонтанное объединение сознанием двух схожих предметов в смысловое единство, аналогизирующий перенос смысла организма с моего тела на объект, кинестетические движения которого сходны с моими собственными. Я воспринимаю тело Другого как психофизическое единство, благодаря «эмпирически сходной манере» управлять телесными движениями. (Заметим, что для Сартра бытие другого открывает не видимое единство взаимосогласованных движений, но *взгляд* человека). Это разновидность ассоциативного мышления, апеллирующего к собственному прошлому опыту, не данному в актуальном восприятии, – ибо феноменологическая теория рефлексии предполагает, что собственный опыт может быть постигнут лишь ретроспективно, как уже истекший, *in modo praeterito*. Именно поэтому восприятие попавшего в сферу принадлежности интенционального объекта с атрибутами поведения нуждается в «осовременивании», в аппрезентации (т.е. актуализации прошлого опыта применительно к настоящему). В результате подобной операции последняя обретает смысл подобного мне психофизического единства, способного управлять собственными телесными («кинестетическими») движениями. Благодаря этому ей можно приписать атрибуты *поведения* (*des Gebaren*). Но обосновывая независимость Alter Ego от «первого творения», Э. Гуссерль подчеркивает, что второе Я является именно Другим, а не просто дубликатом смысла собственного Я, благодаря различию в модусах пространственной данности. Иными словами, Другой не есть Я, потому что Я «здесь» (*hic*), а он «там» (*illic*).

Первой формой общности, существующей между мною и Alter Ego, закладывающей фундамент всех прочих интересубъективных общностей более высокого порядка, является общность *природы*, смысл которой принадлежит не только к примордиальной, но транс-

<sup>4</sup> *Husserl E. Cartesianischerr Meditationen und Pariser Vortrage. Haag, 1950. V Meditation. Sec. 44.*

цендентальной сфере. Социальные общности имеют характер Alter Ego высших порядков, но феноменологически понятая социальность производна от смысла Другого. А поскольку смысл Другого задан принципами феноменологического конституирования, то все возможные формы социальных сообществ априори присутствуют в сознании трансцендентального Эго, а потому лишь случайным образом могут быть не раскрыты. Их описание во всех возможных культурных конфигурациях и есть задача феноменологической культурологии, создание которой Э. Гуссерль завещал своим ученикам.

Нетрудно видеть, что трансцендентально-феноменологическая теория intersубъективности заряжена фундаментальным противоречием. Оно состоит в том, что феноменологически понятая intersубъективность принадлежит исключительно сфере чистого сознания трансцендентального субъекта. Все Пятое «Картезианское размышление» пронизано противоречивым стремлением конституировать Другого как Иного, Alter Ego по отношению трансцендентальному субъекту, и вместе с тем, из ресурсов его собственной Эго-интенциональности. Именно поэтому одна лишь ссылка на различие модусов пространственной данности *hic* и *illic* не смогла удовлетворить последователей Э. Гуссерля.

Так, развивая далее идеи основателя феноменологической школы, М. Мерло-Понти, к примеру, полагает, что само понятие интенциональности следует трактовать более широко. Это не только объект моего мышления, на который оно с необходимостью направлено (т. к. мышление, как показал еще Ф. Brentano, не может не быть интенциональным), но и вся моя жизненная ситуация, проживаемая в «живом настоящем». *Социальное – вариация моей отдельной жизни*, в которой я участвую как Alter Ego. Подобное представление о социальности как варьировании единичного с целью усмотрения его «сущности» восходит к гуссерлевскому представлению о *Wesenschau* – интуитивному усмотрению сущности (вспомним мысленный эксперимент Э. Гуссерля по варьированию таких свойств кубика, как размеры, цвет, материал, плотность как феноменологический метод интуитивного усмотрения сущности «кубичности»). Согласно Х. Ортеги-и-Гасету, человеческая жизнь Другого всегда латентна и гипотетична. Его реальность является реальностью второго порядка, квази-Я, сопричастное, но ни при каких обстоятельствах не данное мне. Я воспринимаю эту реальность как непроблематизируемую, *неспоримую данность*. Способность Другого соответствующим образом реагировать на мои действия является первым социальным фактом, фундаментальным для конституирования общего окружения. Но в целом он весьма критичен в отношении теории аналогизирующей апперцепции Э. Гуссерля, поскольку она, по справедливому замечанию Ортеги, не учитывает того принципиального обстоятельства, что

моему наблюдению открыты лишь *внешние* проявления чужой телесности, в то время, как свое собственное тело я испытываю в опыте *изнутри*. Подобное различие в опыте восприятия своей и чужой телесности не может быть сведено к различию лишь пространственных перспектив «здесь» (Я) и «там» (Другой). Более того, как возможна подобная апперцепция, если, допустим, я – мужчина, а другой – женщина? (сам Э. Гуссерль предусмотрел лишь случай, когда телесность калеки конституируется как отклонение от нормы). Поэтому, убежден Ортега, гуссерлевский Другой – всегда абстрактен. Феноменологическое конституирование предстает не как конституирование диалогового единства Ego и Alter Ego, а как «*репликация монологичности*». Более того, он отвергает гуссерлево понятие коллективности как интересубъективности более высокого порядка, т.е. утверждение о том, что любые формы социальных общностей в зародыше содержатся в чистом сознании трансцендентального Эго. Свой собственный анализ проблемы интересубъективности Ортега осуществляет не в трансцендентально-редуцированной сфере, а в сфере «естественной установки сознания», от которой сам Э. Гуссерль абстрагируется в процессе трансцендентально-феноменологической редукции. Поэтому для него первично столь значимое и для М. Хайдеггера понятие *des Man*, тогда как понятие собственного Я вторично и предстает как его конкретизация.

Но наиболее глубоким и последовательным критиком гуссерлевской теории интересубъективности является А. Шюц<sup>5</sup>. Он не только подвергает сомнению разрешимость проблемы интересубъективности в рамках трансцендентально-редуцированной сферы, но и обнаруживает серьезные натяжки в рассуждениях Э. Гуссерля в пределах собственно феноменологического стиля философствования.

1. Во-первых, он указывает на недостаточность фундаментального для феноменологии понятия трансцендентально-феноменологической редукции применительно к решению проблемы интересубъективности. В результате осуществления трансцендентально-феноменологической редукции мы получаем доступ к собственному потоку сознания, открытому лишь для внутреннего опыта – «монаду без окон». Этот поток сознания интенционально отнесен к моему жизненному миру, от веры в реальное существование которого на трансцендентальном уровне (абстрагированном ото всех бытийных суждений) я воздерживаюсь («заключаю в скобки»), тогда как в самом жизненном мире бытие других является «неоспоримой данностью». И то, что в рамках трансцендентально-редуцированной сферы приходится вновь возвращаться к

---

<sup>5</sup> См. *Natalia Smirnova. A. Schutz's Contribution To Phenomenological Theory of Intersubjectivity // A.T. Tymieniecka* (ed.) *Analecta Husserliana. LXXIX.* Kluwer Academic Publishers. P.P 313-319.

проблеме элиминации конститутивных элементов активности Другого в процессе редукции к сфере принадлежности (примордиальной сфере), скорее подтверждает справедливость данного в естественной уставовке сознания, чем опровергает его.

Но и здесь возникает не менее серьезный вопрос. Как возможно абстрагироваться ото всех значений, отсылающих к Другому? Иными словами, значения, относящиеся к другому, должны с неизбежностью содержаться в *самом критерии несводимости к Другому*. А это делает операцию редукции к сфере принадлежности самопротиворечивой. Тем не менее, А. Шюц убежден в том, что Пятое «Картезианское размышление» Э. Гуссерля вносит весомый вклад в разрешение проблемы интерсубъективности, поскольку независимо от того, в какой степени ему лично удалось совладать с этой проблемой, он продемонстрировал трудности и противоречия, неизбежно возникающие на пути решения проблемы интерсубъективности в рамках трансцендентальной феноменологии.

Серьезный вклад в решение проблемы интерсубъективности вносит М. Шелер<sup>6</sup>. Он убежден в том, что неудачи всех прежних попыток разрешения этой проблемы коренятся в недостаточно четком различении как уровней этой проблемы, так и более частных вопросов, требующих предварительного уяснения. Он различает следующие уровни:

1. Онтологический. Что чему предшествует в философско-онтологическом плане: человек обществу или общество человеку? Имеет ли общество статус реальности («фактуального существования») независимо от фактического существования конкретных индивидов?

2. Логико-эпистемологический уровень. Как возможно обрести доступ к чужому сознанию, не данному непосредственно?

3. Уровень феноменологической психологии. Каков индивидуальный опыт и каковы формы конститутивной активности трансцендентального Эго делают возможным конституирование смысла Другого? Например, предполагает ли он, что я уже обладаю смыслом самосознания, Природы?

4. Уровень ценностных суждений. Нет сомнений в том, убежден М. Шелер, что такие моральные акты, как ответственность, долг, благодарность по самой природе отсылают к Другому. М. Шелер называет их «существенно социальными актами». Они не могут быть конституированы как до- или пред-социальные.

М. Шелер не разделяет ни с концепцию аналогизирующей апперцепции, ни проективную теорию эмпатии. Каковы же его аргументы? Животные и дети, не обладающие способностью к аналогизирующей апперцепции в достаточной мере, не испытывают сомнения в существовании Другого и способны схватывать симптомы его психической

---

<sup>6</sup> Scheler M. The Nature of Sympathy. New Haven. 1954. P.P. 213-264

жизни, о чем свидетельствуют данные генетической психологии и культурной антропологии. Далее, мы обретаем знание о собственных телесных движениях с помощью всего аппарата чувственного восприятия, в то время, как жесты Другого даны нам лишь как оптические явления, не имеющие аналогов в нашем собственном чувственном восприятии. Следовательно, аналогизирующая апперцепция жестов Другого уже предполагает психическое существование Другого.

В свою очередь, проективную теорию эмпатии М. Шелер считает лишь гипотезой, объясняющей причину нашего верования в существование Другого. Ибо интерпретация жестов Другого как выражений его психической жизни основана на неявном предположении, что Другой *уже* существует. Сама возможность подобной интерпретации есть следствие, а не доказательство его существования.

Анализируя аргументы М. Шелера, А. Шюц заключает, что его критика не колеблет главной неявной посылки обеих гипотез, а именно предположения о том, что: 1) наше собственное Я дано нам изначально; 2) первое, что мы схватываем в другом человеческом существе, – это его внешность.

М. Шелер убежден в том, что обе теории недооценивают трудности самовосприятия и переоценивают трудности схватывания чужого мышления. Так, М. Шелер убежден в том, что возможны ситуации, в которых мы даже не в состоянии распознать, является ли та или иная мысль нашей собственной или нет. Мысль сама по себе не содержит знаков принадлежности к чьему-либо сознанию. Аргументация М. Шелера основана на данных современной ему детской психологии, свидетельствующей, что ребенок сравнительно поздно открывает в себе собственную индивидуальность. Поэтому его предположение, что человек живет изначально скорее в опыте других, «присваивая» его в процессе социализации, чем в своем собственном, справедливо. А если так, то и традиционное отождествление внутреннего опыта и собственного опыта ошибочно.

М. Шелер выдвигает перцептивную теорию Alter Ego, артикулирующую роль тела в процессе восприятия в опыте чужого мышления. В соответствии с его представлениями, именно телесные чувства конституируют различия между Я и не-Я (но в ином, чем у Гуссерля, смысле). Но до тех пор, пока человек живет в своих собственных чувствах, он не может найти подхода к жизни Другого. И лишь когда он преступает собственное Эго и возвышается до уровня личности (Person), он действительно обретает доступ к опыту Другого – безо всякой эмпатии и вывода по аналогии.

Связь шелеровского понятия интересубъективности с его антропологией и этикой очевидна: до тех пор, пока человек лишь в собственных телесных чувствах, доступ к жизни другого для него закрыт. Кроме того, М. Шелер подчеркивает, что само восприятие Другого является

очень сложным: мы воспринимаем не чужую телесность, его Self или Ego, а тотальность (феноменологическую целостность), которую невозможно разделить на объекты внутреннего или внешнего опыта.

К сожалению, одно из наиболее глубоких теоретических прозрений М. Шелера – представление о теоретических уровнях проблемы intersубъективности – осталось нереализованным в рамках его теории Alter Ego. Кроме того, аргументы М. Шелера основаны на эмпирических фактах, почерпнутых из детской психологии и культурной антропологии, а не на анализе проблемы в пределах трансцендентальной сферы. Иными словами, М. Шелеру, как и Э. Гуссерлю, не удалось найти удовлетворительного решения проблемы intersубъективности, отвечающего трансцендентальному идеалу всеобщности и необходимости.

Подытоживая эти попытки, можно и вовсе усомниться в том, что проблема intersубъективности в принципе разрешима в пределах трансцендентально-феноменологической сферы. В самом деле, все Пятое «Картезианское Размышление» пронизано напряженным противоречием между требованием конституировать Другого с помощью интенциональных ресурсов собственного сознания и требованием конституировать его как Другого. Поэтому А. Шюц переносит решение проблемы intersубъективности, а следовательно и природы социальности, в сферу жизненного мира человека (*Lebenswelt* – понятие, заимствованное из поздних работ Э. Гуссерля), стремясь обрести не всеобщее, трансцендентально-феноменологическое, но социологически релевантное решение проблемы<sup>7</sup>. Вопрос же о том, разрешима ли проблема intersубъективности как трансцендентально-феноменологическая, остается открытым.

Будучи перенесена из сферы чистого сознания трансцендентального Эго в область жизненного мира человека, проблема intersубъективности не только сохраняет свой социально-философский статус, но и приверженность феноменологической традиции. И если рассматривать проблему intersубъективности в пределах жизненного мира, то первый вопрос, на который следует дать ответ, состоит в том, насколько оправдано предположение М. Шелера, что сфера Мы предшествует сфере Я. А. Шюц убежден в том, что сознание в естественной установке не проблематизирует бытие других, воспринимая его с той же очевидностью, что и существование физических вещей. И только радикальные солипсисты полагают его «слабой данностью». Но и они, обсуждая этот вопрос на конгрессах, не сомневаются в том, что апеллируют друг к другу как к «данности» вполне реальной. Таким образом, убежден А. Шюц, мы можем согласиться с утверждением М. Шелера, что сфера «Мы» – в этом специфическом смысле – действительно первична. Более того, в рамках естественной установки мой собствен-

---

<sup>7</sup> *Schutz A. Collected Papers. The Hague, 1962. Vol. I, P.P.167-172*



ный поток сознания может казаться анонимным. В дорефлексивном мышлении, каким является сознание в естественной установке, нет представления о том, что **Я** мыслю, но есть изначальное представление о потоке сознания как таковом. Мое собственное Я (self) заслонено от нас интенциональными объектами нашего мышления. И лишь в результате рефлексии впервые появляется представление о собственном Я. Подвергая критике шелеровскую концепцию интерсубъективности, А. Шюц отмечает, что ее автор не проводит различий между мышлением в естественной установке и рефлексивным мышлением. Но если мы примем во внимание различие между обеими установками, то против теории М. Шелера могут быть выдвинуты следующие возражения:

1. Утверждение о том, что мы изначально живем скорее в сознании других, чем в своем собственном, справедливо лишь для естественной установки сознания.

2. Не существует такого опыта, о котором нельзя было бы сказать, к чьему потоку сознания он принадлежит. Я могу сомневаться, является ли мысль, пришедшая мне в голову, оригинальной или нет, но у меня не может быть сомнения в том, что сама мысль, равно как и мое сомнения в ее оригинальности, принадлежат моему потоку сознания.

3. Наконец, данные современной детской психологии и культурной антропологии, свидетельствующие, что как ребенок, так и человек дописьменной культуры далеко не сразу осознают свое отличие от других, свидетельствует лишь о том, что усвоение техники рефлексии – процесс длительный, как фило- так и онтогенетически<sup>8</sup>.

А. Шюц обращает внимание и то обстоятельство, что темпоральная структура естественной («наивной») и рефлексивной установок различна. В естественной установке сознания мы живем настоящим и ориентированы на ближайшее будущее, которое предвосхищаем в наших ожиданиях ближайшего будущего (антиципациях). В трансцендентально-редуцированной сфере Э. Гуссерль именует их «про-тенциями», применительно же к естественной установке сознания А. Шюц использует термин «живое настоящее». Но согласно феноменологической теории рефлексии, она в состоянии схватить лишь прошлый, истекший, прожитый опыт, но никак не опыт длящегося мгновения. Наше самосознание конституировано прошлым опытом. Но – и это отличие для А. Шюца принципиально – в ситуации «лицом-к-лицу» поведение другого и симптомы его психической жизни даны нам в «живом настоящем», как «живая одновременность». Именно это различие и лежит в основе определения Alter Ego: Alter Ego – это такой субъективный поток сознания, который я улавливаю в «живом настоящем». Опыт восприятия чужого потока сознания в «живой одновременности» он называет «всеобщим тезисом существования

---

<sup>8</sup> Schutz A. Collected Papers. The Hague, 1962. Vol. I, P.P.170-171

Alter Ego»<sup>9</sup>, а общее нам настоящее – «чистой Мы-сферой» (the pure sphere of the We). И если принять подобное определение, то можно вполне согласиться с мнением М. Шелера о том, что сфера Мы действительно первична по отношению к Я, – но в весьма специфическом смысле, который сам М. Шелер никогда не имел в виду. И все же мне трудно всецело разделить критические замечания А. Шюца в адрес М. Шелера, связанные с тем, что тот не проводит четкого разделения на наивную и рефлексивную установки. Ведь само разделение на наивную и рефлексивную установки принадлежит не естественной («наивной») установке сознания, к которой апеллирует А. Шюц в попытках найти социологическое решение проблемы интересубъективности, но рефлексивной установке как таковой. Кроме того, даже и в естественной установке сознания мы не можем разделить наше бытие на «просто жизнь» и «мышление». Они одновременны, взаимозависимы и не разложимы на упомянутые составляющие даже и в рефлексивном повороте. И вопрос о том, в какой мере «всеобщий тезис существования Alter Ego» может служить достаточным решением проблем интересубъективности, и следовательно, теоретическим фундаментом социальных наук, ждет дальнейших исследователей.

Несомненно плодотворной попыткой теоретического прорыва в философском осмыслении проблемы интересубъективности является предпринятый Ю. Хабермасом синтез «понимающего» и структурно-функционального подхода на основе теоретических наработок лингвистической философии<sup>10</sup>. В методологическом плане теория коммуникативной рациональности Ю. Хабермаса знаменует отход от чистой философии сознания в пользу коммуникативной теории с соответствующим такому подходу переосмыслением понятия интересубъективности. Коммуникативно-рациональная интерпретация интересубъективности полагает субъекта социального действия не только в соотношении с объектом в фундаментальных модусах познания и действия, но и как достигающего договоренности («рационального консенсуса») с другими субъектами познания и действия на основе языковой и коммуникативной компетенции. Основываясь на опыте изучения «идолов языка», интерес к которым существенно возрос в результате «лингвистического поворота» англо-саксонской философии, автор «Теории коммуникативного действия» подвергает критическому осмыслению условия возможности интересубъективной договоренности, используя методы языковой прагматики, в частности аппарат семиотического анализа, для обоснования коммуникативной этики ответственного дискурса. Конститутивным для нее оказывается

---

<sup>9</sup> Schutz A. Collected Papers. The Hague, 1962. Vol. I, P. 174

<sup>10</sup> Хабермас Ю. Модерн - незавершенный проект // Вопросы философии. 1992. N 4. Habermas J. Theorie des kommunikativen Handelns. Bd.1. F/M., 1981.

положение о том, что при безусловном фактическом равенстве договаривающихся сторон коммуникативного действия рациональный консенсус достигается на основе «силы лучшего аргумента». Нетрудно видеть, что подобная констатация придает формальной прагматике Ю. Хабермаса нормативный статус идеального диалога в лишенном диспозиций власти и ситуативных детерминаций социальном пространстве. Но в однополярном глобализирующемся мире аргумент силы, увы, куда сильнее силы аргументов. В собственно когнитивном же отношении девальвация Ю. Хабермасом истины до нередуцируемого многообразия контекстуальных intersубъективных значений, разделяемых членами локальных коммуникативных сообществ, или рационального консенсуса на основе intersубъективных схем толкования, присущих различным жизненным мирам, равно как и низведение «эпистемической очевидности» до уровня решающего хода в языковой игре, представляет собой несомненную уступку Ю. Хабермаса постмодернизму.

В постмодернистской же перспективе социальной философии, равно как и в социальных практиках, ее порождающих, понятие intersубъективности проблематично не столько с точки зрения его теоретической обоснованности, сколько адекватности наличному состоянию социальных связей. Ибо постмодернизм провозглашает не только «смерть субъекта», но и фактическую *деградацию социальных коммуникаций*. «Общество индивидов» (Н. Элиас), «индивидуализированное общество» (З. Бауман), «восстание масс» (Х. Ортега-и-Гассет), наконец, «уходящая в бесконечность симуляция всех систем референции» (Ж. Бодрийяр) – таков далеко не полный диагноз духовных ситуаций постмодерна. Общий «кризис легитимации», виртуализация социальных дифференциаций («класс на бумаге» – П. Бурдьё) и усиление социальной фрагментации диагностируются как «смерть социального», ее превращение в «знак неподлинного» – симулякр (Ж. Бодрийяр). Ощущение деструкции «божественного социального» (Э. Дюркгейм) инспирирует попытки, с одной стороны, «обрести твердую почву в повседневности, когда шоссе грядущего развития теряется в тумане»<sup>11</sup>, с другой – в обращении к реликтам прежней социальности, затерянной в веках. Сегодня мы наблюдаем, как в обличье современных технологических средств возрождаются характеристики глубоко архаичных структур традиционалистского сознания: изначальная слитность публичного и приватного, размывание заповедных зон публичной презентации и артикуляции в культуре, эстетизация примитива. «Средние века уже начались», – констатирует У. Эко<sup>12</sup>, ссылаясь и на возрождение корпоративного духа среди имущих, и на усиление социально-

<sup>11</sup> см. Касавин И. Т., Щавелев С.П. Анализ повседневности. М., 2004

<sup>12</sup> Эко У. Средние века уже начались // Иностранная литература. 1994. №4.

культурного локализма, и на потерю контроля над большинством социальных процессов, ранее казавшихся вполне «прозрачными». В унисон его прозрениям звучат откровения одного из наиболее глубоких критиков современного общества, английского философа З. Баумана, обеспокоенного «бессилием эклессии (сферы публичного – Н.С.) и запустением агоры» (места встречи публичного и приватного)<sup>13</sup>. Постмодернизм заявляет о себе невиданным в Новейшей истории взлетом культурного авторитета архаичных когнитивных практик, укорененных в формах мышления и деятельности доиндустриальных, а подчас и дописьменных обществ (синдром неотрайбализма, неокочевничества, неоязычества и т. д.). Подобного рода практики (совещаться с астрологом, гадать на картах, кофейной гуще и т. д.), ныне широко востребованные в верхних слоях российской политической и культурной элиты, уже нельзя назвать маргинальными. Ибо сегодня подобные маргиналы восседают в креслах советников публичных политиков и активно участвуют в принятии решений, значимых для всего общества. В терминах социальной феноменологии это означает, что налицо универсализация и легитимация социальных значений крайне узкого, а то и вовсе маргинального слоя современного российского общества. И первое, что надлежит сделать в интересах налаживания социальных коммуникаций – незримых нитей ткани социальной жизни, – изучить весь спектр социальных значений, бытийствующих в современном обществе. И социальная феноменология располагает для этого адекватными понятийными и методологическими ресурсами. *Ибо социальная феноменология видит в смысловой структуре социального мира всеобщую матрицу непрерывного воспроизводства социальности, которая, как свидетельствуют ранее приведенные высказывания, сегодня «деградирует».* Поэтому в ситуации постмодернистской деградации социальных коммуникаций феноменологический анализ предельных оснований социальности, поиск фундаментальных жизнемировых констант *интерсубъективности* может стать достойным философским ответом на ситуацию «постмодернистского «землетрясения культуры».

В свою очередь, изучение параметров интересубъективности приближит нас и к решению проблемы конституирования смысла Другого в виртуальном общении, в частности, задав новый, феноменологический, вектор трактовки теста Тьюринга.

---

<sup>13</sup> Бауман З. Индивидуализированное общество. М., 2002.

---

## УЛУЧШЕНИЕ КАЧЕСТВА МЫШЛЕНИЯ: «СОКРАТИЧЕСКИЙ ДИАЛОГ» И ИНТЕРАКЦИЯ С КОМПЬЮТЕРОМ\*

*Н.С. Юлина*

Начиная с 60-70-х гг. прошлого столетия во всем мире лавиной нарастает публикация работ, посвященных осмыслению фактов вторжения в нашу жизнь компьютеров и, конечно, сравнению искусственного и естественного интеллекта. Среди них относительно мало работ, в которых это сравнение проводится с точки зрения педагогики. Точнее, в которых сопоставляется коммуникация с естественным интеллектом и коммуникация с искусственным интеллектом (компьютером) с целью определения эффективности этих двух типов общения в наработке навыков хорошего мышления. Например, в сборнике «Влияние Интернета на сознание и структуру знания» (2004 г.), изданном в Институте философии РАН, авторы обсуждают очень тонкие формы воздействия электронных средств связи на «сознание» (по их мысли революционного), но практически не касаются педагогических, образовательных следствий этого воздействия.

В данной статье мы предприняли попытку провести такое сравнение. Ее цель — педагогически-образовательная. В то же время, как нам представляется, сравнение в какой-то мере позволяет посмотреть на соотношение искусственного и естественного интеллекта под очень важным углом зрения, а именно, эпистемологии человеческого опыта.

Объектами нашего сравнения является не вообще «человек-компьютер» и «человек-человек», а общение с обучающими программами компьютера и общение в форме философского «сократического диалога». Популярное в настоящее время понятие «сократический диалог» толкуется по-разному. Мы будем иметь в виду то его содержание, которое оно имеет в программе «Философия для детей» — как педагогическая форма обучения разумному мышлению учащихся в процессе коллективного (дискуссионного) поиска истины. Чтобы пояснить это содержание, имеет смысл предварить его кратким вступлением, в котором говорится о поиске новой парадигмы образования и основных принципах программы «Философия для детей».

---

\* Работа подготовлена при финансовой поддержке Российского гуманитарного научного фонда (РГНФ), проект № 04-03-00123а

## 1. «Образование для будущего» и программа «Философия для детей»

Одновременно с вторжением компьютеров в нашу жизнь, то есть, с 70-х гг. прошлого столетия, во всем мире активно заговорили о необходимости внесения серьезных корректив в традиционную (информативную) систему образования с тем, чтобы сместить акцент с *количества* усвоения учащимися знания на улучшение *качества* их мышления. Многие авторы (например, Алвин Тоффлер) предрекали кризис цивилизации в случае, если традиционный информационный крен образования (как можно большее усвоение знания) не будет сбалансирован рефлексивным («раскруткой мозгов»). Исправление такого крена он включил в первейшие задачи «стратегии выживания»<sup>1</sup>. Были выдвинуты различные проекты «Образования для будущего». Они имели разные названия («рефлексивное», «исследовательское», «критическое» и др.), но их объединяла общая идея — найти эффективные педагогические средства для повышения качества мышления. Последнее означало приблизительно следующее: *сделать мышление детей самокритичным, творческим, гибким, оснащенным разными методологическими средствами для эффективного применения знания на практике и адаптации к меняющейся среде*. Иначе говоря, вооружить их разнообразными *навыками*, позволяющими *самостоятельно* решать неординарные проблемы, которые может подкинуть новая эпоха.

Утверждение новой образовательной парадигмы — это продолжающийся десятилетиями болезненный процесс апробации различных подходов и стратегий. В рамках дискуссий об «Образования для будущего» возник вопрос, какие *дисциплинарные рычаги* оптимальны для «раскрутки мозгов» и улучшения качества мышления. Вопрос этот, конечно, не новый. Вспомним, что в дореволюционной российской школе основная ставка делалась на лингвистические предметы, в особенности на изучение классических языков — древнегреческого и латыни. Считалось, что они дисциплинируют ум, делая его податливее для усвоения другого знания. Вскоре последовало разочарование, и внимание педагогов обратилось на потенциал математики, исходя из того, что упор на развитие математических навыков убивает сразу двух зайцев: дает знание по этой важной дисциплине и развивает рассуждающие, логические способности интеллекта. Затем энтузиазм педагогов обратился на возможности логики, и она стала интенсивно внедряться в школьные программы. Все эти дисциплины имели свои позитивные эффекты, но существенно не повысили качество мышления и рассуждающей способности интеллекта, что считается разумностью и практической мудростью. И в этом отношении, не смогли нивелиро-

---

<sup>1</sup> *Toffler A.* Future Shock. N.Y.1970, p. 4.

вать негативные стороны того, что сегодня вылилось в «кризис образования». Задача «раскрутки мозгов» — научить детей навыкам разумного рассуждения — основе всех других приобретаемых в образовании навыков, — по-прежнему остается актуальной.

Новый этап в дискуссиях о дисциплинарных рычагах, способных улучшить качество мышления, ознаменован появлением в культуре и образовании невиданного феномена — компьютеров и компьютерных обучающих программ (или дистанционного обучения). Этот феномен дал новую пищу для дискуссий об образовании для будущего. Очень многие педагоги обратили свой взор на эту новацию, увидев в ней долгожданное средство повышения и количества информации, и качества интеллекта. А некоторые увидели в ней лекарство для спасения образования от кризиса.

Конечно, всегда находились люди, предлагавшие передать роль «раскрутки мозгов» философии и ввести преподавание этой дисциплины в школу. Особенности философии — вечно проблемный характер ее сюжетов, допущение поливариантных ответов, задействованность в ней когнитивных, этических и эстетических способностей сознания, делает ее инструменты уникальными для выработки недогматичного, гибкого и одновременно доказательного мышления. Однако для реализации на практике этих предложений существовало много препятствий: представление о философии как «взрослой» дисциплине, которую в лучшем случае можно вводить только в старшие классы (господство установки Ж.Пиаже об обретении способности к абстрактному мышлению только к 12-14 годам), но главное — отсутствие теоретически разработанных и практически работающих программ для всего периода школьного обучения. Препятствием было и справедливое опасение родителей (и деятелей образования), что под видом изучения «философии» дети будут подвергаться *индоктринации*, то есть им будет навязываться та или иная мировоззренческая или идеологическая доктрина и они могут стать объектами идеологической манипуляции со стороны составителей философских программ.

Почва для опасений уменьшилась с появлением в 70-е годы программы «Философия для детей» (Создана в Институте по развитию философии для детей, университет Монтклер, США). Важно подчеркнуть следующее: в программе «Философия для детей» философия используется не как *доктрина*, навязывающая детям какое-либо мировоззрение или идеологию. Ее цель — использовать философию как *инструмент* для обучения будущих граждан навыкам или искусству разумного самостоятельного мышления и ответственного социального и морального поведения. Программа получила самые высокие оценки комиссий по образованию, ООН, Европарламента, Министров по образованию США и других стран. На сегодня в 45 странах мира функционируют центры, ведущие работу по этой Про-

грамме<sup>2</sup>. Конечно, ее распространение в разных культурах было бы невозможно без впечатляющих практических результатов; например, в классах, прошедших подготовку по этой Программе успехи в математике на 33%, а по языковым предметам на 66% выше, чем в классах без философии. Не менее важный показатель — повышение уровня моральной и социальной ответственности учащихся, который невозможно оценить в процентах<sup>3</sup>. Авторитет, который эта программа завоевала в мире, основывается на том, что разработчики Программы правильно угадали доминирующий вектор развития образования в мире — поворот от информационной к «рефлексивной» (или «исследовательской») парадигме образования, — и сами внесли в нее весомый вклад.

## 2. «Сократический диалог» в сообществе исследователей

Каковы же главные педагогические новации программы «Философия для детей»? Ключевым понятием ее педагогической стратегии является *«сообщество исследователей»* (community of inquiry). Сообществом является обычный школьный класс, коллективно ищущий истину в форме исследовательского «сократического диалога». Пусковым механизмом такого диалога является контекстуально представленная проблема и необходимость ее разрешения<sup>4</sup>. Этапы диалога – анализ контекста ситуации, формулировка релевантных к нему философских проблем, конструирование гипотез, обсуждение критериев правомерности выдвигаемых мнений, выработка итогового «хорошего суждения».. То есть, вместо традиционной модели, которая используется во многих школах Западной Европы (сообщения о персоналиях, их высказываниях, «измах» и дисциплинах философии), в Программе используется *исследовательская* модель или проблемно-

---

<sup>2</sup> Особенно успешно она пошла в странах Латинской Америки, Австралии, Испании, Тайване, Болгарии.

<sup>3</sup> Более подробно об этой программе см.: Юлина Н.С. Философия для детей: улучшение качества разумного мышления. М., 2005.

<sup>4</sup> Согласно М.Липману, основному разработчику программы «Философия для детей», для того, чтобы улучшить качество образовательного процесса, следует признать его парадигмальной характеристикой *исследование*, а школьные классы превратить в «сообщество исследователей». «Чтобы стать исследованием, в обучении должно присутствовать какое-то сомнение, что с ситуацией все в порядке, некоторое осознание, что в ней имеются вызывающие тревогу трудности и она в каком-то отношении проблематична. Обучение должно быть самокорректирующим исследованием, принимающим во внимание все соображения, и конструирующим альтернативные гипотезы о возможных путях разрешения этой проблемы. Помимо всего прочего, исследование включает вопрошание, более узко — поиск истины, а более широко — поиск смысла» (Lipman M. Thinking and Education. Cambridge MA, 2003, p. 94-95)



деятельностный метод трансляции знания. Ставится задача обучить **философствованию** (а не информации о философии), и в процессе философствования (с использованием наработанного в истории мысли инструментария), учить навыкам хорошего, грамотного, разумного рассуждения. Или тому, что главный разработчик М.Липман назвал «мышлением высшего порядка». Мышление высшего порядка имеет место не в результате одноактного действия – интуиции или озарения, а в процессе трудоемкой творческой (смыслотворящей) и критической (самокорректирующей) деятельности разума индивидов в коммуникации с другими индивидами<sup>5</sup>. По идее разработчиков школьное «сообщество исследователей», в котором акцент сделан на вербальной дискуссии, является оптимальным местом наработки не только когнитивных и исследовательских, но и этико-социальных навыков. Такой школьный класс может стать главной социальной ячейкой общества, в которой происходит становление субъектов демократии и осваиваются принципы гражданского общества.

Обучение в форме исследовательского «сократического диалога», конечно, предполагает программное и текстовое обеспечение. Обычные учебники и хрестоматии здесь заменены особым рода текстами – нагруженными философскими проблемами художественными повестями для детей, в которых главными персонажами выступают школьники, и сюжет строится на обычных (не сказочных), близких жизненному опыту учащихся, школьно-семейных событиях. Последние – затравка для философского разговора в классе и подсказка возможных его поворотов.

В отличие от европейских программ, где философия преподается только в старших классах и в виде текстов по истории философии, «Философия для детей» рассчитана на весь период обучения в школе (с адаптацией содержания философии к возрастным особенностям детей). Разработчики этой программы не устают повторять две вещи: Во-первых, что философский тренинг (как и математический или языковой) эффективен тогда, когда он осуществляется не спорадически, а в течение всего периода школьного обучения; во-вторых, приобретаемые с помощью философии навыки разумного мышления и рассуждения являются **базовыми навыками**. Когда они наработаны, на их основе легче приобретаются все остальные навыки, которым обучают в школе — навыки чтения, литературного изложения, математические, экспериментальные и другие.

---

<sup>5</sup> «Мышление также есть форма практики, однако ее еще нельзя считать хорошей практикой, пока она не включит в себя самокорректирующую компоненту и тем самым станет формой исследования. Более того, я исхожу из того, что хорошее мышление – главная цель образования» (*Lipman M. Promoting better Classroom Thinking // Educational Psychology, 1993. V. 13, № 3-4, P. 291-292*).

Обычно педагогов-практиков больше всего интересует вопрос, – каков «сухой остаток» той или иной предлагаемой педагогики? В общей форме можно сказать, что главный планируемый результат педагогики «сообщества исследователей» «Философия для детей» – обретение учащимися навыков (умений, искусств, способностей) здравого и разумного мышления, как необходимого условия здравого, разумного и ответственного социального и морального поведения. (Понятно, что измерить разумность и здравость теми же мерками, какими измеряется обычное школьное знание, вряд ли возможно). Условно эти навыки можно подразделить на следующие:

**Лингвистические навыки** (в адекватных понятиях выражать чувства и мысли; различать слова и значения; прояснять оттенки смыслов понятия и видеть их контекстуальные границы, и др.). **Исследовательские навыки** (распознавать проблемную ситуацию и формулировать относящиеся к ней вопросы; фиксировать противоречия в имеющейся информации; использовать элементарные приемы фактуального подтверждения; видеть зависимость результатов от используемых когнитивных средств и др.). **Когнитивные навыки** (проводить различия и устанавливать связи; видеть тождественное содержание в разных словесных выражениях; анализировать скрытые допущения; понимать значимость критериев и приводить основания в их пользу; приводить примеры и контрпримеры; различать сильные и слабые аргументы; подводить факты под правила и законы; избегать обобщений на основе единичных фактов и видеть особенности широких абстракций и др.). **Этические навыки** (умение слушать других и принимать во внимание мнение другого; работать в коллективе и рассматривать свое Я со стороны; проявлять самостоятельность и брать на себя ответственность; проявлять терпение и терпимость к критике; уважать коллектив и заботиться о нем и др.). **Психологически-личностные навыки** (уметь рефлексивно идентифицировать свое Я и отстаивать свою индивидуальность; уметь самоутверждаться и самооценивать себя в дискуссиях со сверстниками, и др.). **Социальные (гражданские, коллективистские) навыки** (уметь различать моральные, правовые, социальные проблемы и грамотно рассуждать о них; владеть навыками разумного разрешения межличностных конфликтов, конфликтов «индивид-коллектив», этнических и религиозных конфликтов; занимать активистскую позицию и брать на себя моральную и социальную ответственность, и др.).

#### 4. «Компьютерное обучение и хорошее мышление

Как мы уже сказали в начале статьи, свое веское слово об «Образовании для будущего» сказали компьютерщики. Мощный рывок в развитии компьютерной техники в 80-х годах сделал возможным дистанционное обучение с использованием специальных обучающих программ. Его разработчики и энтузиасты утверждают, что педагогическая новация в виде обучающих компьютерных программ (или в

форме развивающих игр) является самым эффективным средством тренировки и совершенствования мыслительных навыков учащихся и способности принимать самостоятельное решение. Наиболее оптимистично настроенные из них отвергают бытующее мнение об ущербности коммуникации с персональным компьютером; в ее рамках, считают они, возможен не только обмен информацией, но взаимопонимающий разговор. А при создании более совершенных программ он может стать подлинным «сократическим диалогом». Он будет играть такую же обучающую роль, какую играет коммуникация людей, занимающихся поиском истины. Некоторые из оптимистов пошли дальше, заявляя, что использование в образовании этой машинной новации позволит многое понять в природе человеческой коммуникации. В 1980 году Сеймур Паперт, один из пионеров AI и применения искусственного интеллекта в образовании, опубликовал книгу «Mindstorms», в которой выдвинул необычную для того времени идею о компьютеризации школьных классов. (Он создал первую в США обучающую программу для 5-летних детей, которая строилась на сказках и загадках)<sup>6</sup>. Компьютерные и педагогические идеи Паперта вызвали шквал возражений. В 1993 году он выпустил книгу «Машина детей: переосмысляя школу в век компьютера», в которой, подводя итоги практике компьютерного обучения в США, сделал вывод, что основным препятствием на его пути по-прежнему является косность мышления деятелей образования и консерватизм образовательной системы. Продолжая оставаться поклонником компьютерного обучения, он убежден, чтобы бы его противники не говорили об упрощенном «механическом мышлении компьютеров», компьютер помог более трезво посмотреть и на мышление, и на то, **как следует обучать хорошему мышлению**<sup>7</sup>.

Несмотря на популярность обучающих программ, число скептиков не уменьшается. Многие расценивают образовательный процесс с применением компьютеров как псевдообучение, а некоторые — как бедствие. Не отрицая важную инструментальную и информативную роль компьютера, они утверждают, что в коммуникации с компьютером не задействованы огромные и разнообразные пласты человеческого опыта, без которых невозможен реальный взаимопонимающий диалог — необходимое условие повышения качества мышления учащихся и их социализации. Однако возражения скептиков уже бессильны остановить процесс: с каждым новым годом компьютер (и специальные обучающие программы) занимают все более прочное место в школьном образовании. Во всяком случае, сегодняшние дети

---

<sup>6</sup> Papert S. Mindstorms. N.Y., 1980.

<sup>7</sup> Papert S. The Children's Machine: Rethinking Schools in the Age of Computer. N.Y., 1993.

нередко тратят больше личного времени в общении с компьютером (играя в развлекательные и развивающие игры, или используя школьные обучающие программы), нежели в общении со сверстниками или взрослыми, а тем более в «сократических беседах» в философских классах (даже там, где они есть).

Данный факт нашей жизни на фоне заявлений энтузиастов компьютерного обучения является, конечно, вызовом педагогике «сообщества исследователей» программы «Философия для детей». И не только ей. Ставится под сомнение вообще идея улучшения качества мышления с помощью традиционного философского «сократического» (дискуссионного) обучения. Фактически брошен вызов важнейшей функции философии, о которой говорил еще Сократ, — ее способности учить *мудрости* в процессе критического дискурса.

Вызов, конечно, требует обстоятельного и развернутого ответа. В ограниченных пределах данной статьи мы не претендуем формулировать его. Ее узкая задача состоит в сравнении двух типов коммуникации: коммуникации в форме «сократического диалога» (в той интерпретации, какую он имеет в программе «Философия для детей»), и коммуникации с персональным компьютером с точки зрения их возможностей повышения качества мышления и социализации. У этой задачи есть и философско-эпистемологический аспект: на основе сравнения высвечиваются грани и особенности живого человеческого опыта: абстрактный тезис о коммуникативном, непрерывном, со многими взаимодействующими элементами характере опыта приобретает более осязаемые черты.

Свое сравнение я хотела бы начать с оговорки, что я вовсе не предлагаю выбросить компьютеры из класса и не собираюсь ставить под сомнение важность обучающих компьютерных программ. Сегодня создано множество таких программ (по чтению, грамматике, иностранному языку, математике, истории и др.). Как правило, программы содержат качественный информационный материал и хорошо проработанные методики, учитывающие возраст обучающихся, уровень подготовки и даже их личные особенности. У таких программ, несомненно, есть свои *достоинства*, которые нельзя отбрасывать.

Достоинством информационных технологий является *введение в мыслительные операции новых семиотических средств*. Известно, что изобретение технологии письменной речи радикально расширило диапазон навыков мышления и воображения и существенно увеличило свободу оперирования абстракциями. Надо полагать, что семиотические средства компьютерных информационных технологий в еще большей мере расширят диапазон средств мышления и воображения, делая их более гибкими и пластичными (оперирование с виртуальными объектами усиливают способности обобщения и полет фантазии). Хотела бы оговорить, что когда речь идет о влиянии письменного

сти или семиотических средств новых информационных технологий, сказанное относится прежде всего к обогащению *навыков мышления*, то есть набора языковых (и семиотических) привычек интеллектуальной деятельности, а не к «сознанию» как таковому или «уму». Сознание, с нашей точки зрения, является продуктом миллионов лет генно-культурной эволюции и на него вряд ли могут повлиять какие-либо новые технологии.

Сегодня вряд ли кто отрицает, что с помощью компьютера приобретаются и тренируются очень полезные *когнитивные и исследовательские навыки мышления*. Навыки планировать, связывать поставленные задачи с имеющимися средствами (инструкциями выполнения задачи); строить прогнозы возможных результатов; развивать способность к самокоррекции, получая помощь в виде указаний на ошибки (логические и иные); грамотно анализировать проблемы, разделяя их на подпроблемы; осваивать алгоритмические процедуры; прорабатывать детали, и др. Особенно эффективно нарабатываются навыки точного использования языка, что сказывается на общей грамотности, чуткости к особенностям синтаксиса и семантики. Все они *дисциплинирует* мышление, учат точности и, безусловно, развивают интеллектуальные возможности. Но при этом следует помнить, что все они относятся к категории *когнитивных и исследовательских* навыков, куда не входят навыки оперирования с феноменальными, эмоциональными и ценностными, то есть качественными, сторонами опыта.

Несомненным плюсом в работы с компьютером является *сохранение времени* мыслительных операций за счет быстрого поиска информации, осуществления логических операций, установления обратной связи и др. Одним словом, реализуется выдвинутый Папертом лозунг — «save your time!» — «берегите ваше время».

Компьютер не только хороший *тренажер когнитивных навыков*. В каком то смысле он может играть роль собеседника в разговоре — отвечать на вопросы, не соглашаться, подсказывать новые ходы рассуждений и т.д. Для некоторых людей в определенных ситуациях это единственно доступная для них форма разговора. Все чаще встречаются дети, для которых общение с компьютером (и путешествие по сетям Интернет) стало дороже общения с друзьями и родными. (Не случайно его назвали «персональным компьютером»).

Распространено мнение, что компьютер не годится на роль собеседника по причине отсутствия у него эмоций. Это не совсем так. Хотя машина неэмоциональна, у общающегося с ней человека вполне может быть задействована *эмоциональная компонента*. Многие дети более интенсивно мобилизуют способности воображения, памяти и эмоций в компьютерных играх, нежели в играх со сверстниками или при чтении книг. Играющие с компьютером шахматисты не просто обмениваются с ним информацией и соревнуются в умении делать оптимальные хо-

ды, но и получают удовольствие от игры. Ссылаясь на эти факты, энтузиасты компьютерного обучения считают приемлемым тезис, что в случае общения человека с компьютером имеет место не просто информационные и когнитивные подвижки, а настоящий эмоционально насыщенный понимающий «сократический диалог».

Огромным достоинством компьютера, с нашей точки зрения, является присущий только ему эффективный *способ трансляции знания*. Известно, что вопрос о трансляторах знания, один из важнейших в обучении. Это понимал еще Платон, когда он, не отрицая достоинств «сократической беседы», считал, что как передатчик знания другим людям или другим поколениям устная беседа проигрывает в сравнении с текстом. (Одна из причин, почему в европейской традиции утвердился «платоновский», а не «сократический» метод обучения кроется именно в том, что письменный текст оказался более удобным и надежным транслятором). Сегодня тоже мало кто оспаривает достоинства «сократической беседы» в раскрутке мозгов, вместе с тем даже скептики признают, что как универсальный тренажер и не имеющий граничи передатчик знания компьютер обладает большими мощностями.

В заключение позитивных характеристик компьютера следует сказать еще об одном важном достоинстве — *демократизме*. Дистанционное обучение дает возможность учащимся в любой географической точке получать качественное знание, что, безусловно, демократизирует образование. Что особенно важно для России с ее огромной протяженностью, нехваткой учителей и невозможностью обеспечить должное обучение в отдаленных регионах. Живущий в какой-нибудь глуши ребенок, лишенный возможности учиться в хорошей школе, в принципе может получить квалифицированную учебную помощь, обратившись к обучающим программам. Компьютер дает ему возможность восполнить дефицит библиотек безграничным информационным полем Интернета. Тем самым в какой-то мере компьютер помогает преодолеть *информационное неравенство* детей, живущих в центре и в провинции, или в семьях с низким, и в семьях с высоким уровнем культуры.

И еще одно заключение скорее общефилософского, нежели педагогического порядка. Между типом общения с персональным компьютером и типом общения в сообществе исследователей, ведущих философский (исследовательский) «сократический разговор» имеются сходные черты. Они обусловлены тем, что и в программировании компьютера, и в философской рефлексии исходным пунктом является *опыт мышления о мышлении*. И там, и здесь приведены в действие *когнитивные* и *исследовательские* операции. И там, и здесь действуют механизмы *самокоррекции*. Это очень важные параметры сходства. Тем не менее, несмотря на ряд сходств, различия этих двух типов коммуникации более существенны. Мы постараемся показать, прибегнув к анализу сократического диалога в классе.

#### 4. Особенности коммуникации в философском «сократическом разговоре»

Главное отличие состоит в том, что живой диалог в сообществе исследователей *многомерен*. Опыт сознательного акта индивида в процессе человеческого общения представляет собой сложно запутанный клубок, в котором перекрещивается такое количество постоянно меняющихся векторов, которые невозможно разложить на детали, проанализировать и затем воспроизвести в компьютерной программе. Здесь действуют и биологические диспозиции, сформировавшиеся в результате миллионов лет эволюции, и социальные диспозиции, сложившиеся в результате языково-культурной истории, и реакция на изменчивую ткань контекста, и уникальность индивидуального опыта и памяти, и еще многое другое. Очень важно иметь в виду, что в сознательных актах естественного интеллекта, — а они не могут реализоваться без коммуникации с другими сознаниями и диалога, — задействованы не только когнитивные, исследовательские и самокорректирующие способности, но и феноменальные, то есть реакции на качественную окрашенность опыта и весь фоновый контекст. Можно предположить существование других, сформировавшихся в эволюции слоев, которые не проходят через когнитивную составляющую сознания, но принимают участие в актах рефлексивной дискуссии. Многомерность и контекстуальность человеческого сознательного опыта — та трудность, на которую натываются все философы, занимающиеся природой сознания. И самая большая трудность связана не с объяснением когнитивной составляющей, относительно легкой, а с чувственной стороной. (Дебаты по поводу «архитектуры» сознательного опыта до сих пор являются одними из наиболее острых в философии сознания, расколов ее на два лагеря — на «когнитивистов», отрицающих существование феноменальных «квалиа», и «феноменалистов», считающих феноменальную окрашенность опыта атрибутом когнитивного). Именно сторонники компьютерного подхода пытаются объяснить сознание по аналогии с искусственным интеллектом, ограничивая сознание только его когнитивными (выражаемыми в языке) аспектами за счет элиминации феноменального. (Например, Д. Деннет).

Многомерность живого общения («сократического диалога»), вероятно, является важнейшим условием смыслотворчества и разумного мышления. «Разумность, пишет М.Липман, включает в себя культивирование многомерного мышления»<sup>8</sup>. Занимаясь разрешением проблемной ситуации, человеческий ум временами работает как информационный процессор, однако его сознательная деятельность осуществляется не только с помощью индуктивно-дедуктивных опе-

<sup>8</sup> *Lipman M. Thinking and Education, P. 238.*

раций или жестких алгоритмов. Она включает в себя реакции на меняющийся контекст (в том числе и его чувственную, феноменальную, качественную сторону), на характер собеседников, на ход разговора, на достигнутые в нем результаты. В ней присутствуют интерпретация и самоинтерпретация, эмоциональное, моральное и эстетическое отношение к участникам диалога и многое другое. Конечно, в контекстуально ориентированной человеческой дискуссии случаются формально-логические ошибки, она грешит избыточностью и неточностью понятийных средств, а итоговое суждение о разрешении проблемой ситуации может быть хуже того, которое могло быть выдано совершенным компьютером. Однако дефектность такой дискуссии является источником сомнений, а сомнения толкают к новым поискам. И, что важно, в часто спонтанной и многовекторной деятельности диалогического мышления происходит творческий скачек к новому уровню понимания — разумному пониманию ситуации и даже к *рождению нового смысла*.<sup>9</sup> Иначе говоря, можно сказать, что в опыте «сократического диалога» задействованы более *гибкие, качественно разнообразные критические и творческие навыки мышления*.

Особенность живого диалога — *устная речь*. Эта самая древняя, как говорят ученые, форма коммуникации стимулирует глубинные (имеющие эволюционные корни) резервы интеллекта. В спонтанном акте говорения и сообщения информации другому лицу задействованы не просто лингвистические навыки, но и навыки быстро оформлять мысли с учетом *быть понятым другим*. В речевом акте взаимопонимания присутствует невербальный момент общения; семиотические функции могут выполнять интонации, взгляды, эмоции, мимика и т.д. Понимание часто достигается с полслова, взглядом, чего не осилит самый совершенный компьютер. Несмотря на несовершенство, данные свойства фонетического общения играют огромную роль в мыслительном процессе. Поэтому интеллектуальное приключение в «обществе исследователей», ведущем философский «сократический разговор», не тождественно интеллектуальному приключению во взаимодействии с программой компьютера.

Помимо когнитивных и феноменальных факторов важнейшую роль в «сократическом разговоре» играют *моральные и социально-психологические факторы*. Забота о конечной цели разговора и ответственность за достижение взаимопонимания являются важными навыками в

---

<sup>9</sup> Х. Дрейфус и С. Дрейфус, сравнивая возможности искусственного интеллекта и обыденного, формально неструктурированного опыта социального общения, высказываются в пользу последнего с точки зрения развития понимающих способностей человека. ( *Х.Дрейфус и С.Дрейфус*. Создание сознания vs. моделирование мозга: искусственный интеллект вернулся на точку ветвления // Аналитическая философия: становление и развитие. М., 1998, Р. 422.



интеллектуальном и нравственном росте личности. Трудно вообразить компьютер, заботящийся о конечном результате, или берущий на себя ответственность за качество работы, или испытывающий утрызения совести за плохую работу с пользователем. Одина из целей «сократического разговора», во всяком случае, в том виде, в каком он представлен в программе «Философия для детей», — состоит в **самоидентификации индивида через участие в коллективной деятельности**. В живом «сократическом диалоге» осознание своего Я приобретается в процессе взаимно обмена аргументами, в осознании своих целей, интересов, предпочтений в сравнении с целями, интересами, предпочтениями других равных ответственных личностей. Здесь больше возможностей проявить свою индивидуальность свободу и меньше опасностей стать объектом манипуляции анонимных сил. В случае коммуникации с машиной создается только видимость сохранения индивидуальности, поскольку партнер — машина — лишена Я, свободы и ответственности.

Выше мы отметили как достоинство демократизм новых компьютерных информационных технологий. Однако это замечание требует оговорки: демократизм здесь ограничен сферой трансляции информации. С точки зрения содержания, полученная из персонального компьютера информация для учащегося — это **авторитарная власть** всезнающего ящика, а диалог с ним — это диалог «знайки» и «незнайки». Компьютерный мир информации и игр не является личной принадлежностью учащихся. Он создан анонимными дизайнерами «твердой» и «мягкой» программ и подчинен логике программирования. В этом таится опасность подавления инициативы пользователя. В то время как философский диалог в классе со сверстниками происходит на равных, живая полемика и столкновение аргументов и контраргументов ближе подлинной (демократической) коммуникации.

Отдав приоритет в обучении качественному мышлению философскому «сократическому разговору», я хотела бы высказать ряд сомнений. В теоретическом плане вопрос о педагогических потенциях двух типов коммуникации — человека с человеком и человека с машиной — очень сложный, дать «окончательный» и «однозначный» ответ на него на основе проведенного сравнения затруднительно. И вот почему. Все зависит от понимания природы сознания, опыта, мышления, разумности, что в свою очередь, определяет позицию по вопросу о возможностях естественного и искусственного интеллекта. Все эти сюжеты в высшей степени дискуссионные. Например, сторонники «слабого AI» (Дж. Серль, Т. Виноград и др.) считают, что компьютер способен имитировать (дублировать) когнитивную сторону опыта сознания, но не феноменальную его сторону, поэтому понимание и взаимопонимание для него недостижимы. Сторонники «сильного AI» (например, Д. Деннет и др.) утверждают, что в принципе компьютеры

в состоянии преодолеть этот недостаток, при условии отказа от идеи о реальности ментального, феноменальных качеств опыта (или квалиа). На сегодня нам представляется более убедительной концепция «слабого AI».

Если принять эту концепцию, правдоподобным представляется предположение, что в разговоре человека с человеком имеет место **принципиально иной тип коммуникации**, нежели в общении с персональным компьютером. Приобретаемые в живом диалоге исследовательские, когнитивные и творческие навыки мышления качественно иные, нежели в случае общения с персональным компьютером. Вероятнее всего имеет место и более совершенное качество мышления. (Поэтому мы бы предложили называть общение с персональным компьютером не «сократическим диалогом», а более слабым термином — «интеракция»). Из этого предположения следует вывод, что у нынешних детей гораздо больше шансов стать разумными, интеллектуально и морально ответственными **личностями**, пройдя через практику философского диалога, нежели практику общения с компьютером. Данное предположение, если оно правильное, заслуживает внимание со стороны тех, кто занимается сравнением возможностей естественного и искусственного интеллекта. Еще большее внимание оно заслуживает со стороны тех, кто занимается оптимизацией образовательного процесса и кого тревожит качество мышления российских школьников<sup>10</sup>.

- 
10. В опубликованной в «Московском комсомольце» статье («Вот такая PISA... 13 января, 2005 г., С. 6) говорится о результатах исследования успехов российских школьников, проведенного Международной программой по оценке образовательных достижений учащихся (Program for International Student Assessment—PISA). Согласно этим результатам, **«по умности» российские 15-летние школьники занимают 28 место в мире после Макао и африканских стран**. Оказалось, что ни система математического и естественнонаучного образования, которой мы так гордились, ни введение в классы компьютеров, существенно не повлияли на качество мышления школьников. Исследования показывают, что разница между студентами, которые общаются или не общаются с компьютером и Интернетом — это разница в количестве информации, а не в качестве их мышления. Основные изъяны, которые отмечались у российских школьников — это недостаточное развитие у них самостоятельного мышления: навыков анализа и саморефлексии, умений критически оценивать свою работу, соотносить знание с контекстом, пользоваться гибкими методологическими инструментами.

---

# КАК В НАСТОЯЩЕЕ ВРЕМЯ МОЖНО ОСМЫСЛИТЬ КОНЦЕПЦИЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА?

*В.М. Розин*

В.М. Розин

В Новой философской энциклопедии искусственный интеллект характеризуется как «научно-техническая дисциплина», решающая в основном две задачи – моделирования «процессов познания и мышления» и создания технических устройств, в которых реализуются результаты такого моделирования, но также которые могут быть названы «интеллектуальными» [2; с. 159]. В концептуальной трактовке искусственного интеллекта есть известное противоречие. С одной стороны, познание и мышление здесь рассматриваются как деиндивидуальные процессы, например, получения, передачи и переработки информации, с другой – по-прежнему интеллект понимается как особая инстанция (способности) индивида. Для современного этапа развития исследований в области искусственного интеллекта, отмечает Ю.Петрунин, характерно «изучение и моделирование рациональных структур в связи с навыками и неаналитическими методами обработки образной информации, что сближает современные модели искусственного интеллекта с их естественным человеческим прототипом» [2; 159].

И на первый взгляд, кажется, что победа в поединке с Гаспаровым машины, подтверждает это высказывание. Такая машина, играющая в шахматы и побеждающая сегодня человека, действительно, воспринимается как его интеллектуальный прототип. При этом сторонники так понимаемого искусственного интеллекта локализируют мышление и познание человека в индивиде и считают, что их можно не только рационально описать, но промоделировать в естественно-научном ключе. Последнее означает, что мышление и познание человека может быть представлено в научном исследовании как естественный механизм и затем воспроизведено технически в форме искусственного механизма.

Можно ли согласиться в настоящее время с такими представлениями, лежащими в основании концепции искусственного интеллекта? Вряд ли. При этом стоит развести реальную практику и концептуализацию. В реальной практике изучения интеллекта и конструирования интеллектуальных машин все же речь идет о развитии технических наук и технологий, повышающих производительность вычислительной техники или позволяющих создавать машины и механизмы, частично заменяющие человека, например, роботы. Но это

не означает, что робот или машина, играющая в шахматы, мыслит в человеческом смысле слова. Вспомним историю развития шахмат. Сначала шахматисты просто запоминали партии. Затем они стали их записывать и анализировать, создавая специальные шахматные технологии. Затем настало время подготовки самих шахматистов; были созданы образовательные и тренировочные (к соревнованиям) технологии подготовки шахматистов. Как при этом менялось мышление шахматиста, никто не знает, да и это было не нужно. Был создан обходной путь – мышление шахматиста становилось все более и более эффективным за счет шахматных и образовательных технологий. Кстати, шахматные и образовательные технологии учитывались при разработке программ для шахматных машин. В этом плане такие программы моделировали не реальное мышление шахматиста, а соответствующие шахматные и образовательные технологии.

Но и с точки зрения современных представлений о мышлении (естественно, не всех, а ряда направлений) приходится отказаться и от идей локализации мышления (познания) в индивидуе и понимания мышления как процессов получения, передачи и переработки информации. Изучение мышления показало, что его нужно понимать двояко: с одной стороны, как деиндивидуальные процессы семиотического значения, нормирования и коммуникации, с другой – как реализацию и жизнедеятельность личности. Конечно, мыслит отдельная личность, но условия и детерминанты мышления обусловлены не только ее уникальными мотивациями и реальностями, но и указанными деиндивидуальными процессами и структурами. И локализовано мышление как в отдельном индивидуе, так и в семиозисе, социальных институтах (задающих нормы), социальной коммуникации. Когда устремление и активность личности совпадают с тенденциями и запросами социума, происходит развитие мышления, в противном случае имеет место опережение времени (Архимед) или пустое творчество. Понятно, что говорю я в данном случае о мышлении, которое в оппозиции к искусственному интеллекту можно назвать естественным.

Одним из важнейших механизмов развития естественного мышления выступает технология. Само изобретение знаков и создание социальных институтов представляют собой особые социальные технологии. Но и правила логики и категории, как я стараюсь показать, – это тоже технологии. И образование, в контексте которого формируется мышление – особая технология. Означает ли сказанное, что естественное мышление включает в себя искусственное? Безусловно, но это искусственное, конечно, отличается от технической деятельности. Тогда, каким образом можно охарактеризовать практику искусственного интеллекта? Как формирование одной из интеллектуальных технологий, способствующих развитию современного мышления. Наряду с этой технологией в настоящее время можно указать еще не-

сколько. Во-первых, это методология. Как интеллектуальная рефлексивная и технологическая дисциплина методология, безусловно, способствует развитию современного мышления. Во-вторых, на развитие современного мышления оказывают влияние деловые и имитационные игры. Игровые «машины» сегодня достаточно эффективны. В-третьих, по-прежнему на развитие мышления существенно влияет изобретение новых семиотических средств (знаков, схем, нарративов и пр.). В-четвертых, машинизация (компьютеризация) перечисленных процессов – еще одна тенденция, обуславливающая развитие современного мышления.

Другими словами, я хочу сказать, что кибернетически ориентированное моделирование мышления и познания (в рамках традиционной концепции искусственного интеллекта) – это только одно из направлений технологизации. Есть и другие, оказывающие на развитие мышления не менее существенное влияние. Но тогда, вероятно, необходимо заново определиться и с самой концепцией искусственного интеллекта. Искусственный интеллект – это все основные линии технологизации, существенно влияющие на развитие естественного мышления. С этой точки зрения, к искусственному интеллекту в настоящее время нужно отнести и практику создания игровых машин, и современную методологию, и Интернет, как оказывающий серьезное воздействие на современное мышление (вряд ли последнее кто-нибудь будет отрицать), и семиотические эксперименты, и кибернетически ориентированное моделирование мышления, и, наконец, машинизацию всех перечисленных здесь практик. Думаю, что при таком понимании концепция искусственного интеллекта обретет второе дыхание, стимулируя многочисленные новые исследования и разработки.

Чтобы лучше понять, что я имею в виду, говоря о естественном мышлении, приведу результаты исследования последнего.

## 2

В истории философии первая рефлексия мышления принадлежит Аристотелю, его же можно считать мыслителем, конституировавшим античное мышление. Однако предпосылки мышления прослеживаются, начиная с древнего мира. К ним относятся, с одной стороны, первые способы получения атрибутивных или мифологических знаний, проходивших опытную проверку в сфере общественной практики, с другой – формирование личности человека, обуславливающей индивидуальное видение и понимание событий, с третьей – изобретение в древней Греции рассуждений, позволявших человеку без обращения к общественной практике получать одни (в частности, новые) знания на основе других (уже известных) [3-7].

Аристотель, преодолевая проблемы, связанные с изобретением рассуждений (возможность получать противоречия и утверждать о

действительности все, что угодно), создает **систему норм** (правил) мышления, регулирующих построение рассуждений, а также формирует **представления и средства, обеспечивающие применение этих норм** (к последним относятся категории, картина построения наук, психологическое обоснование новых уже собственно мыслительных способов рассуждения). В такой ситуации и складывается античное мышление как сложная деятельность и реальность, одновременно, индивидуальная, поскольку, выражая свое видение, мыслит отдельная личность, и деиндивидуальная, так как, мысля, человек соблюдает общие правила (то есть следует логике), относит полученные знания к действительности, заданной категориями (то есть онтологией), реализует общие установки и принципы мышления (например, стремится к истине, избегает противоречий).

Построение Аристотелем правил мышления («Аналитики», «Топика», «О софистических опровержениях») и обоснование этих правил и начал («Метафизика») имело колоссальные последствия для всего дальнейшего развития человеческого интеллекта. Человек получил в свои руки мощное орудие мысли: возможность получать знания о действительности, не обращаясь к ней непосредственно. Правила мышления позволяли включать в рассуждение одни знания (ранее доказанные или эмпирические, или же априорно верные – начала) и получать на их основе другие знания (как уже известные, так и новые). При этом новые знания не приводили к противоречиям, и их не нужно было оправдывать опытным путем. Начиная с этого периода, формируются собственно научное мышление и отдельные науки.

В теоретических целях античное мышление может быть представлено в качестве своеобразной семиотической машины. Узлами ее являются знания, правила, категории, описания (схемы) самого мышления. Действительно, все это не только интеллектуальные конструкции, но и семиотические образования: они обозначают определенные фрагменты действительности. Так аристотелевские правила обозначают типы рассуждений или доказательств и различные мыслительные процедуры (например, не приводящие к противоречиям). Категории обозначают типы содержаний (задают схемы объектов), к которым могут быть отнесены знания, полученные с применением правил. Знания обозначают объекты, истолкованные категориально. Получение знаний в рамках античной философии и науки предполагает организацию эмпирического материала и действий человека в соответствии с «логикой» мышления: на основе эмпирического материала (эмпирических объектов) создаются идеальные объекты, а мыслительные действия подчиняются правилам и схемам мышления. В результате и возникает феномен, о котором пишут Кант и Делез – рассудок, предполагающий подведение частного случая под правила и категории, «мышление как узнавание». Кстати, не нужно заблуждаться, сама процедура подведе-

ния может быть достаточно сложной и многоходовой, в чем можно убедиться, анализируя, к примеру, работы Архимеда.

При этом Делез считает, что мышление как узнавание, как мысль, «наполненная собственным образом» – это не мышление. Однако почему? Периодически в развитии мысли и способах построения знания возникают ситуации, требующие «остановки мысли», создания машин мышления. Как правило, это ситуации, в которых возникают противоречия и другие заторы в мышлении, например, складываются принципиально новые способы построения знаний, критикуются, как неэффективные, старые способы и т. п. Создание машин мышления необходимо и для массового распространения в культуре новых способов получения знаний.

Например, деятельность Канта можно рассмотреть как работу по построению машины мышления, соответствующей Новому времени. Действительно, Кант старается разрешить проблемы, возникшие из осмысления философии Локка, Юма, Лейбница, Беркли, и обосновать математическое и естественнонаучное мышление Нового времени. Кроме того, в некотором отношении сложившаяся интеллектуальная ситуация напоминала античную, более точно, преаристотелевскую: были созданы и конкурировали между собой разные философские системы и взгляды, по-разному объяснявшие мир и его явления. Аппелируя к опыту математики и естествознания, Кант предполагает, что именно мыслящий (ученый, философ), рассуждая, связывает знания, порождает опыт, привносит в природу законы. При этом рассудок и направляющий его разум оперируют понятиями и категориями, независимыми от опыта, поэтому Кант называет их априорными. Для объяснения роли опыта, на котором стояло все естествознание, Кант вводит понятия «созерцания» как необходимого условия познания, «вещи в себе», «явления» и «предмета». Суть решения состояла, с одной стороны, в постулировании двух реальностей – «трансцендентальной», где с помощью разума (точнее, «в сфере разума»), как говорит Кант «а priori» создаются научные и философские знания, и «эмпирической», где на основе созерцания имеет место опыт. С другой стороны, именно созерцание и опыт выступают как необходимое условие изучаемого в науке или философии явления и предмета. С третьей стороны, использование априорных представлений (понятий и категорий) рассматривается Кантом как необходимое условие самого опыта.

Мышление Кант понимает иначе, чем Аристотель. Последний еще не осознает роль мыслящего (у Аристотеля действия философа сливаются с действием разума-божества), Кант же, опираясь на декартовскую новоевропейскую традицию личности, уже понимает, что именно мыслящий, как он пишет, «сам является творцом опыта», «сам вкладывает» в объект необходимые априорные характеристики, «сам связывает знания и характеристики объекта». Одновременно

Кант утверждает, что разум человека следует «вечным и неизменным законам» разума. Другими словами, разум Кант понимает двояко: как разум (мышление) отдельного эмпирического человека и разум как таковой, как особую природу, законам которой подчиняется отдельный эмпирический разум, отдельный правильно мыслящий человек.

Можно указать три основные задачи, которые решал Кант, создавая «Критику чистого разума». Первая вполне в духе Аристотеля – охарактеризовать неправильные способы мышления и рассуждения и объяснить, чем они были обусловлены. К подобным способам Кант относит, например, такие, которые приводят к противоречивым высказываниям о Боге, душе, а также бесконечности вселенной (известные кантовские антиномии разума); другой пример – те основополагающие построения Локка, Юма, Лейбница, с которыми Кант был принципиально не согласен. Вторая задача – построить философию по образцу новой науки, естественно, могла относиться только к Новому времени. Кант не скрывает, что идеал такой науки для него задают математика и естествознание. Наличие в идеале новой науки этих двух дисциплин должно было вести к пониманию философии, с одной стороны, как описывающей «формы мышления» (философия как математика), с другой – «законы мышления» (философия как естествознание). Третья задача, естественно вытекающая из второй (построить философию как строгую, точную науку), заключалась в необходимости научно объяснить целый ряд проблем и феноменов, интересовавших философов того времени. Например, отношение между философией и наукой, возможность антиномий разума, отличие математики от естественной науки, отличие феноменов от ноуменов, природу центральных философских представлений (например, онтологии).

Важным методологическим принципом для Канта являлся принцип соответствия, который, по сути, прямо следует из принятой Кантом установки на системное представление разума. Используя этот принцип, Кант устанавливает, например, соответствие отношения «рассудок – разум» с понятиями и правилами мышления: рассудку он ставит в соответствие правила и категории, а разуму – принципы и идеи (концепции). Кроме того, на основе того же принципа Кант строит таблицу категорий, структурно уподобляя ее таблице функций рассудка; аналогично, таблице категорий ставится в соответствие система трансцендентальных идей. Другой, не менее значимый методологический принцип, постоянно реализуемый Кантом – переосмысление и перестройка основных традиционных философских понятий. Этот принцип является естественным следствием установки Канта на построение своей системы как строгой науки. Для науки всякое значимое понятие или представление, полученное вне нее (в данном случае в других философских системах), является только эмпирическим материалом и для



научных целей должно быть специально введено в систему этой науки (проблематизировано в ней, представлено в рамках ее онтологии, доказано и т. д.). Например, Кант показывает, что известное философское понятие онтологии представляет собой аналитику чистого рассудка, а идущее еще от Платона понятие ноумена, есть демаркационное понятие, служащее для ограничения притязаний чувственности и потому имеющее только негативное применение.

На мой взгляд, реконструкция творчества Канта демонстрирует два важных момента: с одной стороны, он, отчасти по образцу Аристотеля, строит новую машину мышления, с другой – как бы сказал Делез, «размышляет» и тем самым создает необходимое условие для интеллектуальной встречи (Канта с математиком, физиком, этиком, современным ему философом), для «мышления-события».

*Мышление-событие, мышление-встреча.* Что ведет мысль Канта? Не столько правила рассудка или основоположения, хотя и то и другое корректирует ход мысли. Мысль Канта ведут необходимость осуществить критику традиционного мышления, решить три рассмотренные задачи, вера в Творца и его роль-разума в мышлении, убежденность Канта, что, с одной стороны, именно мыслящий порождает реальность, а с другой – что это порождение ограничено опытом. Размышляя, рассуждая, создавая мыслительные схемы, Кант удовлетворяет (разрешает) все перечисленные требования, психологически, конечно, исходящие от самого Канта, но объективно – от мышления. Но в данном случае мышление – это не действие машины, а реализация личности Канта, реализация посредством творчества Канта требований Новоевропейской культуры, создание Кантом условий для встречи с ведущими актерами «театра» Нового времени (математиками, физиками, этиками, философами).

Лучше я почувствовал разницу между машиной мышления и мышлением-со бытием, когда стал сравнивать два типа своих реконструкций гуманитарных текстов. Один, действительно, сводился к «пересчету» творчества определенных авторов, то есть я их подводил под уже имеющиеся схемы и представления. Например, анализируя творчество Галилея, я хотя и обсуждал некоторые особенности его личности (склонность к платонизму, интерес к технике и одновременно к науке по Архимеду, маниакальное упорство в отстаивании своих идей, гибкость мышления), тем не менее, и личность и творчество Галилея для меня выступали только как объекты изучения. Никакого личного отношения к Галилею у меня не было. Аналогично, анализируя «Исповедь» Августина, с которым я, естественно, никак не пересекался, я представил его творчество и путь как эзотерические, применив для этого выработанные мной ранее представления об эзотеризме. Как эзотерик Августин критикует и отвергает ценности обычной жизни, которая в данном случае совпадает с античной,

утверждает существование подлинной реальности (христианского Бога и Града Божьего), понимает цель своей жизни как обретение подлинной реальности (возможность придти к Богу), открывает в себе эзотерического человека (он его называет «внутренним человеком»), познает подлинный мир, одновременно порождая его. Говоря словами А. Пузыря, я «пересчитал» Августина, применив к его творчеству отработанную на другом материале схему.

Иная познавательная ситуация имела место в случае исследования творчества А.С.Пушкина. Читая его письма, я как-то поймал себя на мысли, что мне совершенно не понятны ни поступки, ни высказывания великого поэта. В то же время, и игнорировать свое непонимание я не мог, слишком велико в моей душе было значение Пушкина; следуя за Мариной Цветаевой, я вполне мог сказать – «Мой Пушкин». Я не мог и жить с таким пониманием, точнее непониманием, и отмахнуться от возникшей проблемы. Читая дальше письма, я с определенным удовлетворением отметил, что сходная проблема не давала покоя и Петру Чаадаеву. Чаадаев пишет, что Пушкин «мешает ему идти вперед», спрашивается, при чем здесь Пушкин, иди вперед, если хочешь. Но в том-то и дело – если Пушкин мой, во мне, часть моего я, то не могу отмахнуться, если не понимаю или не одобряю его поступки.

В результате я вынужден был начать сложную работу. Вспомнив совет Михаила Бахтина, который писал, что «чужие сознания нельзя созерцать, анализировать, определять как объекты, как вещи, – с ними можно только диалогически общаться, думать о них – значит говорить с ними, иначе они тот час же поворачиваются к нам своей объектной стороной», я предоставил голос самому Пушкину, чтобы он отвечал на мои недоумения. Для этого я искал в его письмах ответы на мои вопросы, пытался встать в позицию Пушкина, увидеть мир его глазами, сам и с помощью Ю. Лотмана реконструировал его время, нравы, обычаи и т.д. Я анализировал поступки Пушкина и старался понять их мотивы, короче делал все, чтобы Пушкин, действительно, стал моим, чтобы Пушкин, как писал Чаадаев, позволил мне идти своим путем, чтобы я смог жить вместе с Пушкиным. Не знаю, как это выглядит со стороны, но психологически мне это, в конце концов, удалось.

При этом я, безусловно, вел исследование творчества Пушкина, но главным было не подведение Пушкина под какую-то известную мне схему или теорию творчества, а движение в направлении к Пушкину. То есть мое исследование, как тип мышления, представляло собой создание условий для нашей встречи, для общения. Структура и «логика» мысли задавались в данном случае не правилами, категориями или построенной ранее схемой, хотя все это я использовал по мере надобности, а именно работой, направленной на встречу и общение с Пушкиным.

Встретившись с Пушкиным, я спросил себя, а нельзя ли сойтись поближе и с Августином? Однако ведь последний не встал поперек моего жизненного пути подобно Александру Сергеевичу. Я не мог сказать – «Мой Августин». Что же делать, значит мне дорога к Августину окончательно закрыта? И тут, вдруг, я вспомнил свой чуть ли не трехлетний диалог с известным философом, ученым и католиком Юлием Анатольевичем Шрейдером по поводу веры. Он отказывал мне в духовности потому, что я не верил в Бога, утверждал, что, поскольку я не имею религиозного опыта, то со мной нечего разговаривать, что истина мне недоступна. Я возражал ему на это, говоря: можно жить духовно и без веры в Творца, что я не понимаю, как это такой ученый и философ, как Шрейдер, может верить в вечную жизнь, чудеса Священного писания, Бога как личность и троицу. Мы спорили много месяцев, наговорили кучу умных вещей, но не продвинулись друг к другу ни на шаг. Каждый точно по Бахтину выражал себя, относясь к другому, но оставался при своем.

Вспомнив все это, я подумал, что моя позиция в диалоге со Шрейдером очень похожа на ту, которую имел Августин вначале своего жизненного пути (он любил жизнь во всех ее проявлениях, был ритором, философом и неверующим человеком), а позиция Шрейдера – на ту, к которой, в конце концов, пришел Августин (он стал не только глубоко верующим, но и активно разъяснял и распространял христианское учение). Тогда я понял, что у меня с Августином, оказывается, идет диалог, что это против меня Августин направляет свои стрелы и обличения, что я должен ему как-то ответить.

Я снова стал читать и анализировать «Исповедь» и увидел то, что я не видел раньше. Увидел, что вокруг Августина все, начиная с любимой матери, приходило к христианству, и это не могло не влиять, заставляя «верить в Бога еще до веры». Августин проделал огромную работу, переосмысляя обыденные представления о Боге. Я прошел шаг за шагом вслед за Августином и понял, как идея Творца сначала для него совершенно неприемлемая и неправдоподобная, постепенно становилась все более понятной и необходимой и, в конце концов, превратилась в реальность, в которой нельзя было усомниться. При этом опять я использовал различные представления и схемы, размышлял, но вела меня забота сделать шаг навстречу Августину (и Шрейдеру), а также дать им полноценный голос.

Думаю, моя мысль стала понятнее. Мышление как событие и встреча – это определенная *форма жизни личности, осуществляемая с помощью рассуждений и размышлений, создающая условия для встречи данной личности с другими*. Одновременно это может быть и *форма социальной жизни, реализуемая через творчество личности и размышления*. Создавая «Исповедь», Августин способствовал реали-

зации средневекового социального проекта – созданию новой культуры на основе христианства.

Но почему, спрашивается, не сказать иначе: рассмотренные здесь моменты, а именно, личность, культура, проблемы общения и коммуникации и т. д. – это всего лишь факторы мышления, само же оно есть активность и творчество личности, рассуждающей, размышляющей, использующей правила и категории? Но ведь тогда придется понятийно перегрузить саму личность и машину мышления, приписав им в качестве имманентных свойств все культурное многообразие западной цивилизации; придется все новообразования, весьма различающиеся в разных культурах и типах человека, заложить в человеке с самого начала, пусть даже в потенциальной форме.

Если машина мышления строится именно так, чтобы возможно было мыслить в некотором смысле, не думая, то мышление как встреча и событие – это всегда **уникальное негарантированное действие**, состоится оно или нет, зависит не от законов мышления, а от того, как «здесь и сейчас» сойдутся различные элементы и обстоятельства. Тем не менее, после того как мышление состоялось, событие случилось (почему, как раз здесь и на этом человеке – это всегда тайна), то становится возможным отрефлексировать структуру мысли, попытаться понять, что ее обусловило и предопределило. Понятно, что полученные при этом знания могут быть использованы при осуществлении, разворачивании новых действий и порывов мысли, но в качестве чего? Не законов мышления, а всего лишь для сценирования и конституирования новой мысли, понимая, что помимо этих «знаний о состоявшемся мышлении» действуют и другие не менее существенные и обычно слабо осознаваемые факторы. Дальше я приведу несколько примеров подобных знаний о уже состоявшемся мышлении.

Мой анализ показывает, что смена типов культуры (например, переход от античности к средним векам и далее к Возрождению, а затем и к новому времени) обуславливает не только смену машин мышления, но и формирование «ситуаций мышления-встречи, мышления-события». Действительно, например, в средние века задачи мышления кардинально изменились. Главным теперь становится не познание областей бытия и упорядочение рассуждений, что было характерно для античности, а критика на основе христианских представлений античных способов объяснения и понимания мира и человека, а также уяснение и объяснение новой реальности, зафиксированной в текстах Священного писания. Обе эти задачи можно было решить только на основе мышления, поскольку формирующийся средневековый человек перенимает от античности привычку рассуждать и мыслить, а также потому, что новая реальность хотя и выглядела привлекательной и желанной, но одновременно была достаточно непонятна. Что собой представлял Бог, как он мог из ничего создать мир и человека, почему он

одновременно Святой Дух, Отец, и Сын, как Бог воплотился в человека Христа и что собой Христос являл – Бога, человека или их симбиоз, как понимать, что Христос воскрес – эти и другие сходные проблемы требовали своего разрешения именно в сфере мысли.

Как показывает известный российский медиевист С. С. Неретина, на средневековое мышление существенно влияли два фактора: сервилитская роль мышления по отношению к христианской религии (задачам спасения) и необходимость удовлетворить «логике» отношений «сакральное – мирское». Действие первого фактора приводит к этической нагруженности средневекового мышления, а второго – к присущей средневековым понятиям «двуосмысленности» [1]. Когда, например, Иустин (II век) пишет, что «Бог не есть имя, но мысль, всаженная в человеческую природу, о чем-то неизъяснимом», то здесь «мысль» понимается двояко: как относящаяся к Богу и к человеку; в первом своем значении понятие «мысль» указывает на трансцендентальную сущность, во втором – на содержание обычного человеческого мышления. Средневековое мышление, если его сравнивать с античным и современным, весьма необычно и выглядит противоречивым. Но как показывают современные исследования, средневековое мышление вполне целостно и органично своему времени. Оно основывается на двух типах схем: заимствованных из Священного писания и переосмысленных на их основе схемах античного мышления. Соответственно, двуосмысленны также средневековые категории и онтология. Чтобы создать новую машину мышления и вообще осуществлять индивидуальную и социальную жизнь, средневековые мыслители, начиная от отцов церкви и философов, размышляют подобно Августину, Боэцию, Абеляру, создавая мыслительное пространство и поле, в котором только и может разворачиваться средневековая жизнь (неверующие приходят к Богу, начинают действовать в соответствии с требованиями христианства, готовятся к Страшному суду и встрече с Творцом и прочее).

Внутри каждой культуры складываются свои типы коммуникации и личности, что тоже предполагает и обуславливает новые формы мышления. Уже изобретение рассуждений в ранней античной культуре позволило конституировать совершенно новый тип коммуникации: не человека с богами, а одной личности с другой. Личность типа Сократа с помощью рассуждений может приводить другого человека к себе, ведь в рассуждении человека заставляют перейти от его собственного знания к другому знанию, которое разделяет рассуждающая личность.

Создание Аристотелем машины мышления – это следующий тип коммуникации и личности, когда эти движения с помощью рассуждений начинают подчиняться общезначимым правилам и позволяют получать непротиворечивые знания о действительности. То есть но-

вая личность – это человек не только действующий самостоятельно (как Сократ), но и подчиняющийся общему порядку.

Переход к средним векам знаменует собой переструктурирование коммуникаций и самостоятельного поведения: человек ориентируется теперь не только и не столько на себя, но не меньше – на другого человека, бескорыстную помощь (любовь к ближнему), а также на целое (общину, государство, Град Божий). Этическая нагруженность (например, та же идея христианской любви) и двусмысленность средневековой мысли как раз и обеспечивают этот новый тип коммуникации и личности.

В новое время потребовалась естественнонаучная и инженерная мысль, чтобы передать власть новоевропейской личности, основывающей свои действия и жизнь на вере в законы природы. Потребовалась гуманитарная мысль, чтобы дать слово личности по Бахтину. Социально-психологическая мысль, чтобы создать условия для личности и коммуникации по Шебутани, основанных на идее согласованного поведения и ожиданиях (когда все основные структуры личности – «Я-Образы», ценности, мотивация и прочее формируются в ответ на требования и ожидания Других). Постмодернистская мысль и деконструкция, чтобы возвести вокруг личности стену до небес, а также заблокировать претензии других на власть. Сегодня формируются новая коммуникация и личность: помимо задач приведения другого к себе и самовыражения все более настоятельны требования **приведения себя к другому** (встречи-события), а также **ориентация самостоятельного поведения человека на других, сохранение природы, безопасное развитие человечества**.

Существенно меняется структура мысли и условия для мысли-встречи, мысли-события, когда принимаются и начинают осуществляться новые «социальные проекты». Одним из первых социальных проектов можно считать задачу Аристотеля и его школы: нормировать рассуждения и доказательства и затем заново, опираясь на построенные нормы, получить знания об отдельных областях бытия; последнее, как известно, вылилось в создание античных наук.

Второй проект – перестройка античного органа и мировоззрения на основе текстов Священного писания. Третий, относящийся к XVI-XVII вв., не менее грандиозный – овладение силами природы, создание новых наук о природе и новой практики (инженерной). Четвертый, складывающийся уже в настоящее время – **перевод цивилизации на путь контролируемого и безопасного развития**. Сакраментальный вопрос – удастся ли этот проект реализовать без прохождения «точки Конца Света»?

Распад существующей культуры или становление новой создают широкие поля для мышления-встречи, мышления-события. Как пра-

вило, в этот период необходима критика традиционных способов мышления и представлений и формирование новых подходов.

Мишель Фуко и Мераб Мамардашвили, а задолго до них Платон утверждали, что необходимым условием подлинного мышления является работа человека над собой, направленная на конституирование и изменение собственной личности. Например, Фуко считает, что современный философ – это человек критически относящийся к себе, ко всему, что он делает, к тому, как он мыслит и чувствует, человек постоянно себя воссоздающий, конституирующий себя, анализирующий и уясняющий свои границы (критика, замечает Фуко, в работе «Что такое просвещение?» собственно и есть анализ границ и рефлексия над ними). Мераб Мамардашвили в своих «Лекциях о Прусте» трактует эту работу как особый духовный опыт спасения, в ходе которого человек уясняет и открывает другой мир (миры) и учится жить в нем.

Однако, спрашивается почему, разве нельзя просто философствовать или мыслить, не работая над собой, не меняя себя, не прорываясь в какие-то другие реальности? Если речь идет о становлении новой культуры, то, вероятно, нет, ведь человек подобно змею должен сбросить с себя старую кожу и нарастить новую. Например, реформа Декарта заключалась в создании ценностных ориентиров и реальности, переключавших мыслителя с традиционного средневекового мироощущения на новое. Декарт создает такие представления, которые, с одной стороны, позволяют мыслить самостоятельно, не оглядываясь на Бога и Священное писание, с другой – рисуют мир, вполне отвечающий идеалам новой науки. Как пионер, открывающий новый мир и самого себя в качестве человека Нового времени, Декарт не мог не быть эзотеричным, поэтому его философия, так же и философия Платона или Августина, может быть названа «философией спасения». Мой анализ показывает, что при смене культур философия спасения предвзает и готовит, так сказать, нормальную философию. Другая роль философии спасения – создание схемы и образа нового человека, способного мыслить в соответствии с реалиями складывающейся культуры Нового времени.

Современные исследования все больше подводят нас к пониманию, что картина, в которой человек и мир разделены, неверна. Сегодня мир – это созданные нами технологии, сети, города, искусственная среда, которые в свою очередь создают нас самих. Говоря о работе человека над собой, я имею в виду одновременно и работу, направленную на изменение нашей деятельности и жизни, что невозможно без изменения культуры и социума как таковых. Вопрос, как можно помыслить эти изменения, как мы должны при этом размышлять? Суммируем теперь основные характеристики мышления.

Мышление – это способ построения человеком одних знаний на основе других и построение представлений о действительности (схем, понятий). Будем считать это первой характеристикой мышления.

Вторая характеристика мышления – это одновременно способ познания действительности, обеспечивающий становление и функционирование культуры, и необходимое условие реализации личности, разрешающей в коммуникации несовпадение общепринятых и собственных представлений о мире и о себе.

Третья характеристика, мышление – это такой способ приведения в движение (смены, изменения) представлений о действительности, который помимо знаний о действительности и реализации личности выступает условием согласованного социального поведения.

Четвертая характеристика – мыслительная активность становится мышлением только в том случае, если эта активность нормирована (мышление как включающее в себя образцы мышления, правила логики, категории). При этом мышление задает самостоятельную реальность, второй, идеальный и конструктивный, мир относительно обычного. В его рамках перевоссоздаются и основные познаваемые явления. Нормирование мышления обеспечивает возможность, с одной стороны, строить знания без противоречий и других затруднений, с другой – получать знания, которым можно приписать свойство прагматической адекватности (истинности).

Необходимым условием реализации самого мышления является приписывание действительности определенного строения (пятая характеристика). Так рождаются картины мира и представления личности о себе (антропологические представления). В рамках этих картин и представлений осуществляется познание действительности и самопознание. Особенностью традиционного понимания действительности является неразличение трех указанных функций мышления (в отношении социальной и персональной реальности, а также коммуникации) и примат его социальной роли. Постмодернистская идеология, напротив, настаивает на приоритете личности и персональной реальности.

Шестая характеристика – в ситуациях становления новой культуры или решения социальных задач, а также ситуациях кризиса и становления личности: ***мышление – это новый опыт жизни, конституирование новой реальности***, при этом мышление перестраивается. В ситуациях функционирования культуры и личности мышление работает как машина.

Кризис современной жизни (глобальные проблемы, гипертрофированное развитие личности, обособление отдельных сторон самого мышления и прочее) обуславливает необходимость современного этапа конституирования мышления. Сюда входит, во-первых, пере-



ориентация мышления на решение нового социального проекта – *сохранение жизни на земле, безопасное развитие, поддержание природного, культурного и личностного разнообразия (многообразия) и сотрудничества, способствование становлению новой цивилизации, в рамках которой складываются метакультуры, новая нравственность, новые формы жизни и мышления..* Во-вторых, восстановление равновесия между социальным и личностным планами мышления. В свою очередь, это предполагает ограничение своеобразия современного человека, принятие им новых уровней ответственности, более решительный поворот к нуждам общества. В-третьих, сюда же относится *работа по созданию новых норм мышления*: не только образцов, правил, категорий, но и методологии. Именно методология позволяет, с одной стороны, направить и конституировать мысль, с другой – обеспечить ее разнообразие.

#### ЛИТЕРАТУРА

1. *Неретина С.С.* Верующий разум. К истории средневековой философии. Архангельск. 1995
2. *Петрунин Ю.Ю.* Искусственный интеллект // Новая философская энциклопедия. Т. 2. М., 2001
3. *Розин В.М.* Специфика и формирование естественных, технических и гуманитарных наук. Красноярск. 1989
4. *Розин В.М.* Культурология. М., 1998, 1999, 2000, 2001, 2002, 2003
5. *Розин В.М.* Любовь и сексуальность в культуре, семье и во взглядах на половое воспитание. М., 1999
6. *Розин В.М.* Человек культурный. Введение в антропологию. Москва-Воронеж, 2003
7. *Розин В.М.* Семиотические исследования. М., 2001

---

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ПСИХИКА, ТВОРЧЕСТВО

*А.Н. Кочергин*

При всем разнообразии трактовок понятия искусственного интеллекта общим для многих из них является признание того, что системы искусственного интеллекта моделируют функции человеческого мышления. При этом моделирование понимается не как воспроизведение мыслительных функций, а как их имитация [5]. Воспроизводить – значит воссоздавать то, что было или есть, т.е. при воспроизведении воссоздаются свойства, в совокупности сохраняющие сущность прототипа на всех его уровнях (субстратном, структурном, функциональном). Н. Винер и А. Розенблут не случайно определяли моделирование через *representation* (подобие), а не через *reproduction* (воспроизведение) [9, p. 317]. Такое понимание моделирования мышления позволило некорректно сформулированную проблему «может ли машина мыслить?» перевести в форму «какие мыслительные функции можно моделировать?», в рамках которой психические процессы рассматриваются как частная форма информационных процессов. Однако и подобный подход к психике встретил возражения части психологов.

Мозг является сложной самоуправляющейся информационной системой. Если учесть, что управление строится на основе переработки информации, то доминирующая роль информационных процессов в деятельности мозга становится понятной. Однако, хотя информационные процессы и являются существенной стороной психики, моделирование которых дает ключ к пониманию структур психических явлений, трудно объяснить все особенности психических феноменов с позиции лишь какого-то одного принципа. Так, например, на весьма существенные трудности наталкивается описание психических актов, связанных с эстетической деятельностью. Имеются данные о том, что в эстетической деятельности информационный процесс протекает иначе, чем в познавательной.

О.К. Тихомировым описан ряд функциональных механизмов, от которых, по его мнению, отвлекается информационный подход и которые должны быть объектом собственно психологического анализа: «характеристики операционального смысла ситуации для решающего, смысла конкретных попыток решения, смысла переобследования, смысла отдельных элементов в ситуации и отличие от их объективного значения; характеристики процессов возникновения и развития смыслов одних и тех же элементов ситуации и ситуации в целом

на разных стадиях процесса решения задачи, соотношение невербализованных и вербализованных смыслов различного рода образований в ходе решения задачи; процессы взаимодействия смысловых образований, роль смысловых образований в организации исследовательской деятельности, в определении ее объема (избирательности) и направленности; процесс возникновения и удовлетворения поисковых потребностей; изменение субъективной ценности, значимости одних и тех же элементов ситуации и действий, выражающееся в изменении их эмоциональной окраски (при константной мотивации); роль меняющейся шкалы субъективных ценностей в организации протекания поиска; формирование, динамика личностного смысла ситуации задачи и его роль в организации деятельности по решению задачи» [8, с. 296-297]. Действительно, приведенные данные ставят под сомнение «обоснованность критерия количества перерабатываемой информации как основного фактора, создающего трудность в решаемой задаче» [8, с. 231] – необходимо учитывать «такие реальные функциональные образования, как смысл (операциональный и личностный) и ценность информации...» [8, с. 297].

Однако все это говорит об ограниченности шенноновской теории информации, оперирующей понятием «количества информации», но не о принципиальной невозможности описания психических явлений в терминах концепции информации вообще. Понимание психики как информационного процесса дает возможность ее изучения с более широких позиций. Кроме того, «просвечивание» специфики общим позволяет эту специфику рассматривать более основательно. Д.И. Дубровским достаточно убедительно обосновано, что информационный подход позволяет не только уяснить в общем сущность субъективных феноменов, но и конкретизирует две важные взаимосвязанные проблемы: проблему расшифровки нейродинамического кода и проблему объяснения управляющей функции субъективных феноменов на уровне личности [4, с. 277].

Переработка информации – не синоним мышления, хотя любой познавательный акт есть процесс переработки информации. Но игнорирование информационной стороны мышления не дает возможности понять, как субъективные феномены влияют на физические процессы. Речь идет не о воспроизведении мышления, а о его имитации, т.е. о приравнивании информационных и психических процессов на уровне определенной абстракции, не снимающем задачи выявления специфики переработки информации человеком. Иначе говоря, информационное объяснение психики вовсе не исключает ее объяснения в иных аспектах.

Следует, однако, признать, что возражения против информационного подхода к психике, сформулированные О.К. Тихомировым, не столь просты, чтобы быть опровергнутыми приведенными выше со-

ображениями. Вопрос может быть поставлен так: в какой системе понятий сравнивать функционирование мозга и ЭВМ, что дает для психологии знание закономерностей переработки информации? Конечно, всякая психическая деятельность есть деятельность информационной, но не всякая информационная деятельность является деятельностью психической. Мозг и ЭВМ, моделирующая работу мозга, являются информационными системами – с позиции теории информации описываются общие для мозга и ЭВМ процессы. Кроме того, та и другая системы имеют какие-то специфические черты, отличающиеся друг от друга. Таким образом, психика есть информация плюс нечто, что и составляет специфику, свойственную лишь психическому. Если в предмет психологической науки включать лишь эту «специфику», то, естественно, изучение ее с позиции информационных процессов ничего не даст (вопрос, как видно, упирается в неразработанность самой проблемы предмета психологии). Поэтому с позиции психологии теория, например мышления, фиксирующая лишь те стороны мышления, которые являются общими для него и ЭВМ, действительно является существенно неполной. Но ведь никто не требует, например, от ножа быть одновременно и пилой. Точно также, вообще говоря, не имеет смысла требовать от теории, объясняющей одну сторону объекта, одновременно объяснять и другую сторону, находящуюся в «юрисдикции» другой теории. Аргументация сторонников необходимости включения в одну теорию объяснения и тех сторон, для объяснения которых существуют другие теории, применительно к обсуждаемой проблеме такова: игнорирование смысловой стороны при исследовании мышления означает уход от собственно психологического исследования, следовательно, исследование с позиций концепции информации не есть исследование психологическое.

Однако такое рассуждение справедливо лишь с позиций традиционной психологии, не включавшей в свой предмет информационный аспект психики. С позиции информационного подхода к психике предмет психологии должен быть уточнен на счет включения в него информационного аспекта психики.

«Специфичность» психики (предметная соотнесенность с внешним миром, субъективность, активность и т.д.) не может существовать «сама по себе», не основываясь на процессах переработки информации. Утверждать, что мышление является категорией психологической, а информация – категорией кибернетической, не совсем точно. Выделение специфического, свойственного лишь для человеческой психики, безусловно, важно и необходимо. Но важно и знание общего. Конечно, общие понятия отвлекается от многообразия структур конкретного типа, но они могут направлять поиски конкретных структур. Поэтому есть резон в предложениях определять психоло-

гию как науку о сохранении информации в памяти человека, о возможности обращения человека к информации, хранимой в его памяти, о воздействии этой информации на принимаемые человеком решения и на его поступки [6, с. 224].

Понять общее – не менее важная задача, чем понять специфическое, тем более, что знание общего открывает новые перспективы в понимании фундаментальных структур, лежащих в основе познавательной деятельности человека. При этом, разумеется, не следует забывать и различий, поскольку познать специфику психики без знания конкретных механизмов и структур невозможно. Как показал П.К. Анохин, в живой системе информационный процесс в любом его звене содержит в себе в разных кодах черты и признаки исходного объекта, поскольку передача информации подчиняется следующему правилу: между начальным и конечным звеном передачи информации должна быть «точная и адекватная эквивалентность». Вся приспособительная деятельность живых систем оказывается возможной потому, что все воздействия внешнего мира «входят в организм в форме тончайших информационных процессов, весьма точно отражающих основные параметры этого объективного внешнего мира» [1, с. 46].

Именно поэтому анализ психических функций мозга можно проводить с позиции концепции информации. Нервная система, как известно, осуществляет в организме функцию управления. Управление же без информации невозможно – процесс управления есть процесс информационный.

Свести полностью психическую деятельность к разнообразным электро-нейрофизиологическим процессам, имеющим, по существу, дискретный характер, конечно, нельзя. Психика синтезирует из дискретных сигналов (оптических, акустических, тактильных и т.д.) непрерывные образы внешнего мира. Это относится также к образам, созданным во «внутреннем мире» (картина в воображении художника, музыка в «голове» композитора и т.д.), т.е. субъективность психики не исчерпывается информационным ее аспектом. Субъективность дана в интроспекции, всякое описание субъективного означает его объективацию. С позиции концепции информации понять субъективное также трудно, как, например, понять творчество Ф.М. Достоевского, не привлекая в предмет исследования ничего, кроме словаря и грамматики русского языка – всегда будут исчезать уникальные моменты психической деятельности, для выражения которых (а не наоборот) только и существуют правила, знаки. Можно сказать, что психика есть информация плюс субъективное. Но это не означает, что в принципе невозможно расшифровать (хотя бы вероятно) коды даже самой сложной психической структуры – психика

вообще становится предметом изучения лишь в том объеме, в каком она выявлена в информационных процессах.

Однако специфика той частной формы информации, которая реализуется в психических актах, не может быть объяснена при полном абстрагировании от энергетических компонентов этих актов – для этого необходимо учитывать взаимосвязь информационных и энергетических аспектов психической деятельности. Ряд свойств психики, возможно, является по своей природе энергетическими (например, утомляемость или некоторые аномалии психической деятельности при перегрузках). Можно предположить, что эмоции управляют, прежде всего, энергетическими процессами психики, обеспечивая их активизацию или торможение. В целом для психической деятельности информационный аспект, видимо, доминирует над энергетическим, но последний является необходимым условием первого. Вполне возможно, что будут обнаружены новые универсальные принципы, увязывающие информационные и энергетические характеристики.

Смысл возражений против использования концепции информации в изучении психики сводится к тому, что нельзя использовать эту концепцию там, где вопрос о смысле, значении, ценности и истинности сигналов имеет первостепенное значение, т.е., иными словами, нельзя абсолютизировать универсализм этой концепции. Применение концепции информации в иных областях, по сравнению с теми, для нужд которых она была сформулирована, оказалось делом несравненно более трудным, чем простой перенос терминов концепции информации из одной области знания в другую.

Действительно, понятие информации в том смысле, какой был ему придан К. Шенноном (как количественная характеристика процессов управления и регуляции), оказалось недостаточным для понимания даже самого существа информационных процессов в живых системах, поскольку не учитывает их качественного аспекта. Живые системы оказались более «требовательными» к содержанию понятия информации, поскольку «количество информации», которым оперирует шенноновская теория информации, для изучения живых систем оказалось малопродуктивным – для них главное значение приобретает качество информации, ее полезность. Получая информацию из внешней среды, организм должен, прежде всего, оценить ее качественно, иначе организм не сможет приспособиться к внешней среде. В этих условиях возникает задача такого развития концепции информации, которое преодолело бы ограниченность его прежнего состояния.

Шенноновский вариант теории информации не учитывает смысла и ценности сигналов, несущих информацию, т.е. не включает в себя семантический и прагматический аспекты. Описывая количество информации на языке символов и статистических соотношений, она вы-

стует как синтаксическая теория. Для уточнения роли информационных процессов в психической деятельности имеет смысл сформулировать вопрос в следующем виде: какие изменения следует внести в определение понятия информации, чтобы при его помощи можно было бы понять те стороны психики, которые не находят объяснения в рамках шенноновской теории информации. Объяснение психической деятельности будет углубляться по мере уточнения понятия информации. В принципе же моделирование психики должно представлять собой не простое «отношение к ближайшему классу», не интерпретацию психики в терминах технической кибернетики, а разработку моделирования, позволяющего имитировать (не воспроизводить!) самые сложные психические функции, включая творческие.

В качестве аргументов против возможности моделирования творческих психических функций чаще всего выдвигаются следующие.

Аргумент первый: существование алгоритмически неразрешимых задач делает невозможным моделирование творческих функций – творческий акт включает в себя интуитивную стадию, которая принципиально недоступна моделированию. К такого рода задачам относятся: распознавание выводимости (А. Черч), установление тождества теории групп (П.С. Новиков), распознавание эквивалентности слов в любом исчислении (А.А. Марков и Э. Пост). Однако аргумент, основанный на существовании алгоритмически неразрешимых задач, уязвим. Дело в том, что человек способен у алгоритмически неразрешимых задач находить частные случаи. С.Л. Соболев и А.А. Ляпунов указывают на возможность решения такого рода задач достаточно мощным моделирующим устройством. В.М. Глушков обосновал возможность моделировать процесс решения любых задач, если этот процесс доступен описанию.

В конце 50-х гг. XX века в теории алгоритмов А.А. Ляпуновым, А.П. Ершовым, Ю.И. Яновым и др. [7] были проведены исследования, показавшие, что данная стадия решения задачи может быть записана в виде последовательности связанных между собой определенным образом операторов. Если затем отвлечься от конкретных операторов и заменить обозначающие их постоянные символы на переменные, то мы получим логическую схему алгоритма. Такая логическая схема может быть преобразована с помощью равносильных преобразований по определенным аксиомам в другую, эквивалентную ей схему. Таким образом, с помощью равносильных преобразований может быть получено множество различных логических схем одного и того же алгоритма и, следовательно, множество различных решений одной и той же задачи. Поэтому утверждение о том, что интуиция связана с какими-то особенностями психики, принципиально недоступных логическому анализу, оказывается необоснованным. Такой логический анализ проводится в теории логических схем про-

грамм, и в рамках этих исследований интуиция получает свое теоретическое объяснение.

Разработанный аппарат анализа логических схем алгоритмов и логических схем программ позволяет точно описывать процессы решения различных задач. Благодаря такому точному алгоритмическому описанию процессы решения задач могут быть представлены в виде программы для ЭВМ. Это позволяет передавать ЭВМ выполнение процессов, которые прежде связывались с интуицией – и здесь можно говорить о моделировании интуиции.

Аргумент второй: для моделирования процесса решения задачи его необходимо формализовать, а поскольку полная формализация невозможна, то решение не всех задач может быть осуществлено на ЭВМ. В самом деле, теорема Геделя показывает, что на основе формального исчисления не может быть изложено даже учение о целых числах. Если имеется достаточно мощная непротиворечивая система, то в ней при помощи математических средств, выражающих эту систему, можно сформулировать такие утверждения, которые в ее рамках нельзя ни доказать, ни опровергнуть, т.е. рамках данной системы эти утверждения считаются неразрешимыми. Иными словами, творческие мыслительные процессы не могут быть полностью формализованы. Существует также мнение, что теоремы неполноты и неразрешимости вообще выдвигают преграду исчерпывающему познанию любой системы, которая описывается аксиоматическим формальным языком.

Но действительно ли теорема Геделя является достаточным доказательством невозможности моделирования творческих задач? Существует обоснованное мнение о том, что ограничения, вытекающие из теоремы Геделя, относятся лишь к машинам Тьюринга, не получающим из внешней среды никакой информации. При условии, если машина получает информацию из внешней среды, являющейся бесконечной информационной системой, она оказывается способной решать неконструктивные проблемы, относительно которых можно было бы доказать их алгоритмическую неразрешимость. Иначе говоря, неразрешимость относится к абстрактному мышлению, а не к процессу познания в целом [3, с. 21]. Что же касается рассуждений о неформальности человеческого мышления, то не являются ли они преувеличенными? В.М. Глушков указывал на две причины этой кажущейся неформальности. Одна из них – малоизученность подсознательной деятельности, в следствие чего интуиция и представляется неформальной. Другая причина вытекает из факта взаимосвязи и взаимовлияния различных видов умственной деятельности [3, с. 19]. Если представить интуицию в виде логического процесса, то это позволит вскрыть закономерности, лежащие в основе интуитивного акта, и описать их, т.е. превратить в формальные. Интуитивная дея-



тельность, как и всякая другая, не может не быть управляемой какими-то естественными, познаваемыми закономерностями. Что же касается взаимовлияния отдельных видов духовной деятельности, то, действительно, творческий акт неотделим, например, от эмоциональной сферы, которая стимулирует процесс творчества. Чтобы промоделировать сложные творческие процессы, потребуется промоделировать, в частности, и эмоциональную сферу.

В связи с этим встает еще одна проблема: можно ли формализовать сам процесс перехода от одной системы к другой, более мощной. По-видимому, это возможно в определенных рамках, так как из теоремы Геделя вытекает не только неполнота той или иной системы, но также указание на то, что не охватывается данной системой. Формальная система, получив «толчок» для своего существования со стороны человека, приобретает известную самостоятельность. Но поскольку характер такой системы чисто формальный, то эта система не может определить, в каком направлении она должна быть расширена, если какая-либо задача, содержащаяся в ней, неразрешима. Для этого необходимо учитывать ее содержательное отношение к другой, более мощной системе. Если процесс творчества определить как такое «расширение», то он больше относится к области психологии, нежели логики, так как для подобного содержательного анализа необходим опыт субъекта познания, приобретенный им во взаимодействии с объектом. Следовательно, направление этого расширения в конечном счете определяет сам человек.

Аргумент третий: творческие функции мышления не могут быть описаны с помощью математических средств. Поскольку устройства, моделирующие мышление, имеют дело с формальными логическими отношениями, а для обширных областей человеческого мышления характерны связи, относящиеся к диалектической логике, то эти области не могут быть познаны с помощью математических методов. Однако дело заключается не в том, что это невозможно в принципе, а в том, что для этого недостаточно существующих математических средств.

Существующий математический аппарат, как известно, возник преимущественно для нужд физики и инженерной техники, где количество переменных и вариабельность незначительны по сравнению с биологическими и социальными системами. Действительно, уже сейчас можно указать на ряд проблем, для решения которых еще нет достаточных математических средств: разработка методов, позволяющих оперировать с большим числом взаимосвязанных переменных (А.И. Берг); разработка специального алгебраического аппарата, позволяющего работать с неопределенными (расплывчатыми) множествами в смысле Л. Заде; разработка методов оптимизации недифференцируемых и плохо формализованных функционалов (Н.Н.

Моисеев); разработка методов сравнения и оценки сложных алгоритмов, написанных на различных языках; разработка кодов, основанных на комбинированной пространственно-временной локализации данных (О. Шмитт); разработка аппарата, учитывающего время как фундаментальную характеристику управления в живом организме (В.В. Парин); разработка аппаратов, дающих возможность описать процесс построения динамических моделей действительности (В.Н. Пушкин).

В настоящее время можно говорить о том, что определились некоторые направления поиска в разработке нового математического аппарата: разработка средств, обеспечивающих переход с дискретного языка на непрерывный (Л. Маккей); разработка так называемой «логики спора» (О. Шмитт, П. Лоренцен); «биологики» (Г. Ферстер); «серой логики» (О. Шмитт); «психологики» (Ж. Пиаже) и др. Представляется, что пессимизм, основывающийся на утверждении о возможности моделирования мышления лишь некоторых элементарных областях мыслительной деятельности, будет рассеиваться в зависимости от успехов выращивания новых глав математики.

Аргумент четвертый: творческий процесс предполагает постановку цели и интерпретацию достигнутого результата, машины же в принципе не могут обладать такой способностью. Как известно, процесс решения всякой задачи можно разбить на несколько этапов. Первый из них – постановка задачи. Второй этап – решение задачи. Для этого используются известные ранее методы решения подобных задач, в случае если их недостаточно, осуществляется детерминированный или случайный поиск недостающего для решения задачи элемента. Третий этап – интерпретация полученного результата. Данный аргумент основывается на утверждении, что можно моделировать лишь второй этап, который считается нетворческим.

Однако данный аргумент весьма уязвим, поскольку его принятие означает невозможность моделировать и второй этап решения задачи, поскольку он включает в себя такие элементы творчества, как разработка критериев поиска и критериев отбора. Конечно, ЭВМ сама по себе, по своей внутренней потребности не может ни поставить задачу, ни определить критерии поиска и отбора, что свойственно лишь человеку. Имитировать же эти способности ЭВМ в состоянии. Доказательство теорем – деятельность творческая. Ван Хао, как известно, удалось ее промоделировать, причем были доказаны заново не только теоремы, содержащиеся в «Принципах математики» Уайтхеда и Рассела, но и ряд новых теорем. Известна программа В.М. Глушкова по проверке доказательств теорем алгебры, программа для доказательства или опровержения теории на основе алгоритма А. Тарского. Известны также успешные шаги в области моделирования некоторых творческих задач, связанных с сочинением музыки, стихов, определения авторства про-

изведений и т.д. Число таких задач постоянно расширяется. При этом утверждения о том, что ЭВМ не дает нового знания, не могут считаться справедливыми. Если считать новым то, что не было предусмотрено конструктором и что ему не было известно, то можно сказать, что ЭВМ действительно выдает нечто новое. В этом отношении показательное использование систем искусственного интеллекта для доказательства теоремы, известной как теорема четырех красок (ТЧК).

ТЧК была сформулирована Фрэнсисом Гутри в 1852 году. Им было высказано предположение, что любую географическую карту, изображенную на листе бумаги, можно раскрасить в четыре цвета таким образом, чтобы страны, имеющие общую границу, были раскрашены в разные цвета. Последовавшие за этим попытки математиков обосновать эту догадку более столетия оказывались безуспешными. Вместе с тем развитие математики породило у многих математиков конца XIX века уверенность в том, что на любой вопрос, сформулированный на языке математики, может быть дан ответ, если будут использованы достаточно мощные математические идеи, и что любой компетентный математик может проверить правильность решения задачи в разумный промежуток времени. Однако полученные в 30-е годы XX столетия К. Геделем и А. Черчем результаты свидетельствовали, что существуют утверждения, истинность или ложность которых не может быть доказана в рамках данной системы, и что в системе могут быть теоремы, доказательство которых не может быть записано в разумный промежуток времени. Эти выводы были использованы одними математиками для заключения о том, что ТЧК не может быть ни доказана, ни опровергнута, а другими – для заключения о невозможности записать доказательство, даже если оно существует, в разумный период времени.

Часть математиков не теряла надежды найти решение ТЧК. Результаты, полученные А. Кемпе и П. Хивудом в XIX веке, а также Д. Биркгофом и Ф. Франклином в XX веке, послужили тем фундаментом, который позволил Г. Хишу формализовать известные методы доказательства сводимости конфигураций и показать, что, по крайней мере, один из них (прямое обобщение метода, использованного А. Кемпе) представляет собой процедуру, которую может осуществить ЭВМ [10]. Теория сводимых конфигураций, развитая Г. Хишем, к концу 80- годов XX века позволила уяснить необходимые для доказательства ТЧК идеи сводимости. Однако на пути доказательства ТЧК стояли еще два препятствия: недостаточный для доказательства ТЧК объем памяти существующих ЭВМ и неясность того, каков должен быть верхний предел необходимости для доказательства ТЧК сводимых конфигураций. Решающий шаг был сделан К. Аппелем и В. Хакеном.

Вначале К. Аппелем и В. Хакеном было найдено множество сводимых конфигураций, «размер кольца которых был достаточно мал,

чтобы машинное время на доказательство сводимости было в пределах разумного» [10, р. 115]. Это позволило к концу 1972 года составить машинную программу для выполнения специального типа разряжающей процедуры. Прогонки этой программы позволили ее модифицировать при сохранении основной структуры. Это позволило на основе новой разряжающей процедуры построить множество сводимых конфигураций, в результате чего в конце 1976 года ТЧК была доказана (на что потребовалось 1200 часов машинного времени трех ЭВМ).

Таким образом, оказалось, что существуют такие математические утверждения, проверка которых требует объема вычислений и затрат времени, превышающих возможности не только отдельного человека, но и большого коллектива. Поэтому использование ЭВМ для решения подобных задач не только возможно, но и необходимо. Известно, что в традиционных доказательствах могут быть опущены некоторые моменты по причине их очевидности. Допускаются также ссылки на работы, в которых можно найти то, что пропущено в доказательстве. В том и другом случае подразумевается, что пропущенное может быть при желании восполнено. Использование же ЭВМ в математическом доказательстве создает ситуацию, в которой такой очевидной возможности нет.

Многие математики исходят из того, что доказательство должно обладать такими характеристиками, как убедительность, обзорность и формальность. Убедительность доказательства – это свидетельство понимания математики как человеческой деятельности. Обзорность – свидетельство возможности проверить (обозреть) его во всей полноте человеком, т.е. обзорность конкретизирует убедительность, связывая процесс доказательства с его субъектом. Традиционно источником убедительности доказательства считается ясность, т.е. возможность проверить его квалифицированным математиком без использования ЭВМ, хотя некоторые доказательства могут быть весьма длинными, и потребовать много сил и времени. Формальность доказательства означает представимость его в виде конечной последовательности формальной теории, удовлетворяющей некоторым условиям, т.е. это вывод заключения из аксиом теории с помощью правил логики. Формализованность доказательства, в свою очередь, конкретизирует обзорность, разбивая ее на конечные обзорные модели.

Когда в доказательстве присутствуют обзорность и формальность, это решающий аргумент в пользу признания доказательства математиками. Большинство их полагает, что все обзорные доказательства могут быть формализованными, хотя интуитивисты отрицают возможность замены актуальных конструкций математических доказательств формальными системами. Кроме того, теорема Геделя свидетельствует о том, что ни одна теория не в состоянии формализовать каждое доказательство. В таком случае ясно, что обзорное

доказательство можно формализовать в более сильной формальной теории, но в последней будут содержаться новые обозримые доказательства, которые в ее рамках не могут быть формализованы.

Таким образом, не существует системы, в которой любое доказательство может быть формализовано, поэтому формализованность называют локальной характеристикой доказательств, а не глобальной. Но, как было показано Р. Томом, для любого доказательства существует некоторая подходящая формальная система, в которой оно может быть формализовано [11]. Однако все ли формализованные системы обозримы? Существуют формализованные доказательства, которые не могут быть обозримы вследствие ограниченности человеческой жизни. Поэтому возможны формализованные утверждения без обозримого доказательства. На практике же математики обычно приходят к знанию формального доказательства через посредство обозримых доказательств. Убедительность и обозримость могут быть связаны как сопряжения целого и части. Если убедительность – это в известной мере осуществимость доказательства как целого, завершенного, но в котором особо выделены исходный и заключающий его пункты, то обозримость – это осуществимость доказательства в каждом пункте сцепления доказательства. Формализованность же – упрощающая процедура, делающая доказательство более лаконичным и универсально выраженным, а также делающая его доступным для задания ЭВМ.

В какой же мере указанные характеристики применимы к доказательству ТЧК с помощью ЭВМ? Прежде всего, является ли оно убедительным? К. Аппель и В. Хакен подметили любопытную особенность, заключающуюся в том, что признание убедительности доказательства ТЧК с помощью ЭВМ зависит от характера полученного математического образования: математики, получившие образование до появления ЭВМ, к признанию убедительности данного доказательства относятся, как правило, оппозиционно [10].

Является ли доказательство ТЧК обозримым? Отрицание возможности обозримости основывается на том, что ни один математик не в состоянии проверить его шаг за шагом. По мнению некоторых исследователей, доказательство ТЧК может быть признано обозримым лишь в том случае, если применение ЭВМ считать новым методом доказательства. Но тогда само понятие доказательства должно измениться [11]. Отсюда делается вывод, что обозримость сохраняется за традиционным доказательством, но не доказательством с помощью ЭВМ.

Наконец, является ли доказательство ТЧК формализованным. Оно убедительно, формализовано, но необозримо. Существенное различие между доказательством ТЧК и традиционным заключается в том, что первое требует обращения к ЭВМ для заполнения бреши в доказательстве, которое в остальных отношениях традиционно [11]. Иными словами, доказательство ТЧК осуществлено на основаниях,

часто являющихся эмпирическими. Поэтому применение ЭВМ в математике вводит эмпирические эксперименты в математику. Доказательство с помощью ЭВМ не является традиционным, т.е. это не априорный вывод утверждения из посылок, поскольку используются данные эксперимента. А это делает доказательство ТЧК первым доказательством *a posteriori*, что снова поднимает проблему отличия математики от естественных наук.

Таким образом, принятие тезиса о том, что положения теоретической математики могут быть установлены с помощью эмпирического доказательства, вводит, по мнению ряда авторов, в математику эмпирические методы, что имеет весьма серьезные последствия для философии математики, в частности влечет необходимость пересмотра или уточнения ряда положений, таких как: 1. все математические теоремы известны *a priori*; 2. математика (в противоположность естественным наукам) не имеет эмпирического содержания; 3. математика полагается только на доказательства, тогда как в естественных науках ставится эксперимент; 4. математические теоремы определены в такой степени, которой не может соответствовать ни одна теорема естественных наук [11].

Вопрос, следовательно, упирается в надежность работы ЭВМ. Правильность результатов математического эксперимента в рамках той или иной теории оказывается зависимым от того, насколько правильно ЭВМ осуществляет свои операции. А это определяется правильностью конструкции ЭВМ, работы ее элементов, воздействием внешних условий и т.д., т.е. в конечном счете – правильностью определенных положений естественных наук и технических расчетов. А это означает, что зависимость математики от естественных и технических наук становится все более важной. По словам Н. Бурбаки, математики всегда были уверены, что они доказывают «истины» или «истинные высказывания» [2]. Вопрос, таким образом, в том, аналогичны ли эти «истины» тем результатам, которые получены с помощью ЭВМ. Если принять, что использование ЭВМ в математических доказательствах обосновано и что ТЧК доказана, то последняя должна быть признана математической истиной. А это с необходимостью ведет к переоценке роли формального доказательства в философии математики. Этот вывод, конечно, не ставит под сомнение все то, что связано со статусом формальной теории как ветви математической логики. Однако, начиная с доказательства ТЧК, формальные доказательства оказываются не единственной исследовательской программой в математике. Но все дело в том, готовы ли математики признать математикой результаты, подобные ТЧК, а метод доказательства ТЧК включить в методологию математики.

Таким образом, включение ЭВМ в процедуру математического доказательства вызывает изменение нормативов этой процедуры,

ломку традиционных канонов философии математики. Осознание этого факта обуславливает возникновение нового представления о математической истине в плане ее увязывания с возрастанием степени вероятности утверждений математики. Различие ситуации с включением в процедуру доказательства ЭВМ и ситуации с включением других технических средств типа арифмометра является принципиальным, поскольку в первом случае нет возможности проверить доказательств непосредственно человеку. Использование ЭВМ в процессе доказательства приводит к тому, что происходит расширение средств доказательства: к чисто человеческому фактору добавляется машинный фактор, привносящий в толкование математической истины определенную долю вероятности, зависящую от уровня развития ЭВМ и степени контроля их работы.

Итак, можно признать, что моделирование творческих функций мышления возможно. Более того, возможно моделирование процессов решения таких задач, решение которых традиционными методами невозможно. И суть совсем не в том, что «машина начинает мыслить», поскольку речь идет не о воспроизведении мыслительных актов, а об их имитации. Но и при таком понимании систем искусственного интеллекта их возможности ставят перед человеком сложные философские проблемы. С развитием искусственного интеллекта были сняты возражения против возможности моделирования многих творческих функций человеческой психики. О возможностях ЭВМ будущих поколений мы можем только догадываться. Здесь могут быть прорывы, меняющие наши представления о возможностях познания.

#### ЛИТЕРАТУРА

1. *Анохин П.К.* Теория отражения в современной науке о мозге. - М., 1970.
2. *Бурбаки Н.* Очерки по истории математики. - М., 1963.
3. *Глушков В.В.* Кибернетика и умственный труд. - М., 1965.
4. *Дубровский Д.И.* Психические явления и мозг. - М., 1971.
5. *Кочергин А.Н.* Философские вопросы моделирования функций мозга. - Новосибирск, 1973.
6. *Ляпунов А.А.* О рассмотрении биологии с позиции изучения живой природы как системы // Проблемы методологии системного исследования. - М., 1970
7. Проблемы кибернетики. Вып. 1. - М., 1958.
8. *Тихомиров О.К.* Структура мыслительной деятельности человека. М., 1969.
9. Rosenblueth and Wiener N. The Role of Models in Science // Philosophy of Science. 1945, vol. 12.
10. *Appel K., Haken W.* The solution of the colour map problem // Sci. Amer. 1977, vol. CXXXVII, N 10.
11. *Тымoczko T.* The four-colour problem and its philosophical significance // J. Philos. 1979, vol. LXXVI, N 2.

---

# ВОЗМОЖНОСТИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: МОЖНО ЛИ ПРОЙТИ ТЕСТЫ ТЬЮРИНГА?

*А.Ю. Алексеев*

## Введение

Рассмотрим мысленный пример *«Девушка по переписке»*:

Молодой человек ищет спутницу жизни и, как принято, обращается к интернетовской службе знакомств. Задаёт параметры запроса – возраст, рост, объёмы, цвет волос, знак зодиака, привычки и пр. В ответ – ряд кандидаток. Начинает переписываться. Особенно нравится девушка Элиза. Переписка длится более полгода – в выборе нельзя ошибиться! Беседы ведутся на самые разные темы: семьи, детей, секса, быта, отдыха, зарплаты, экономики, политики, литературы, китайской кухни и пр. Играют в шахматы с переменным успехом. Девушка поражает объёмом знаний, лёгкостью в общении, изысканностью стиля. И глаза у неё выразительные. Наконец молодой человек не выдерживает и назначает встречу на площади у памятника. Приезжает с букетом роз. Ждёт. Час, два, три. Никто не приходит. Вернувшись ни с чем, пишет гневное письмо. В ответ – «Прости, милый! Я – Электронная Лиза, компьютерная программа»<sup>1</sup>.

## 1. Проблема тестирования компьютеризированной реальности

Молодому человеку не повезло – но он не убедился в адекватности своих представлений действительному положению дел. В шутку замечено, что в Интернете выдавать себя за человека может не только программа, но и животное, например, натренированная собака. Питер Стейнер в знаменитой карикатуре *«В Интернете никто не знает, что ты – собака!»* (1993 г.) [22] изобразил пса, сидящего перед компьютером и нажимающего на клавиши. *«И в самом деле!»* – восклицает Вильям Рапопорт по поводу этой карикатуры [24, Р.469]:

---

<sup>1</sup> На выбор имени «девушки» повлияло название широко известной программы понимания естественного языка, - ELIZA, разработанной Джозефом Вейзенбаумом в 1966 г. [30]. Программа состояла всего из двух сотен операторов на языке Бейсик и незатейливого алгоритма «понимания» задаваемых ей вопросов. При этом ELIZA настолько умело имитировала речевые способности собеседника, что, к примеру, секретарша Вайзенбаума при отсутствии шефа очень любила с нею «поболтать». В основе концепции программы лежала психотерапевтическая техника диалога с пациентом. Сегодня программы, у которых в основе диалога лежит такая же концепция, принято называть «ELIZ'a- подобные программы» [30, Р.504-506].



«В Интернете никто не знает, кто с кем на самом деле общается. Это неведение может привести к самым серьёзным последствиям в реальной жизни. Если я ввожу номер кредитной карточки через Web-сайт с целью покупки книги, то не известно, куплю ли я эту книгу или выдам свой секретный код мошеннику. Компьютерная безопасность определяет способ социального благополучия и защиты личности, и даже жизнь или смерть. Если моя дочь общается посредством чата с лицами противоположного пола примерно её возраста, то кто знает, может быть, она общается с серийным убийцей. Обычно, правда, мы предполагаем такие опасности и полагаем, что общаемся не с мошенниками, убийцами, собаками. Или – как в нашем случае, с компьютерными программами».

Сегодня утверждения В. Рапопорта (2000 г.) несколько устарели на фоне экспоненциального развития информационно-коммуникационных средств, сопровождаемого, в свою очередь, громкими судебными делами по поводу преступлений, «соучастниками» которых выступают компьютеры. По всей видимости в недалёком будущем воплотится и сценарий «Девушка по переписке». Тогда актуальнейшей проблемой станет поиск методов и средств различения между человеком и компьютерной системой. Подобная проблема имеет давнюю традицию.

## 2. Тест Декарта

Считается (см., например, [13, 19]), что впервые на проблемы различения способностей человека и машины обратил внимание Р. Декарт (см.: «Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках. Часть V» [3, С. 250-296.]). Он считал невозможным различение между нечеловеческим животным и машиной, если она ни внешним видом ни внутренним строением не будет отличаться от животного. Так как человек обладает разумом, в отличие от других существ, то даже если машина будет иметь полнейшее сходство с нашим телом и поведением, различение возможно. Для этого следует протестировать машину: 1) на её способность пользоваться языком так же, как пользуемся мы, сообщая другим свои мысли – она провалит тест, так как нельзя представить, что она должным порядком будет располагать слова; 2) на её способность работать в самых различных контекстах – тест также не будет пройден, так как только человеческому разуму под силу охватить всё многообразие всевозможных поведенческих ситуаций, даже если в некоторых частных случаях машина может действовать лучше нас.

Изменилось бы мнение Р. Декарта о возможности пройти такой тест машиной, если бы он сегодня познакомился с программами

ИИ в среде Интернет?<sup>2</sup> Следует, конечно, учесть уровень технических достижений XVII в. [19]. В контексте же научно-технических успехов XX в. подобный тест был предложен Аланом М. Тьюрингом (1912-1954 гг.) в статье 1950 г. «Вычислительные машины и интеллект» [28], ставшей *самой цитируемой* в литературе по философии ИИ [8, Р.391]. В отличие от Р. Декарта, А. Тьюринг положительно решает вопрос построения мыслящих машин. Достаточным условием интеллектуальности системы он считает возможность компьютера «подражать» человеку в условиях игры в имитацию.

### 3. Игра в имитацию и канонический Тест Тьюринга [28]<sup>3</sup>

А. Тьюринг заменяет вечный вопрос «Могут ли машины мыслить?» (в силу его бессмысленности по причине неопределённости терминов) на «игру в имитацию» (по сути, лингвистическую игру<sup>4</sup>), которая состоит из трёх сценариев (обозначим их ИИ1 – ИИ3)<sup>5</sup> и последующих модификаций (ТТ1-ТТ2).

**ИИ1.** Играют три человека – мужчина (М), женщина (Ж) и бесполоый судья (С)<sup>6</sup>. Игроки располагаются в разделённых непроницаемой перегородкой комнатах, в одной – С, в другой – М и Ж. Диалог между ними ведётся средствами телетайпа. Цель С – определить

---

<sup>2</sup> См., например, бот «АЛИСА» <http://alicebot.org/>. При входе на сайт начинает говорить девушка.

<sup>3</sup> В данной работе нет возможности осуществить подробный анализ Теста Тьюринга. Оригинал [28] можно скопировать с сайта А. Тьюринга <http://www.turing.org.uk/turing/>. (Главный редактор сайта – широко известный биограф А. Тьюринга – Эндрю Ходгис). На русском языке перевод работы появился в 1960 г. (под редакцией Б.В. Бирюкова), вызвал широкие дискуссии, включая конференции Всесоюзного масштаба. Затем ТТ был забыт и долгое время книга была библиографической редкостью. За последние пять лет книга много раз переиздавалась различными издательствами, поэтому будем считать, что читатель знаком с этой работой, классической для философии ИИ.

<sup>4</sup> Неизвестно, был ли А. Тьюринг знаком с концепцией языковых игр Л. Витгенштейна. Игра в имитацию появилась незадолго до появления английского перевода «Философских исследований» Л. Витгенштейна в 1953 г. В этих играх можно отметить много общих моментов, например, отказ от определения понятий, поиск значения слов в поведенческих сценариях и др.

<sup>5</sup> Здесь следует возразить Дж. Лассажу [20], который указывает всего на два сценария игры в имитацию. Если следовать буквальной трактовке статьи, к чему призывает Дж. Лассаж, то можно достаточно утвердительно сказать, что Тьюринг описывает две фигуры в п.п. 1 своей работы, а третья фигура введена в конце п.п. 5 при описании универсальной машины.

<sup>6</sup> Даются обозначения, отличные от оригинала в целях лучшего восприятия. Также слово «Interrogator» было переведено «судьей» вместо принятого в [1] громоздкого перевода «кто-нибудь задающий вопрос»

посредством вопросов пол игроков. Вначале игры М и Ж – неизвестные X и Y. По условиям игры Ж говорит правду и только правду, помогая судье, а М может говорить и правду и ложь, стремясь обмануть С путем подражания Ж. Игра заканчивается, когда С утверждает одно из двух: «X – это М, а Y – это Ж» либо «X – это Ж, а Y – это М».

**ИИ2.** Место М занимает компьютер К – универсальная цифровая вычислительная машина, на которой запущена программа типа анализа и синтеза связного текста. Тьюринг задаётся вопросом: «Будет ли С ошибаться столь же часто, как и в игре, где участниками являются только люди?». Этот вопрос и заменяет исходный – «Может ли машина мыслить?». Если судья будет ошибаться столь часто, как и в ИИ1, то Тьюринг считает, что К демонстрирует интеллектуальные способности.

**ИИ3.** Играют К и М. Мужчина М заменяет Ж. Цель С остаётся прежней – определить пол игрока. В такие игры, по мнению Тьюринга, можно будет играть в 2000 г. При наличии памяти  $10^9$  бит, достаточной скорости операций и соответствующей программы, у С будет не более 70 % правильной идентификации игроков (кто из них – М, а кто – Ж), т.е. К сможет обманывать судью, как минимум, в 30% случаев.

**ТТ1:** Играют человек Ч и компьютер К. Преследуются цели: для К – имитация Ч, для Ч – доказательство человечности, для С – решение вопроса: «X – это Ч, а Y – это К» либо «X – это К, а Y – это Ч».

**ТТ2:** Играет только К, цель которого обмануть С путем подражания человеку. Цель С состоит в различении механического и человеческого интеллекта (кто с ним играет – Ч или К?).

Собственно варианты ТТ1 и ТТ2 в настоящее время принято называть **Тестом Тьюринга** (ТТ). Само выражение «Тест Тьюринга» появилось в середине 1970-х годов. Об этом свидетельствует Дж. Лассаж [20] на основании работы Д. Мичи [11, Р. 65]. Далее он утверждает, что А. Тьюринг стал наиболее известным в среде философов «компьютерной ориентации» за то, что никогда и не предлагал. В самом деле, при буквальном прочтении статьи слово «тест» используется всего четыре раза, три из которых приходится на случай (особо интересный для нас), связанный с анализом возможности идентификации «другого сознания». <sup>7</sup> Здесь вопросы, которые задаёт судья, отождествляются с тестом. <sup>8</sup>

<sup>7</sup> Дж. Лассаж ошибается, утверждая, что Тьюринг употребил слово «тест» два раза. На самом деле «тест» употребляется трижды в п.п. 6.4 при опровержении «Аргумента от “сознания”» и в п.п. 6.9 при опровержении «телепатического» аргумента.

<sup>8</sup> Помимо этого, исходя из этических правил научной деятельности, не принято называть собственное творение своим именем. Стоит вспомнить, что А. Тьюринг также не назвал своим именем и знаменитую «машину Тью-

Прекрасная метафора интеллектуальной машины привлекает большое внимание исследователей, служит концептуальным каркасом ИИ, является генератором новых идей, различных интерпретаций и толкований. Исследуются самые разные параметры теста: 1) длина – интервал времени, в течение которого осуществляется тестирование; 2) содержание тестовых вопросов; 3) «стена Тьюринга» – параметры скрытости игроков от судьи; 4) характеристики и особенности игроков и др. На многообразие новых трактовок повлияла также неопределённость игры в имитацию, «чехарда» с ролями мужчины, женщины, машины – почему нельзя было сразу поставить К против С?) [25].

На сегодняшний день насчитывается более тысячи крупных статей и специальных работ, в которых осмысливается ТТ. Поэтому стоит говорить о **тестах Тьюринга**, а не об одном единственном каноническом тесте, который, в свою очередь, достаточно неявно был определён самим Тьюрингом.

Рассмотрим ряд наиболее крупных модификаций ТТ. Вначале обратим внимание на те интерпретации ТТ, которые утверждают, что ТТ по замыслу не является тестом на интеллект.

#### 4. Неинтеллектуальные интерпретации Теста Тьюринга

4.1. «Параноидальный» ТТ [25, Р.33]. Предложен К. Колби в работе «Искусственная паранойя» (1971 г.). Используется схема ТТ1. В роли С – врачи-психиатры, в роли Ч – параноики. Опыт тестирования показал, что почти в половине случаев психиатры не могли отличить ответы компьютера от ответов больных. На мой взгляд, успех К в прохождении данного теста обусловлен тем, что ответы непадают – это правило психического отклонения, для реализации этого теста не требуется сложная техника, вполне приемлема лингвистическая недостаточность ТТ. По всей видимости, параноидальный тест могут успешно пройти программы генерации «постмодернистских стихотворений».

4.2. Социальный ТТ [25, Р.31-32]. ТТ служит для оценки социального интеллекта. В роли С выступает мать-природа (которую «не проведёшь!»), в роли игроков – сообщество людей, которые пройдут тест, если выживут. Тест интересен в концептуальном отношении – в плане идентификации и тестирования общественных форм сознания и «интеллектуальности» социальных систем.

4.3. Гендерный ТТ [25, Р.32-33]. ТТ служит решению вопроса – «Чем мужчина отличается от женщины»? Имеются различные трактовки: 1) Тьюринг выражает склонность к транссексуальности<sup>9</sup>; 2) Тьюринг – ан-

---

ринга» (предложенная им дескрипция интуитивного понятия алгоритма, опубликованная в 1936 г. получила собственное наименование в журнале «Journal of Symbolic Logic» в 1937 г. [20]).

<sup>9</sup> Тьюринг был обвинен в гомосексуализме, не выдержал тяжести обвинения и покончил самоубийством в 1954 г.

тифеминист [16], для него важно – «Может ли женщина мыслить?»); 5) Напротив, Тьюринг считает подражание женщине верхом совершенства [6; Р. 3-6]; 5) Тьюринг просто шутил, так как в нём гармонично сочетались ум, юмор и секс (А. Хотгис [18]). С. Яновская (1960 г.) [1, С. 6] считает, что гендерный ТТ принципиально невозможен – судья будет правильно отличать женщину от мужчины в 50 %.

### **5. Тест леди Лавлейс (ТЛ) – можно ли пройти тест на творческие способности?**

С. Брингсфорд, П. Белло, Д. Феруччи в работе «Творчество, Тест Тьюринга и (улучшенный) Тест Лавлейс» [27] продолжают возражение леди Лавлейс, предложенное А. Тьюрингом [28, п.п. 6.6]. Используется схема ТТ2. Тест задаётся следующим образом. Искусственный агент К, разработанный человеком Р, проходит ТЛ тогда и только тогда, когда: 1) К создаёт на выходе некий артефакт о (для леди Лавлейс – дочери Байрона – в качестве о выступала бы, по всей видимости, поэма); 2) о – это не результаты случайного стечения обстоятельств или сбоя машины. К всегда может воспроизвести (повторить) о; 3) Р не может объяснить, как К выдал о даже в условиях полного представления о К – структуре баз данных, алгоритмах функционирования и пр. Авторы считают, что: 1) ТЛ невозможно пройти путем обмана; 2) Все существующие модели ИИ не пройдут ТЛ, так как разработчики знают их особенности, а «от себя» такие программы прибавить ничего не могут; 3) Необходимым условием прохождения ТЛ является способность К самостоятельно модифицировать свою программу, т.е. К должен быть самообучаемым. Однако это условие не достаточно для прохождения ТЛ, так как правила самообучения не являются по сути оригинальными – они формируются в контексте «известного» для Р; 4) Вариант наделения К «свободой выбора» путём моделирования, например, мотивационно-волевых механизмов творчества также не подходит, так как артефакт о будет представлен в некотором пространстве решений, что исключает его оригинальность.

Как известно, Тьюринг в ответ на возражение Лавлейс заменяет выражение «способность к творчеству» на «способность удивлять». Он считает, что любая ошибка в программном коде или аппаратный сбой «могут удивить». В ТЛ это исключается, так как: 1) условие «ошибки» оговорено в определении ТЛ и запрещено; 2) к К применяются жесткие условия контроля вычислительного процесса даже в случае сбоя. Тем не менее, ТЛ не убедителен в силу неопределённости «творческой способности» и требований к артефакту. Его можно пройти.

### **6. Тест Френча (ТФ) – можно ли пройти тест на субкогнитивные способности?**

Тест предложен Р. Френчем в 1990 г. в работе «Субкогнитивные способности и границы Теста Тьюринга» [31]. Р. Френч считает, что

ТТ умело обошёл бескрайнее философское болото проблемы дух/тело, однако он не способен исследовать глубинные, подсознательные области человеческого интеллекта. Поэтому в ТТ следует включить так называемые субкогнитивные вопросы, т.е. любые вопросы, позволяющие идентифицировать низкоуровневую когнитивную структуру, например, подсознательную сеть ассоциаций (а эти вопросы, считает Р. Френч, неявно включает любой ТТ). Сеть ассоциаций состоит из нечётного количества статистически связанных друг с другом образов, которые воспринимались человеком в прошлом и которые могут быть активированы в любой момент. Какие именно ассоциативные ансамбли возникнут при некотором вопросе – зависит от социокультурной обусловленности и форм конкретной жизнедеятельности игрока; т.е. ТФ способен пройти лишь тот, кто живет и ощущает окружающий мир так же, как и люди. ТФ – это не тест на интеллект «вообще», а тест на социокультурно-обусловленный человеческий интеллект.

Р. Френч предлагает ряд способов идентификации подсознательной сети ассоциаций, среди которых наиболее показательным, на наш взгляд, представляется *метод ассоциативных заучиваний*. Суть метода в следующем. У человека в ходе конкретной жизнедеятельности развиваются определенные ассоциативные связи между понятиями. Сила этих связей – величина варьируемая и её можно оценивать временем, в течение которого происходит связывание одного понятия с другим (с понятием «хлеб» быстрее свяжется понятие «масло», нежели «собака»). Судья С может воспользоваться этим методом следующим образом: 1) за день до тестирования проводится контрольный опрос среди интервьюируемых (людей), подбирается произвольный набор слов и фиксируется статистически среднее время связывания понятий. 2) На следующий день при проведении теста судья задаёт кандидатам (среди которых Ч и К) тоже самое задание, собирает результаты тестирования и сравнивает с результатами контрольного опроса. Анализ позволяет достаточно чётко определить, кто из кандидатов – К, а кто – Ч. Человеком будет тот, чьи результаты окажутся более схожими с результатами контрольного опроса. Компьютер К будет постоянно проваливать ТФ, так как невозможно запрограммировать некий априорный способ ассоциативной связи понятий. Есть только единственный способ различения машиной всей данной совокупности ассоциативных связей между понятиями – это окунуться в поток конкретной жизнедеятельности людей. Однако ТФ представляется неубедительным в силу ряда причин: 1) сомнительна возможность выявления значимых статистических зависимостей между понятиями на этапе контрольного опроса; 2) разработчик Р может воспользоваться различного рода словарями, где учитывается лексическая сочетаемость слов, например [5].

## 7. Тест Блока (ТБ) – можно ли пройти антибихевиористский тест?

Н. Блок в работе «Психологизм и бихевиоризм» (1981) [9] критикует канонический ТТ, показывая, что «игра в имитацию» является разновидностью оценки поведения в контексте лингвистического бихевиоризма. Так как предметом бихевиористского анализа является не ментальная структура, а объективно фиксируемые параметры поведения (реакции), определяемые внешними воздействиями (стимулами) (а в *лингвистическом бихевиоризме* стимулы и реакции – суть вербальные выражения), то здесь имеется ряд стандартных антибихевиористских возражений на возможность осуществления адекватной оценки: 1) Аргумент Чизголма и Гича, согласно которому невозможно выделить конкретное психическое состояние в поведенческой диспозиции без учёта *всей совокупности иных психических состояний*. Анализ каждой реакции может продолжаться бесконечно в континууме рассмотрения внутренних процессов и результаты данного анализа невозможно зафиксировать в схеме «стимул-реакция»; 2) Аргумент «совершенный актер» Х. Патнэма, согласно которому *различные* классы психических явлений могут продуцировать *одинаковые* поведенческие диспозиции и, наоборот. Например, в силу принятых в обществе законов «сверх-супер-спартанцы» не будут выказывать боль, даже если на самом деле люди её испытывают. Напротив, человек, не испытывающий боли, может притворяться, что ему больно; 3) Аргумент «паралитики» и «мозги в бочке»: эти «пациенты» вообще не выказывают определенных признаков, означающих, что они испытывают боль.

В свою очередь, интеллектуальное поведение, как правило, продуцируется комбинацией: {интеллект + обычная склонность к мышлению}. Однако невозможна комбинация {отсутствие интеллекта + желание обмануть, что интеллект имеется}.

Отсюда вытекает: 1) нужно изменить ТТ, так как он рушится под натиском антибихевиористских аргументов; 2) следует искать иное определение ТТ.

Предлагается новая формулировка ТТ (нео-ТТ), или тест Блока, который означает способность продуцирования осмысленной последовательности вербальных реакций на некоторую последовательность вербальных стимулов, но не акт продуцирования. Новая постановка усиливает ТТ, ТБ пройти намного сложнее, так как снимаются вышеперечисленные аргументы: 1) можно представить много факторов того, что интеллектуальная система *не будет расположена* к выдаче *осмысленных* ответов, однако эти факторы не скажутся на её *способности*; 2) невозможно притворяться разумным и не быть таковым; 3) у «паралитиков» и «мозгов в бочке» имеется *способность*

реагировать осмысленно, однако им недостаточно средств для реализации.

Для прохождения ТБ Н. Блок предлагает машину, которая способна производить осмысленную последовательность вербальных реакций на вербальные стимулы [9, Р. 9-13]. Поведение машины полностью задаётся человеческим коллективом – проектировщиками, инженерами, программистами и др., которые упорно трудятся, применяют свой интеллект для реализации всевозможных последовательностей осмысленных ответов на предполагаемые последовательности вербальных стимулов. На поведенческом уровне данная машина выглядит интеллектуальной. Однако знание её внутреннего устройства убеждает в полном отсутствии у неё «интеллекта» - *интеллект, «проявляемый» ею как в форме актуального поведения, так и в форме диспозиции интеллектуального поведения - суть интеллект разработчиков.*

Н. Блок, по сути, подсказал возможность прохождения ТБ – надо включить в сценарий имитационной игры разработчика Р. Тогда на любой вопрос со стороны С может быть получен ответ со стороны К при содействии Р. Единственный способ для С различить К от Ч (и тем самым дискредитировать компьютерную программу) можно назвать «анатомическим» - т.е. использовать принцип «вскрытие покажет». Только тогда программа не сможет пройти ТБ.

### **8. Тест Серля (ТС): можно ли пройти тест на «понимание»?**

В 1980 г. Дж. Серль предлагает знаменитый мысленный эксперимент «Китайская комната» в работе «Разумы, мозги, программы» [26] Здесь отрицается концепция сильного ИИ, согласно которой при прохождении ТТ компьютер не просто моделирует человеческую способность к пониманию (это характерно для позиции слабого ИИ), но и: 1) на самом деле *понимает* текст; 2) компьютер и её программа объясняют человеческую способность «понимания».

Сценарий Дж. Серля разыгрывается по схеме ТТ1, но теперь роль Ч и К совмещает в одном лице сам Серль-в-комнате (Searle-in-the-room) – мысленный эксперимент осуществляется с позиции 1-го лица. Судей двое – китаец и англичанин. Когда задаёт вопросы судья-китаец, то Серль-в-комнате, играя роль К и выполняя чисто синтаксические преобразования над китайскими иероглифами (он снабжен инструкциями перевода на родном английском языке), выдаёт китайцы ответы, которые тот воспринимает как осмысленные. Когда вопросы задаёт судья-англичанин, Серль-в-комнате играет роль Ч, непосредственно понимая задаваемые вопросы и не выполняя никаких синтаксических преобразований. Отличие между первым контуром игры (когда понимания нет) и вторым контуром игры (когда понимание очевидно для Серля-в-комнате) заключено, по мнению Серля, в интенциональности, так как человек – это биологическое существо, способное ощущать, действовать, понимать, обучаться и т.д. Ин-



тенциональные же феномены (в частности, квалиа) не редуцируемы к формальной схеме. Формальная модель по причине отсутствия самостоятельных казуальных сил не самодостаточна для воспроизводства феномена интенциональности. Интенциональность, «имеющаяся» у К, на самом деле имеется лишь в головах у Р - кто их программирует, и у С - кто интерпретирует ответы.

Условия прохождения ТС по сути воспроизводят условия прохождения ТБ, так как здесь также концептуально задействован Р. То есть пройти ТС можно, ничего принципиально нового относительно ТБ здесь нет.

### **9. Тест Ватта (ТВ) – можно ли пройти инвертированный ТТ?**

Стюарт Ватт в работе «Наивная психология и инвертированный Тест Тьюринга» [29] считает, что успех прохождения ТТ связан с естественной, наивно-психологической установкой приписывать ментальные свойства любым системам – другим людям, животным, природным явлениям, в том числе и компьютерам. Чтобы избежать этого, предлагается *инвертированная* версия ТТ, то есть следует «обратить» роли игроков. Судья должен быть судим, так как именно он приписывает интеллект системе, которая таковым может и не обладать. То есть теперь тестированию подлежит судья. Для этого схема ТТ2 модифицируется – вместо одного судьи теперь двое – судья-компьютер (С-К) и судья-человек (С-Ч). Тестирование осуществляется как прежде. Но теперь С-Ч (в роли его должен выступать опытный эксперт) должен сличать ответы от игрока К или Ч с показаниями С-К. Если С-К начинает приписывать интеллект тестируемой системе, то это является поводом для индуктивного заключения, о том, что С-К – машина.

По поводу невозможности организации «инвертированного» тестирования существует ряд убедительных мнений: 1) тесту присуща *проблема дурной бесконечности*, так как для каждого С-Ч, судящего С-К, должен стоять другой С-Ч', который судит первого С-Ч и т.д. (Р. Френч [15]); 2) тесту присуща *проблема параллелизма*, так как невозможно обнаружить момент, когда С-Ч начинает различать между собой и С-К (С. Брингсфорд [10]). Таким образом вопрос о возможности прохождения ТВ должен решаться после вопроса о самой возможности тестирования.

### **10. Тесты Харнада (ТХ) – как пройти ТТ, если «Стена Тьюринга» разрушена?**

Многоуровневую классификацию тестов Тьюринга предлагает С. Харнад в работе «Разумы, машины и Тьюринг: неразличимость неразличимостей» [17] (2001 г.). Выделяются следующие уровни:

ТХ0<sup>10</sup>: Это – уровень «игрушечных» ТТ – не полноправных тестов, а лишь некоторых фрагментов, ограниченных как по длине, так и по содержанию. Такие тесты не отвечают исходному замыслу А. Тьюринга. Однако все попытки моделирования интеллекта, известные на сегодняшний день, выше данного уровня не поднялись.

ТХ2 – общепринятое понимание ТТ. Именно его и имел в виду А. Тьюринг. Иногда ТТ данного уровня называют тестом «друг по переписке». Длина ТХ2 равна протяжённости человеческой жизни С. Программа «Элиза», из примера «Девушка по переписке» принадлежит к этому классу программ.

ТХ3 – это так называемая «роботизированная» версия ТХ2. Имеется возможность манипуляции предметами внешнего мира. Процедура идентификации систем, которые проходят данный тест, требует реализации принципа «Вскрытие покажет».

ТХ4 – это компьютерные системы, неотличимые как в плане ТХ3-неотличимости, так и в плане микрофизической организации системы. Здесь имеет место «тотальная неотличимость» компьютерной системы от человека, включая мельчайшие внутренние нюансы телесного строения. На данном уровне можно разместить так называемый «Тест Тьюринга с мозгом» («Brainy-Turing-Test»), где в дополнение к поведенческой неотличимости требуется неотличимость на уровне нервной системы и предполагается использование коннекционистских моделей [14].

ТХ5. На этом уровне происходит отбор ТХ4-систем на предмет соответствия Единой Теории Реальности, которая способна, если она вообще возможна, объединить все возможные знания о природе, обществе, человеке. Данный уровень С. Харнад ввёл, по всей видимости, для того, чтобы допустить возможность истинности нескольких отличающихся друг от друга конкретных теоретических описаний одной и той же реальности (т.е. ТХ4-систем). Истинность в данном случае имеет различные метафизические предпосылки у создателей различных теорий.

В соответствии с классификацией С. Харнада, современный уровень достижений ИИ не превысил «игрового» уровня ТХ0. Однако то, что прохождение ТТ возможно, пускай даже на этом уровне, демонстрируют факты. Рассмотрим эти факты.

## 11. Факты прохождения Теста Тьюринга

В январе 2000 г. в Dartmouth College состоялась юбилейная конференция «Будущее Теста Тьюринга: следующие пятьдесят лет» теоретическими задачами которой явились оценка влияния ТТ на современ-

---

<sup>10</sup> Мы вводим свои символы для обозначения уровней, отличных от харнадовских, с целью достижения единообразия нашей классификации ТТ.

ность и определение прогнозов программирования ТТ на будущее. Практическая часть конференции состояла в розыгрыше сценария ТТ1. Участвовали шесть программ. В целом результаты не подтвердили предсказаний Тьюринга – судьи правильно отличали компьютер от человека в 91% после 5 минут и в 93% после 15 минут соревнований [8]. Тем не менее, в настоящий момент в сфере разработки ТТ-программ сформировалась своеобразная индустрия, в основном, игрового характера (уровень ТХ0). С 1991 г. введена специальная премия, соразмерная нобелевской – Лойбнеровская премия, которую учредил Хью Лойбнер (Hugh Loebner). Премия в 100000\$ присуждается за выигрыш, 2000\$ - за участие. Установлена длина теста – 3 часа. Тестирование проводится по сценарию ТТ1. На вопросы десяти отвечают пять К и пять Ч. Пока ещё нет победителя, которому можно было бы присудить полную премию [33]<sup>11</sup>. Однако не следует отождествлять «игру в имитацию», которая прослеживается на состязаниях Лойбнера с ТТ. Тест Тьюринга следует характеризовать как теоретический план. Например, под определение ТТ подпадает любая экспертная система, функционирующая, например, в узко-профессиональной сфере деятельности. Такие системы *могут* проходить ТТ (формулируемый в процессе задания запросов к этой системе). Исходя из широкого толкования ТТ следует более скрупулёзно разобраться с «возможностью» прохождения ТТ.

## **12. Физическая, логическая и метафорическая возможности прохождения Теста Тьюринга**

В контексте проблемы сознания, философии искусственного интеллекта и везде, где используются мысленные эксперименты к слову «возможность» следует относиться осторожно [19]. Здесь, конечно, неприемлемы жаркие диалектические дебаты на тему «возможности – вероятности – действительности», с характерными для них суждениями типа «возможность есть будущая действительность», но при этом «возможность уже существует в действительности – именно как возможность».

Для позитивного решения вопросов о возможности прохождения ТТ следует ввести ряд ограничений: 1) судья – обычный человек (как молодой человек из примера «Девушка по переписке»), а не некий абсолютный и всезнающий субъект; 2) длина теста ограничена; 3) границы теста заданы некоторой фиксированной предметной областью, например, сферой профессиональной деятельности судьи; 4) разработчик программы – эксперт в этой области; 5) стена Тьюринга непроницаема, например, диалог ведётся посредством электронной почты.

---

<sup>11</sup> См. обзор программирования ТТ в [6, С.49-53]

Выделим следующие виды «возможности» прохождения компьютерной системой тестов Тьюринга: физическую, логическую и метафорическую.

**Физическая возможность** – всё то, что согласуется с формальными условиями опыта, например, наглядная демонстрация или конкретный факт. Действительным в таком случае будет всё то, что согласуется с материальными условиями опыта – ощущением (по И. Канту). Обычно исследователи апеллируют к неким всеобщим и предельным формам доказательства физической невозможности, начиная от общеизвестного примера с вероятностью написания обезьяной сонета Шекспира. При изучении проблематики ТТ можно выделить следующие интересные примеры доказательства физической невозможности:

1) **Комбинаторный взрыв** (Дж. Миллер). Существует порядка  $10^{30}$  грамматических предложений длиной в 20 слов. Если произвольно допустить, что  $10^{15}$  также семантически корректны, то для прохождения ТТ длительностью в час может понадобиться порядка 100 таких предложений. А это  $10^{1500}$  строк, число, превышающее количество элементарных частиц во вселенной [9].

2) «Прыжок на Луну» (Франк Типлер). Машина Н. Блока (выше мы видели, что данная машина реализует ТТ силами разработчиков) физически невозможна, так как для её реализации (если информационные процессы осуществлять «вручную»), потребуется больше энергии, чем та, которая требуется человеку для прыжка на Луну [13, п.п.4.2].

Такие доказательства в условиях наших ограничений и допущений не убедительны. Мы имеем реальные факты прохождения программами тестов Тьюринга (это показывают лойбнеровские состязания) и, следовательно, наглядную демонстрацию возможности прохождения ТТ. Отсюда можно заключить о *физической возможности прохождения ТТ*.

**Логическая возможность** – всё то, что можно ясно помыслить и чётко описать (так, по крайней мере, логическую возможность определяет Д. Чалмерс в аргументе мыслимости зомби [19]). Логическая возможность или возможность помыслить некоторую ситуацию или положение дел обычно противопоставляется воображимости, которая, в свою очередь, обуславливает физическую возможность. Например, невозможно вообразить или представить бесконечное число. Однако его можно непротиворечиво помыслить, например, в системе арифметики Пеано. Если мысленный эксперимент с прохождением программой ТТ логически корректен, все компоненты его – отчётливые мысли, связанные стандартными логическими принципами, тогда ТТ логически можно пройти. В рамках традиционной логики логическая невозможность прохождения ТТ апеллирует к противоречиям, кото-

рые возникают в процессе самоописания и самореференции с характерными для него вопросами, на которых невозможно получить ответов (unanswerable question). Логическая невозможность ТТ подмечена А. Тьюрингом в ответе на «математическое возражение» [28, п.п. 2.3]). Он коротко ответил, указывая на то, что людям так же при-суще впадать в логические противоречия. Столь быстрый ответ показался неудовлетворительным для Лукаса и Пенроуза, возникла полемика по поводу логической возможности ТТ, получившая название «Аргумент Гёделя». Приложение теоремы Гёделя применительно к проблематике ТТ кратко изложено в [13]: 1) Пусть К – цифровой компьютер; 2) Так как К подпадает под аргументацию Гёделя (точнее, под возражение Лукаса-Пенроуза), то имеется вопрос q для К, на который нет ответа; 3) Если некоторое E не является предметом возражения Лукаса-Пенроуза, то для E на любой вопрос имеется ответ. 4) Человеческий интеллект нельзя подвести под возражение Лукаса-Пенроуза; 5) Таким образом, нет «безответных» вопросов для человеческого интеллекта; 6) Поэтому вопрос q не является «безответным» для человеческого интеллекта; 7) Задавая вопрос q, человек (судья) может определить, кто перед ним – компьютер (то есть ответа он не получит и тестируемая система «зависнет») или человек; 8) Таким образом компьютер К не пройдет ТТ. Добавим к этому: 9) логически невозможно, чтобы К прошёл ТТ.

На наш взгляд, очевидна правота Тьюринга – если человека попросить разрешить софизм, то он впадет в противоречие, для него вопрос q также будет невозможен и судья не отличит компьютер от человека. Широко известен подход к преодолению софизмов, предложенный Б. Расселом. Он предлагает вводить целый ряд ограничений на условия суждения – в первую очередь, устанавливать иерархию языков, типизировать язык [7, С.65-66]. В принципе, большинство интеллектуальных систем построено на основе выделения типов сущностей в предметной области. При этом, несомненно, система ИИ должна обладать адаптивными способностями к перестройке своих формализованных «знаний» - к их самонастройке, самоорганизации, самообучению (см., например, у Дж. Маккарти в [3]). Но здесь имеется в виду не адаптация «вообще», а приспособление ИИ-системы к классам конкретных состояний внешней и внутренней среды системы, которые, конечно, должен предусмотреть разработчик. Поэтому не будет логической ошибкой предположить наличие у разработчика достаточных средств – мета-языковых для конкретной системы ИИ, способных обеспечить пере-программирование системы с целью работы в тех предметных областях, для которых у неё недостаточно средств к самообучению, самоописанию и т.п. Здесь также предполагается рассматривать не пред-

метную область «вообще», а как предлагал Рассел, концептуально её ограничивать.

Таким образом, в условиях допущений и ограничений на ТТ, представляется *логически возможным* его прохождение.

**Метафорическая возможность.** Рассмотрим ещё одну разновидность возможности прохождения ТТ – метафорическую. Метафорически возможно всё то, что является убедительным благодаря применению метафоры. Общепринято считать, что ИИ – это метафора естественного интеллекта. Более того, *ценой неадекватного принятия (вернее, отторжения) метафоры ИИ в среде отечественных учёных явилось отставание нашей науки в столь важнейшей для развития страны сфере научно-технического знания* (Д.А. Поспелов [2, С.5-7]). Для многих исследователей ИИ, которые принимают метафору ИИ, характерно игнорирование более фундаментального вопроса – а что именно значит «метафора». Так, например, В. Рапопорт, опровергая аргумент «Китайской комнаты» Дж. Серла, призывает к метафорическому пониманию слов, применяемых в сфере исследований ИИ [51, Р. 471]. Он вполне убедительно раскрывает, в частности, что выражение «полет самолёта» - это метафорическое употребление слова «полет птицы», и, по аналогии, «искусственный интеллект» - метафора «интеллекта». Далее он достаточно отчётливо разъясняет преимущества для разработки искусственных агентов своей концепции «семантической семантики», согласно которой состав «знаний» агента должен включать разделяемые синтаксические структуры. Разделение происходит, как минимум, на две составляющие, одна синтаксическая структура начинает играть роль семантики для другой структуры. Однако далее он совершенно не убедительно полагает, что самореференция такой двумерной структуры (т.е. отображение синтаксической структуры на самое себя, приводящее к двuasпектному, синтаксическо-семантическому образованию) может послужить метафорой «понимания», например, понимания Серлем-в-комнате китайского языка. Принятие понятия «метафоры» без какого-либо предварительного анализа в дальнейшем вызывает ряд трудностей в понимании того, каким образом самореференцирующая синтактико-семантическая структура может послужить *метафорой* для «интенциональности», «понимания», «ментального свойства» - всего того, чем владел Серль-в-комнате, говоря об этих понятиях от первого лица.

Подобного рода зыбкость суждений характерна во многом и для других исследователей ИИ, призывающих к метафорическому, расширенному значению слова «интеллект». Поиск обоснованности понятия «искусственный интеллект» и, следовательно, всех иных понятий, с которыми оно соотносится, требует применения лингво-философского анализа самого слова «метафора».

Воспользуемся с этой целью теорией метафоры Д. Дэвидсона [4]. Согласно ей, метафоры – это: не грезы, не «сон языка», не расширенное или удвоенное значение, не средство передачи необычных идей, не способ избежать неопределённости понятия, не сравнение, не неявное когнитивное содержание, передаваемое автором для того, чтобы получатель уловил его с целью понимания сообщения. Метафора не объяснима путем обращения к ее скрытому содержанию. Интерпретации метафоры недопустимы, а содержание невыразимо.

Для чего же тогда нужна метафора, если она не способна отражать истинное или ложное положение дел? Метафора, как полагает Д. Дэвидсон, пригодна для серьезного научного или философского разговора; целиком принадлежит сфере *употребления* (но не определения понятий); привлекает внимание к тому, что автор хочет сообщить и к новому и неожиданному сходству между предметами; заставляет нас заметить то, что иначе могло бы остаться незамеченным. Метафора связана с образным использованием слов и предложений и всецело зависит от *обычного* или *буквального* значения слов (и состоящих из них предложений). Метафора подобна *речевым актам*: утверждению, намеку, лжи, обещанию, выражению недовольства и др. Но она служит не для передачи содержания понятия, а для убеждения в истинности данного понятия. Метафора *не несёт какое-то содержание или имеет какое-то значение, кроме, конечно, буквального*.

В свете данной теории метафорическая возможность того или иного положения дел означает убедительность для социокультурного сообщества метафоры, которая, в частности, обеспечивает единообразное понимание слов, не подлежащих верификации или фальсификации (к таким словам относится любой ментальный термин, например, «интеллект»).

Так как метафора «означает только то (или не более того), что означают входящие в них слова, взятые в своем буквальном значении» [4], то использование метафоры предполагает апелляцию к здравому смыслу. Убедительность в том или ином положении дел осуществляется путем демонстрации, состоящей из очевидных для сообщества вещей, событий и процессов.

В контексте данной теории становится возможным исследование метафорической возможности прохождения ТТ. Если метафора, предназначенная для понимания устройства системы, проходящей ТТ будет для нас убедительна, то тогда имеется метафорическая возможность прохождения ТТ.

### 13. Неубедительность метафоры «человек-компьютер» А. Тьюринга

Для объяснения возможности создать систему, которая способна играть в игру имитации, в статье 1950 г. А. Тьюринг предлагает метафору «человек-компьютер» (human computer). Она, в свою очередь, является производной от метафоры «машина Тьюринга» (1936 г.), убеждающей в том, что устройство с конечным числом состояний, с лентой памяти и записывающе-считывающей головкой эквивалентно понятию алгоритма. Метафора алгоритма была принята научным сообществом и оказалась достаточно убедительной (см., например, отзыв С. Яновской в «Предисловии к русскому изданию» перевода статьи Тьюринга [1, С.3-18]). Однако если интуитивное понятие алгоритма хорошо описывалось машиной Тьюринга, то ещё более интуитивное понятие «интеллект», задаваемое игрой в имитацию (т.е. тестом Тьюринга) для С. Яновской показалось не убедительным. Она считает, что физически невозможно отличить М от Ж путём тестирования, поэтому программа не может иметь «не менее 30% шансов обмануть человека при продолжительности игры в пять минут», так как заведомо гендерный тест имеет шансы 50%. Логическая невозможность (и, соответственно, неубедительность) С. Яновская объясняла противоречием, которое возникает в сценарии ИИ2: вопрос о возможности запрограммировать мышление Ж должен быть решён с помощью «игры в имитацию», так как с помощью её и определяется то, чем является «мышление». Как видим, С. Яновская воспринимала игру в имитацию не как метафору, а как буквальное руководство к определению условий определения «интеллекта».

На наш взгляд, несостоятельность ТТ проистекает из-за неубедительности собственно внутреннего строения метафоры, неочевидности такого строения. Метафора «Человек-компьютер» 1950 г. состоит из машины Тьюринга (МТ) и «Книги правил», которую человек-компьютер «берёт» для выполнения тех или иных заданий. В Книге правил содержатся руководства для действий МТ. Возникает вопрос: как возможен обмен информацией, которая заложена в «Книге правил» с лентой МТ? Как правила преобразуются в символы на ленте МТ? Этот вопрос Тьюринг пытается решить путем «семейной аналогии» [28, п.п.4]. Книга правил заменяется запиской, которую пишет женщина своему сыну. В записке представлены правила, которым должен следовать мальчик и которые представляют «интеллект» матери, передаваемый сыну. Однако при этом мальчик уже действует в соответствии с правилами, записанными на ленту МТ. Опять встаёт вопрос – как возможно соотнесение в рамках данной схемы информации из «записки» (т.е. «книги правил» человека-компьютера) на ленту МТ. Как мальчик «вживит» символы из «записки» в свою



«ленту» и что они будут значить при их считывании и переходе МТ из одного состояния в другое?

Таким образом, если запись символов на ленту МТ и считывание их с неё достаточно очевидный процесс для нас, а не только для шифровальщика, в роли которого некоторое время служил А. Тьюринг (для Декарта, однако, такая метафора была бы не очевидной), то совершенно не ясно, каким образом происходит считывание-запись «знаний», представленных в «Книге правил» на ленту МТ.

Метафора «человек-компьютер» *не убедительна. Поэтому ТТ метафорически невозможен даже в тех условиях и ограничениях на условия проведения тестирования, в которых и физически и логически можно пройти ТТ.* Следовательно, не убедительны и модели искусственного интеллекта, разрабатывающиеся в контексте существующей парадигмы компьютерной технологии и современной (фоннеймановской) архитектуры компьютерных систем.

Выделенная нами метафорическая невозможность прохождения ТТ не означает необходимости сворачивания проблематики ТТ. Напротив, нужен поиск новых, более убедительных метафор интеллекта – ведь «дряхлый, слепой компьютер с достаточно искусной программой, способной обмануть тест Тьюринга, есть научная фантастика наихудшего сорта» (Д. Деннет) [12].

Следует намечать и новые перспективы применения ТТ, разрабатывать новые модификации ТТ, которые позволят значительно расширить существующие и открыть неизведанные концептуальные и конструктивные горизонты в изучении интеллекта, ментальных свойств, сознания. Крайне важен в этой связи призыв Д.И. Дубровского к изучению возможности применения ТТ для *диагностики субъективной реальности* – целостного осознаваемого психического состояния, способного включать различные составляющие: ощущения, чувственные образы, эмоциональные переживания, мыслительные процессы, волевые побуждения и т.д. [32]

Такой тест Тьюринга, если он окажется возможен, по всей видимости, помог бы разобраться молодому человеку с «девушкой по пеписке».

## ЛИТЕРАТУРА

1. Алан М. Тьюринг. Может ли машина мыслить? (Под ред. Б.В. Бирюкова). М., 1960. С. 19-158
2. Будущее искусственного интеллекта. – М. Наука, 1991. – 302 с.
3. Декарт Р. Сочинения: В 2 т. – Т. I. – М., 1989;  
<http://www.philosophy.ru/library/descartes/method/05.html>
4. Дэвидсон Д. Что означают метафоры // Теория метафоры. М., 1990. С. 172-193. Пер. М.А. Дмитривской, <http://kant.narod.ru/davidson.htm>
5. Мельчук И.А. Русский язык в модели «Смысл-Текст» - Москва-Вена: Школа «Языки русской культуры», 1995. – 682 с.
6. Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2006 г. Под ред. А.Ю. Алексеева – М.: МИЭМ, ИИнтелЛП, 2005. – 208 с.
7. Рассел Бертран. Исследование истины и значения. Пер. Ледникова Е.Е., Никифорова А.Л. – М.: Идея-Пресс, Дом интеллектуальной книги, 1999. – 440 с.
8. Akman V., Blackburn P. Editorial: Alan Turing and Artificial Intelligence. Journal of Logic, Language, and Information 9: 391–395, 2000;  
<http://www.cs.bilkent.edu.tr/~akman/jour-papers/jolli/jolli2000.pdf>
9. Block N., (1981). Psychologism and Behaviorism, Philosophical Review 90, pp. 5–43
10. Bringsjord S., (1996). The Inverted Turing Test is Provably Redundant. Psychology 7(29).
11. D. Michie. On Machine Intelligence, New York: Halsted Press, 1974
12. Dennett, D. C., (1984). Can machines think? In (M. Shafiq, ed) How We Know. Harper & Row; <http://www.kurzweilai.net/articles/art0099.html?printable=1>
13. Dowe D., Oppy G., (2003). The Turing Test. Stanford Encyclopedia of Philosophy; <http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=turing-test>
14. Flanagan O., Polger T., (1995). Zombies and the function of consciousness <http://homepages.uc.edu/~polgertw/Polger-ZombiesJCS.pdf>
15. French R. M., (1995). Refocusing the Debate on the Turing Test: A Response. Behavior and Philosophy 23, pp. 59–60;  
<http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
16. Genova, J., (1994). Turing's Sexual Guessing Game, Social Epistemology 8(4), pp. 313–326.
17. Harnad, S., (2001). Minds, Machines and Turing: The Indistinguishability of Indistinguishables. Journal of Logic, Language, and Information.  
<http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/Harnad00.turing.html>
18. Hodges A., (2002). Alan Turing; <http://www.plato.stanford.edu/entries/turing>
19. Kirk R. (2004). Zombies. Stanford Encyclopedia of Philosophy;  
<http://plato.stanford.edu/entries/zombies/>
20. Lassègue Jean, (1996). What Kind of Turing Test Did Turing Have in Mind? Tekhnema 3 / "A Touch of Memory" / Spring 1996
21. McCarthy J., (1995). Artificial Intelligence and Philosophy;  
<http://cogprints.ecs.soton.ac.uk/archive/00000420>

22. On the Internet, Nobody Knows You're a Dog. Page 61 of July 5, 1993 issue of The New Yorker, (Vol.69 (LXIX) no. 20); <http://www.unc.edu/depts/jomc/academics/dri/idog.html>
23. *Platt C.*, (1995). What's It Mean To Be Human, Anyway?, Wired. <http://www.wired.com/wired/archive/3.04/turing.html>
24. *Rapoport William J.*, (2000). How to Pass a Turing Test/ Journal of Logic, Language, and Information 9: 467-490, 2000. 467, <http://www.cse.buffalo.edu/~rapoport/Papers/TURING.pdf>
25. *Saygin A.P., Cicekli, I., Akman V.*, (2000). Turing test: 50 years later. Minds and Machines 10:463-518; <http://www.cs.bilkent.edu.tr/~akman/jour-papers/mam/tt50.ps>
26. *Searle J. R.*, (1980). Minds, Brains, and Programs. (<http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>)
27. *Selmer Bringsjord, Paul Bello, David Ferrucci.* Creativity, the Turing Test and the (Better) Lovelace Test. <http://www.rpi.edu/~faheyj2/SB/SELPAP/DARTMOUTH/tt3.pdf>
28. *Turing A.*, (1950). «Computing Machinery and Intelligence». Mind 59, 236, 433-460.
29. *Watt Stuart*, (1996). Naive Psychology and the Inverted Turing Test, *Psychology*: 7, #14
30. *Weizenbaum J.*, (1996). 'ELIZA—A Computer Program for the Study of Natural Language Communication Between Men and Machines', *Communications of the ACM* 9, pp. 36–45.
31. *French R.* (1995), The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It, *Psychology* 7(39); <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
32. *Дубровский Д.И.* Тест Тьюринга, парадигма функционализма и проблема сознания // В кн. «Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции “Философия искусственного интеллекта”», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева» – М.: МИЭМ, ИИнтелЛИ, 2006. – 192 с. – С. 104-109

---

# IV

## Проблемы моделирования

---

---

### ЗАДАЧА МОДЕЛИРОВАНИЯ КОГНИТИВНОЙ ЭВОЛЮЦИИ\*

*В.Г. Редько*

Аргументируется, что моделирование когнитивной эволюции может служить естественнонаучной основой разработок искусственного интеллекта (ИИ). Характеризуются философские предпосылки исследований когнитивной эволюции. Анализируется направление исследований Адаптивное Поведение (Анимат-подход к ИИ), модели которого могут рассматриваться, как задел исследований когнитивной эволюции. Кратко характеризуется оригинальный проект «Мозг Анимата». В порядке обсуждения предлагается программа будущих исследований когнитивной эволюции.

#### **1. Философские предпосылки исследований когнитивной эволюции – пролегомены к теории происхождения мышления**

##### **1.1. О термине «пролегомены»**

Когда И. Кант написал свою знаменитую «Критику чистого разума» (1781 г.) [1], то он ожидал, что с ним будут спорить, его будут критиковать, ругать или, наоборот, будут соглашаться с ним. Но ожидаемого им отклика не было: не было ни возражений, ни одобрений, а все жаловались на трудность понимания этого сочинения. И тогда Кант написал небольшую брошюру «Пролегомены ко всякой будущей метафизике, могущей появиться как наука» (1783 г.) [2], в которой популярно растолковал свою позицию. Термин пролегомены

---

\* Работа выполнена при финансовой поддержке РАН (Программа "Интеллектуальные компьютерные системы", проект 2-45) и РФФИ (проект 04-01-00179).

означает предварительные рассуждения, введение в изучение. Данный раздел как раз хорошо соответствует термину «пролегомены».

## 1.2. Сверхзадача

Основные проблемы, на которых хотелось бы заострить внимание, – это проблема применимости человеческого мышления в научном познании и связанная с ней проблема эволюционного происхождения мышления. Здесь мы не будем особенно акцентировать внимание на термине «мышление». Примерно с равным успехом можно было бы использовать вместо термина «мышление» термины «интеллект», «логика», «логика человеческого мышления», можно было бы пытаться уточнить эти термины. Но можно пока воспринимать эти термины интуитивно, причем для определенности остановимся на термине «мышление».

Итак, в чем проблема?

Наука – это познание природы. Но способен ли человек познавать законы природы? рассмотрим физику, наиболее фундаментальную из естественнонаучных дисциплин. Мощь физики связана с эффективным применением математики. Но математик строит свои теории независимо от внешнего мира (в тиши кабинета и т.п.). Почему же результаты, получаемые математиком, применимы к реальной природе?

Можно ли конструктивно подойти к решению этих вопросов? Скорее всего, да. По крайней мере, можно попытаться это сделать. Рассмотрим одно из элементарных правил, которое использует математик в логических заключениях, правило *modus ponens*: «если имеет место  $A$ , и из  $A$  следует  $B$ , то имеет место  $B$ », или  $\{A, A \rightarrow B\} \Rightarrow B$ .

А теперь перейдем от математика к собаке И.П. Павлова. Пусть у собаки вырабатывают условный рефлекс, в результате в памяти собаки формируется связь «за УС должен последовать БС» (УС – условный стимул, БС – безусловный стимул). И когда после выработки рефлекса собаке предъявляют УС, то она, «помня» о хранящейся в ее памяти «записи»: УС  $\rightarrow$  БС, делает элементарный «вывод»  $\{УС, УС \rightarrow БС\} \Rightarrow БС$ . И у собаки, ожидающей БС (скажем, кусок мяса), начинают течь слюнки.

Конечно, применение правила *modus ponens* (чисто дедуктивное) математиком и индуктивный «вывод», который делает собака, явно различаются. Но можем мы ли думать об эволюционных корнях логических правил, используемых в математике? Да, вполне можем – умозаключение математика и индуктивный «вывод» собаки качественно аналогичны.

Мы можем пойти и дальше – можем представить, что в памяти собаки есть *семантическая сеть*, сеть связей между понятиями, образами. Например, мы можем представить, что у собаки есть понятия

«пища», «опасность», «другая собака». С понятием «пища» могут быть связаны понятия «мясо», «косточка». При выработке пищевого условного рефлекса, например, на звонок (скажем, УС = «звонок», БС = «мясо») у собаки, по-видимому, формируется простая семантическая связь: «за звонком следует мясо».

Можно далее представить процессы формирования разнообразных семантических сетей в процессе жизни собаки и накопления ей жизненного опыта. Такие семантические сети, формируемые в памяти собаки, по-видимому, аналогичны семантическим сетям, исследуемым разработчиками искусственного интеллекта [3].

Итак, мы можем думать над эволюционными корнями логики, мышления, интеллекта. И более того, было бы очень интересно попытаться строить модели эволюционного происхождения мышления. По-видимому, наиболее четкий путь такого исследования – построение математических и компьютерных моделей «интеллектуальных изобретений» биологической эволюции, таких как безусловный рефлекс, привыкание, классический условный рефлекс, инструментальный условный рефлекс, цепи рефлексов, ..., логика [4]. То есть, целесообразно с помощью моделей представить общую картину эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека.

Естественно, что такие исследования – это огромный фронт работы, и задачу построения теории происхождения мышления, задачу моделирования когнитивной эволюции можно рассматривать как сверхзадачу. Исследования этой проблемы могли бы обеспечить определенное обоснование применимости нашего мышления в научном познании, т. е. укрепить фундамент всего величественного здания науки. Чтобы вести эту работу серьезно, целесообразно идти именно по пути построения математических и компьютерных моделей когнитивной эволюции.

Но прежде чем строить модели, давайте посмотрим, кто еще думал над близкими вопросами. Проследим цепочку: Д. Юм -> И. Кант -> К. Лоренц.

### **1.3. Д. Юм -> И. Кант -> К. Лоренц**

В «Исследовании о человеческом познании» (1748) Давид Юм подверг сомнению понятие причинной связи [5]. А именно, он задался вопросом: почему когда мы видим, что за одним явлением *A* постоянно следует другое *B*, то мы приходим к выводу, что *A* является причиной *B*? Например, когда мы многократно наблюдаем, что Солнце освещает камень, и камень нагревается, то мы говорим, что солнечный свет есть причина нагревания камня.

Фактически Юм задался вопросом: что нас заставляет делать выводы о происходящих в природе явлениях? Что лежит в основе этих

выводов? Юм попытался понять, откуда мы берем основание заключать, что *A* есть причина *B*. Он посмотрел на этот вопрос, как он пишет, со всех сторон и не нашел никакого другого основания, кроме некоторого внутреннего чувства привычки.

Юм взглянул на наш познавательный процесс со стороны, извне. Он как бы вышел на некий мета-уровень рассмотрения наших собственных познавательных процессов и задался вопросом о том, откуда взялись эти познавательные процессы и почему они работают.

Острота сомнений Юма была в том, что он задался вопросом о принципиальной способности человека познавать мир.

**Остроту** сомнений Юма очень хорошо почувствовал Иммануил Кант. Но Кант также видел мощь и силу современной ему науки. Тогда уже была глубокая, серьезная и развитая математика, мощная ньютоновская физика, дающая **картину мира**, позволяющая объяснить множество явлений на основе немногих четких предположений, использующая многозвенные и сильные математические дедуктивные выводы. Что было делать Канту? Подвергнуть сомнению все эти познавательные процессы? И развивая сомнения Юма дальше, отвергнуть всю науку? Ведь на самом деле – драма!!!

Конечно же, Кант, как научно образованный человек, не стал отвергать современную ему науку, а постарался разобраться, как же работают познавательные процессы. В результате появились знаменитая «Критика чистого разума» [1] и ее популярная интерпретация – «Пролегомены ко всякой будущей метафизике, могущей появиться, как наука» [2]. Кант провел исследование познавательных процессов в определенном приближении – приближении фиксированного мышления взрослого человека. Он не задавался вопросом, **откуда** берутся познавательные способности, он просто констатировал факт, что они существуют, и исследовал, **как** они работают. В результате этого анализа Кант пришел к выводу, что существует система категорий, концепций, логических правил и методов вывода (таких как заключенная относительно причинных связей между событиями), которые используются в познании природы. Эта система «чистого разума» имеет априорный характер – она существует в нашем сознании прежде всякого опыта – и является основой научного познания природы.

Естественно, что приближение фиксированного мышления человека наложило свой отпечаток: Кант утверждает – и вполне логично (!) – что так как «чистый разум» априорен, то наш рассудок в познавательном процессе предписывает свои законы природе:

«... хотя в начале это звучит странно, но, тем не менее, верно, если я скажу: рассудок не черпает свои законы (*a priori*) из природы, а предписывает их ей» [2].

Наверно, во времена Канта было разумно ограничиться приближением фиксированного мышления взрослого человека – все сразу не ох-

ватишь. Кроме того, не было еще теории Чарльза Дарвина. Скорее всего, если бы Кант знал теорию происхождения видов, то он явно бы задумался об эволюционном происхождении «чистого разума». Тем более что эволюционные идеи были явно не чужды Канту – вспомним его знаменитую гипотезу происхождения Солнечной системы. Но приближение фиксированного мышления взрослого человека накладывает свои ограничения – оно не позволяет ответить на вопросы – откуда же взялись познавательные способности, познаем ли мы истинные законы природы или наш рассудок «предписывает их ей». Фактически Кант ушел от наиболее острой части вопроса, поставленного Юмом – он не задавался вопросом, откуда взялся «чистый разум», а только тщательно и детально исследовал свойства «чистого разума» и применение его в научном познании.

Естественно, что после появления теории происхождения видов Дарвина должна была произойти ревизия концепции априорного «чистого разума». И она произошла. Очень четко ее выразил Конрад Лоренц в знаменитой статье «Кантовская доктрина априорного в свете современной биологии» [6]. Согласно Лоренцу, кантовские априорные категории и другие формы «чистого разума» произошли в результате естественного отбора: «Наши категории и формы восприятия, данные до индивидуального опыта, адаптированы к внешнему миру точно по той же причине, по какой копыто лошади адаптировано к почве степи и плавник рыбы адаптирован к воде до того, как рыба вылупится из икринки» [6].

То есть, составляющие «чистого разума» возникали постепенно в процессе эволюции, в результате многочисленных взаимодействий с внешним миром. В эволюционном контексте «чистый разум» совсем **не априорен**, а имеет явные эволюционные **эмпирические** корни.

Но это – только общая критика позиции Канта, которая лишь намекает, как подойти к решению проблемы, поставленной Юмом, но далеко не решает эту проблему.

#### 1.4. Немного о взглядах современных философов

Есть такое направление в современной философии – эволюционная эпистемология, два основных тезиса которой состоят в следующем (цитируем работу Карла Поппера, которая так и называется «Эволюционная эпистемология») [7]:

«**Первый тезис.** Специфически человеческая способность познавать, как и способность производить научное знание, являются результатами естественного отбора. Они тесно связаны с эволюцией специфически человеческого языка».

«**Второй тезис.** Эволюция научного знания представляет собой в основном эволюцию в направлении построения все лучших и лучших теорий. Это – дарвинистский процесс. Теории становятся лучше при-



способленными благодаря естественному отбору. Они дают нам все лучшую и лучшую информацию о действительности. (Они все больше и больше приближаются к истине.) Все организмы – решатели проблем: проблемы рождаются вместе с возникновением жизни.

При этом первый тезис считается почти тривиальным, а второй – разворачивается и всесторонне исследуется. То есть, эволюционная эпистемология занимается изучением того, каковы познавательные процессы и насколько их можно сопоставить с процессами накопления информации в процессе биологической эволюции. Но она практически не занимается изучением эволюционного происхождения познавательных способностей человека.

В какой-то степени этим занимается философское направление, которое можно назвать исследования когнитивной эволюции. На эту тему есть хорошая книга И.П. Меркулова «Когнитивная эволюция» [8]. Предмет этой книги частично перекрывается с известной книгой В.Ф. Турчина «Феномен науки» [9]. Особое внимание в [8] уделяется анализу процесса формирования логического мышления на этапах перехода от примитивного мышления первобытных племен к формальному логическому (от племен охотников до Аристотеля).

Но в работах философов как-то не ощущается понимания остроты проблемы о принципиальной способности человека познавать мир. Проблемы: почему с помощью нашего **человеческого** мышления, нашей логики, нашего интеллекта, нашего «чистого разума» мы в принципе способны познавать **природу**. 4. Как правило, много говорится о том, каковы методы познания, формализуются ли эти методы, говорится о том, что трудно формализовать все их особенности, но нигде не ставится задача – разобраться в том, почему они применимы в принципе. Нет прямо такой постановки проблемы ни в «Феномене науки» Турчина, ни в «Когнитивной эволюции» Меркулова. Хотя, конечно, в обеих книгах есть очень хорошие и глубокие подходы к исследованию эволюционного происхождения познавательных способностей человека.

А на самом деле ведь это драма! Примерно такая же, какую, возможно, ощутил И. Кант перед созданием «Критики чистого разума». Надо либо подвергнуть сомнению все научные знания – которые получены с помощью человеческого мышления, в применимости которого к научному познанию можно сомневаться – либо заняться обоснованием самого мышления. И естественный подход к решению проблемы – исследовать биологические корни наших познавательных способностей и постараться разобраться, почему эти способности возникли, и почему в процессе их эволюционного возникновения появилась возможность познания природы. Наиболее четкий путь такого исследования – построение математических и компьютерных моделей когнитивной эволюции. То есть, хотелось бы с помощью

моделей представить общую картину эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека. Причем здесь, как это ни удивительно, можно попытаться поставить эпистемологию – на твердую естественнонаучную почву.

Как же конкретно подойти к исследованию происхождения интеллекта? Есть ли задел таких исследований? Что здесь сделано и делается сейчас? В следующем разделе мы обсудим такой задел – направление исследований «Адаптивное поведение».

## **2. Модели адаптивного поведения – задел исследований когнитивной эволюции**

### **2.1. From Animal to Animat – исследование общих принципов поведения животного и робота**

С начала 1990-х годов активно развивается направление «Адаптивное поведение» (АП) [10]. Основной подход этого направления – конструирование и исследование искусственных (в виде компьютерной программы или робота) «организмов», способных приспосабливаться к внешней среде. Эти организмы называются «аниматами» (от англ. animal + robot = animat). Также часто используется термин «автономный агент».

Поведение аниматов имитирует поведение животных. Исследователи направления «Адаптивное поведение» стараются строить именно такие модели, которые применимы к описанию поведения *как реального животного, так и искусственного анимата* [11,12].

*Программа-минимум направления «АП» – исследовать архитектуру и принципы функционирования, которые позволяют животным или роботам жить и действовать в переменной внешней среде.*

*Программа-максимум этого направления – попытаться проанализировать эволюцию когнитивных способностей животных и эволюционное происхождение человеческого интеллекта* [13].

Программа максимум как раз и соответствует очерченной выше задаче моделирования когнитивной эволюции.

Для исследований АП характерен *синтетический подход*: здесь конструируются архитектуры, обеспечивающие «интеллектуальное» поведение аниматов. Причем это конструирование проводится как бы с точки зрения инженера: исследователь сам «изобретает» архитектуру, подразумевая конечно, что какие-то подобные структуры, обеспечивающие адаптивное поведение, должны быть у реальных животных.

При этом направление исследований АП рассматривается как *бюнический подход* к разработке систем искусственного интеллекта [14].

Хотя «официально» направление АП было провозглашено в 1990 году, были явные провозвестники этого направления. Приведем примеры из истории отечественной науки.

В 1960-х годах блестящий кибернетик и математик М.Л. Цетлин предложил и исследовал модели автоматов, способных адаптивно приспосабливаться к окружающей среде. Работы М.Л. Цетлина инициировали целое научное направление, получившее название «коллективное поведение автоматов» [15,16].

В 1960-70-х годах под руководством талантливого кибернетика М.М. Бонгарда был предложен интересный проект «Животное», направленный на моделирование адаптивного поведения искусственных организмов с иерархией целей и подцелей [17,18].

Хороший обзор ранних работ по адаптивному поведению, представлен в книге М.Г. Гаазе-Рапопорта, Д.А. Пospelова «От амёбы до робота: модели поведения» [18].

В исследованиях АП используется ряд нетривиальных компьютерных методов:

- нейронные сети,
- генетический алгоритм [19,20] и другие методы эволюционной оптимизации,
- классифицирующие системы (Classifier Systems) [21],
- обучение с подкреплением (Reinforcement Learning) [22].

## 2.2. Исследователи адаптивного поведения

Исследования по адаптивному поведению ведутся в ряде университетов и лабораторий, таких как:

- AnimatLab (Париж, руководитель – один из инициаторов данного направления Жан-Аркадий Мейер) [13,23]. Подход AnimatLab предполагает, что система управления анимата может формироваться и модифицироваться посредством 1) *обучения*, 2) индивидуальной *развития* (онтогенеза) и 3) *эволюции*.
- Лаборатория искусственного интеллекта в университете Цюриха (руководитель Рольф Пфейфер) [24,25]. Основной подход этой лаборатории – познание природы интеллекта путем создания («understanding by building»). Он включает в себя 1) построение моделей биологических систем, 2) исследование общих принципов естественного интеллекта животных и человека, 3) использование этих принципов при конструировании роботов и других искусственных интеллектуальных систем.
- Лаборатория искусственной жизни и роботики в Институте когнитивных наук и технологий (Рим, руководитель Стефано Нолфи) [26,27], ведущая исследования в области эволюционной роботики и принципов формирования адаптивного поведения.

- Лаборатория информатики и искусственного интеллекта в Массачусетском технологическом институте (руководитель Родни Брукс) [28,29], которая ведет исследования широкого спектра интеллектуальных и адаптивных систем, включая создание интеллектуальных роботов.
- Институт нейронаук Дж. Эдельмана, где ведутся разработки поколений моделей работы мозга (Darwin I, Darwin II, ...) и исследования поведения искусственного организма NOMAD (Neurally Organized Mobile Adaptive Device), построенного на базе этих моделей [30,31].

В России исследования АП пока ведутся скромными усилиями ученых-энтузиастов, среди этих работ следует отметить:

- модели поискового адаптивного поведения на основе спонтанной активности [11,12,32] (В.А. Непомнящих, Институт биологии внутренних вод им. И.Д. Папанина РАН);
- концепции и модели автономного адаптивного управления на основе аппарата эмоций [33] (А.А. Жданов, Институт системного программирования РАН);
- разработку принципов построения систем управления антропоморфных и гуманоидных роботов [34] (Л.А. Станкевич, Санкт-Петербургский политехнический университет);
- разработку нейросетевых моделей поведения роботов и робототехнических устройств [35] (А.А. Самарин, НИИ нейрокибернетики им. А.Б. Когана РГУ);
- модели АП на основе эволюционных и нейросетевых методов, в частности, модели эволюционного возникновения целенаправленного адаптивного поведения [36-38] (В.Г. Редько, М.С. Бурцев, О.П. Мосалов, Институт оптико-нейронных технологий РАН, Институт прикладной математики им. М.В. Келдыша РАН).

### **2.3. Проект «Мозг Анимата» [39]**

Анализ исследований в рамках анимат-подхода показывает, что разработанные модели пока еще очень фрагментарны и иллюстрируют только отдельные стороны адаптивного поведения. Поэтому было бы целесообразно предложить общую «платформу» для систематического построения моделей адаптивного поведения. В работах [39,40] предложен проект «Мозг Анимата», который нацелен на формирование общей схемы построения таких моделей. Проект основан на теории функциональных систем П.К. Анохина [41]. Кратко опишем данный проект, следуя [39], где предложена архитектура системы управления аниматом, основанная на нейросетевых адаптивных критиках. Нейросетевые адаптивные критики – схемы автономного адаптивного управления, основанные на методе обучения с подкреплением [22].

Предполагается, что система управления аниматом имеет иерархическую архитектуру. Базовым элементом системы управления является отдельная функциональная система (ФС). Верхний уровень соответствует основным потребностям организма: питания, размножения, безопасности, накопления знаний. Более низкие уровни соответствуют тактическим целям поведения. Блоки всех этих уровней реализуются с помощью ФС. Управление с верхних уровней может передаваться на нижние уровни (от «суперсистем» к «субсистемам») и возвращаться назад. Предполагается, что система управления аниматом функционирует в дискретном времени  $t = 1, 2, \dots$ , и в каждый такт времени активна только одна ФС.

Предполагается простая формализация ФС на основе адаптивных критиков. Формальная ФС моделирует следующие важные особенности ее биологического прототипа: а) прогноз результата действия, б) сравнение прогноза и результата, и в) коррекцию прогноза путем обучения в соответствующих нейронных сетях.

**Схема адаптивного критика.** Рассматриваемая схема адаптивного критика состоит из двух нейронных сетей: Модель и Критик. Предполагается, что производные по весам синапсов нейронных сетей могут быть вычислены обычным методом обратного распространения ошибки [42]. Также предполагается, что адаптивный критик предназначен для выбора одного из нескольких действий. Например, при управлении движением действиями могут быть: двигаться вперед, поворачивать вправо, поворачивать влево, стоять на месте. В каждый момент времени  $t$  адаптивный критик должен выбрать одно из возможных действий.

Цель адаптивного критика – максимизировать функцию суммарной награды

$$U(t) = \sum_{j=0}^{\infty} \gamma^j r(t_j) \quad , \quad t = t_0, t_1, t_2, \dots,$$

где  $r(t_j)$  – текущее подкрепление (награда  $r(t_j) > 0$ , или наказание  $r(t_j) < 0$ ), полученное адаптивным критиком в данный момент времени  $t_j$ ,  $\gamma$  – коэффициент забывания,  $0 < \gamma < 1$ .  $\tau = t_{j+1} - t_j$  – шаг по времени. Коэффициент забывания учитывает, что чем дальше анимат «заглядывает» в будущее, тем меньше у него уверенность в оценке награды («рубль сегодня стоит больше, чем рубль завтра»).

Модель имеет два типа входов: 1) входы, характеризующие текущую ситуацию  $\mathbf{S}(t)$  (сигналы из внешней и внутренней среды анимата), и 2) входы, характеризующие действия. Предполагается, что число возможных действий  $n_a$  невелико. Роль Модели – прогноз следующей ситуации  $\mathbf{S}^{\text{Pr}}_i(t+\tau)$  для всех возможных действий  $a_i, i=1,2,\dots, n_a$ .

Роль Критика – оценка качества ситуаций  $V(\mathbf{S})$ , в частности, прогнозируемых ситуаций  $\mathbf{S}^{\text{Pr}}_i(t+\tau)$  для всех возможных действий. Вели-

чина  $V(S)$  есть оценка аниматом функции суммарной награды  $U(t)$ , которую можно получить в будущем, исходя из данной ситуации  $S$ . Эта оценка постепенно уточняется в процессе обучения.

При выборе действия применяется  $\varepsilon$ -жадное правило [22]:

- с вероятностью  $1 - \varepsilon$  выбирается действие с максимальным значением  $V(S^{\text{pr}}_i(t+\tau))$ :

$$k = \arg \max_i \{V(S^{\text{pr}}_i(t+\tau))\},$$

- с вероятностью  $\varepsilon$  выбирается произвольное действие  $a_k$ ,  $0 < \varepsilon \ll 1$ ,

$k$  – индекс выбираемого действия.

Применение этого правила означает, что с большой вероятностью выбирается действие, для которого максимальна оценка суммарной награды  $U(t)$ , ожидаемой в будущем.

Обучение Модели проводится обычным методом обратного распространения ошибки [42]. Смысл обучения Модели – уточнение прогнозов будущих ситуаций.

Обучение Критика проводится методом ошибки временной разности [22]. Смысл обучения Критика состоит в том, чтобы итеративно уточнять оценки качества ситуаций  $V(S(t))$  в соответствии с поступающими подкреплениями  $r$ .

#### **Функционирование системы управления аниматом.**

Предполагается следующая схема работы ФС в рамках функционирования всей системы управления аниматом. Каждая ФС представляет собой адаптивный критик, имеющий связи с некоторой суперсистемой (системой более высокого уровня иерархии) и, возможно, одной или несколькими подсистемами (системами более низкого уровня иерархии). ФС активизируется командой от суперсистемы; Модель и Критик функционируют так, как описано выше. В результате осуществляется выбор действия  $a_k$ . Дальнейшее зависит от вида действия  $a_k$ . Если действие – команда для исполнительных элементов, то такое действие выполняется сразу. Далее анимат получает подкрепление  $r$  из внешней или внутренней среды, и производится обучение в нейронных сетях Модели и Критика.

Другой тип действий – команды для подсистем. Для такого действия подается команда активизации определенной подсистемы (выбор конкретной подсистемы определяется номером действия  $a_k$ ). В этом случае сравнение прогноза и результата, оценка подкрепления  $r$  и обучение нейронных сетей откладывается до получения отчета от подсистемы.

После выполнения всех этих действий ФС посылает отчет об окончании своей работы суперсистеме.

Описанный способ работы ФС представляет собой обычный режим функционирования. Вводится также экстраординарный режим, который имеет место, если прогноз существенно отличается от фак-

тического результата. В экстраординарном режиме величина  $\varepsilon$  (вероятность выбора случайного действия) в данной ФС и ее подсистемах резко возрастает, и поиск новых решений включает большую случайную компоненту. Этот поиск может сопровождаться случайным формированием и селекцией новых функциональных систем, аналогично селекцией нейронных групп в теории нейродарвинизма Дж. Эдельмана [43]. Таким образом, обычный режим функционирования может рассматриваться как тонкая настройка системы управления аниматом, в то время как экстраординарный режим – это грубый поиск подходящего поведения в чрезвычайных ситуациях.

Подробнее о работе данной схемы Мозга Анимата см. в [20,39].

#### **2.4. Выводы по моделям адаптивного поведения**

Исследования адаптивного поведения – актуальное, содержательное и конструктивное направление, которое непосредственно связано с моделированием когнитивной эволюции. Также это направление исследований важно как биологически инспирированная научная основа разработок систем искусственного интеллекта. Это направление использует серьезные математические и компьютерные методы, и здесь построено множество интересных и содержательных моделей. Однако, в целом, результаты этих исследований пока еще далеки от решения стратегических задач, поставленных при иницировании этого направления.

Один из значительных и достаточно неожиданных выводов этих исследований состоит в том, что часто нетривиальное поведение может быть сформировано простой системой управления [11]. Причем, такой системой управления, о которой сам исследователь может и не догадаться – система управления (в частности, нейронная сеть) может формироваться в процессе эволюционной самоорганизации, например, с помощью генетического алгоритма.

#### **3. Моделирование когнитивной эволюции – естественнонаучная основа разработок искусственного интеллекта**

Направление исследований Искусственный интеллект (ИИ), скорее всего, можно рассматривать как прикладное – применение принципов естественного интеллекта в искусственных практически важных для человека компьютерных системах. Судьба прикладных разработок зависит от наличия достаточно серьезного научного фундамента, на котором базируются такие разработки. Например, научной базой развития микроэлектроники во второй половине 20-го века была физика твердого тела. При этом для физиков чисто научные исследования твердого тела были интересны фактически независимо от применения их исследований, в результате чего научная основа микроэлектроники интенсивно развивалась. И результаты микроэлектроники, как наукоемкой технологии, впечатляющи.

Моделирование когнитивной эволюции чрезвычайно интересно и важно с точки зрения научного миропонимания. Следовательно, можно ожидать, что такие исследования, которые предоставили бы картину возникновения и эволюционного формирования естественного интеллекта, будут очень интересны для ученых. Но эти исследования могут быть тесно связаны и с разработками ИИ. И, следовательно, могло бы быть взаимное обогащение фундаментальных и прикладных исследований природы интеллекта. И, тем самым, исследования когнитивной эволюции могли бы служить научной основой разработок систем ИИ.

#### **4. Контуры программы будущих исследований**

Итак, очерчены модели и задел исследования когнитивной эволюции. Видно, что хотя проделана большая работа, ученые еще очень далеки от понимания того, как возникали и развивались системы управления живых организмов, как развитие этих систем способствовало эволюции когнитивных способностей животных, и как процесс когнитивной эволюции привел к возникновению интеллекта человека. Тут огромная область чрезвычайно интересных исследований, которые только-только начинаются.

Очень хороший задел к исследованию когнитивной эволюции – модели адаптивного поведения (раздел 2), дальняя цель которых как раз и состоит в осмыслении эволюции когнитивных способностей животных в контексте эволюционного происхождения человеческого интеллекта. Поэтому предложим план исследований, отталкиваясь от работ по моделям адаптивного поведения.

**Разработка схем и моделей адаптивного поведения анимата на базе проекта «Мозг Анимата».** Воплощение в конкретные модели конструкций Мозга Анимата разумно начать с анализа целостного адаптивного поведения простых агентов, имеющих естественные потребности: питания, размножения, безопасности. Эволюционная схема формирования нейросетевой системы управления подобных агентов, обеспечивающей достаточно нетривиальную структуру целей и подцелей, была исследована М.С. Бурцевым [37]. Теперь было бы полезно промоделировать подобные системы управления в рамках проекта «Мозг анимата».

Дальнейшая работа могла бы включать в себя анализ интеллектуальных изобретений биологической эволюции, таких как привыкание и условные рефлексы (раздел 1), на основе исследований проекта «Мозг анимата».

**Исследование перехода от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов.** Такой переход можно рассматривать, как появление в «сознании» животного свойства «понятие». Обобщенные образы можно представить как мысленные аналоги наших слов, не произносимых животными, но реально используемых ими. Например, у собаки явно есть понятия



«хозяин», «свой», «чужой», «пища». И важно осмыслить, как такой весьма нетривиальный переход мог произойти в процессе эволюции.

**Исследование процессов формирования причинной связи в памяти животных.** По-видимому, запоминание причинно-следственных связей между событиями во внешней среде и адекватное использование этих связей в поведении – одно из ключевых свойств активного познания животным закономерностей внешнего мира. Такая связь формируется, например, при выработке условного рефлекса: животное запоминает связь между условным стимулом (УС) и следующим за ним безусловным стимулом (БС), что позволяет ему предвидеть события в окружающем мире и адекватно использовать это предвидение.

Естественный следующий шаг – переход от отдельных причинных связей к «базам знаний», к логическим выводам на основе уже сформировавшихся знаний.

**Исследование процессов формирования логических выводов в «сознании» животных.** Фактически, уже на базе классического условного рефлекса животные способны делать «логический вывод» вида: {УС, УС --> БС} => БС или «Если имеет место условный стимул, и за условным стимулом следует безусловный, то нужно ожидать появления безусловного стимула». Можно даже говорить, что такие выводы подобны выводам математика, доказывающего теоремы (раздел 1). И целесообразно разобратся в системах подобных выводов, понять, насколько адаптивна логика поведения животных и насколько она подобна нашей, человеческой логике.

**Исследование коммуникаций, возникновения языка.** Наше мышление тесно связано с языком, с языковым общением между людьми. Поэтому целесообразно проанализировать: как в процессе биологической эволюции возникал язык общения животных, как развитие коммуникаций привело к современному языку человека, как развитие коммуникаций и языка способствовало развитию логики, мышления, интеллекта человека.

Конечно же, перечисленные пункты формируют только контуры плана будущих исследований. Тем не менее, уже сейчас видно, сколь широк фронт исследований, и как много нетривиальной, интересной и важной работы предстоит сделать.

#### ЛИТЕРАТУРА

1. Кант И. Критика чистого разума. Соч. в 6-ти томах. Т.3. М.: Мысль, 1964. С. 69-695.
2. Кант И. Прологомены ко всякой будущей метафизике, могущей появиться как наука. Соч. в 6-ти томах. Т.4, часть 1. М.: Мысль, 1965. С. 67-210.
3. Semantic Networks in Artificial Intelligence, Lehmann, Fritz, ed., Pergamon Press, Oxford, 1992.
4. Воронин Л.Г. Эволюция высшей нервной деятельности. М.: Наука. 1977. 128 с.

5. Юм Д. Исследование о человеческом познании. Соч. в 2-х томах. Т.2. М.: Мысль, 1966. С. 5-169.
6. Lorenz K. Kant's doctrine of the a priori in the light of contemporary biology (1941) // In: ed. Plotkin H., Learning, Development and Culture. N.Y. 1982.
7. Поппер К. Эволюционная эпистемология // Сб. Эволюционная эпистемология и логика социальных наук: Карл Поппер и его критики. Составление Д.Г. Лахути, В.Н. Садовского, В.К. Финна. М.: Эдиториал УРСС, 2000.
8. Меркулов И.П. Когнитивная эволюция. М. Наука, 1999.
9. Турчин В.Ф. Феномен науки. Кибернетический подход к эволюции. М.: Наука, 1993. 295с. (1-е изд.). М.: ЭТС, 2000. 368с (2-е изд.). См. также <http://www.refal.ru/turchin/phenomenon/>
10. Meyer J.-A., Wilson S. W. (Eds) From animals to animats. Proceedings of the First International Conference on Simulation of Adaptive Behavior. The MIT Press: Cambridge, Massachusetts, London, England. 1990.
11. Непомнящих В.А. Аниматы как модель поведения животных // IV Всероссийская научно-техническая конференция «Нейроинформатика-2002». Материалы дискуссии «Проблемы интеллектуального управления – общесистемные, эволюционные и нейросетевые аспекты». М.: МИФИ, 2003. С. 58-76. См. также <http://www.keldysh.ru/pages/BioCyber/RT/Nepomn.htm>
12. Непомнящих В.А. Поиск общих принципов адаптивного поведения живых организмов и аниматов // Новости искусственного интеллекта. 2002. N.2. С. 48-53.
13. Donnarumma J.Y., Meyer J.A. Learning reactive and planning rules in a motivationally autonomous animat // IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 1996. V. 26. N. 3. PP.381-395. See also: <http://animatlab.lip6.fr/index.en.html>
14. Wilson S.W. The animat path to AI // In: [10]. PP. 15-21.
15. Цетлин М.Л. Исследования по теории автоматов и моделирование биологических систем. – М.: Наука, 1969. 316 с.
16. Вавиловский В.И., Поспелов Д.А. Оркестр играет без дирижера. М.: Наука, 1984.
17. Бонгард М.М., Лосев И.С., Смирнов М.С. Проект модели организации поведения – «Животное» // Моделирование обучения и поведения. М.: Наука, 1975. С.152-171.
18. Гаазе-Ранопорт М.Г., Поспелов Д.А. От амебы до робота: модели поведения. М.: Наука, 1987.
19. Holland J.H. Adaptation in Natural and Artificial Systems. – Ann Arbor, MI: The University of Michigan Press, 1975 (1st edn). Boston, MA: MIT Press., 1992 (2nd edn).
20. Редько В.Г. Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики. М.: УРСС, 2005. 224 с.
21. Holland J.H., Holyoak K.J., Nisbett R.E., Thagard P. Induction: Processes of Inference, Learning, and Discovery. Cambridge, MA: MIT Press, 1986.
22. Sutton R., Barto A. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 1998. See also: <http://www.cs.ualberta.ca/~sutton/book/the-book.html>
23. Сайт AnimatLab: <http://animatlab.lip6.fr/index.en.html>
24. Сайт AI Laboratory of Zurich University: <http://www.ifi.unizh.ch/groups/ailab/>
25. Pfeifer R., Scheier C., Understanding Intelligence. MIT Press, 1999.

26. Сайт Laboratory of Artificial Life and Robotics: <http://gral.ip.rm.cnr.it/>
27. *Nolfi S., Floreano D.* Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines. Cambridge, MA: MIT Press/Bradford Books, 2000. 384 p.
28. Сайт MIT Computer Science and Artificial Intelligence Laboratory: <http://www.csail.mit.edu/index.php>
29. *Brooks R.A.* Cambrian Intelligence: The Early History of the New AI. MIT Press, 1999.
30. Сайт Neuroscience Institute: <http://www.nsi.edu/>
31. *Krichmar J.L., Edelman G.M.* Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device, *Cerebral Cortex*, 2002, V. 12. PP. 818-830.
32. *Nepomnyashchikh V.A., Podgornyj K.A.* Emergence of adaptive searching rules from the dynamics of a simple nonlinear system // *Adaptive Behavior*. 2003. V.11. N.4. PP.245-265.
33. *Жданов А.А.* Метод автономного адаптивного управления // *Известия Академии Наук. Теория и системы управления*. 1999. N. 5. С. 127-134.
34. *Станкевич Л.А.* Нейрологические средства систем управления интеллектуальных роботов // VI Всероссийская научно-техническая конференция «Нейроинформатика-2004». Лекции по нейроинформатике. Часть 2. М.: МИФИ, 2004. С. 57-110.
35. *Самарин А.И.* Модель адаптивного поведения мобильного робота, реализованная с использованием идей самоорганизации нейронных структур // IV Всероссийская научно-техническая конференция «Нейроинформатика-2002». Материалы дискуссии «Проблемы интеллектуального управления – общесистемные, эволюционные и нейросетевые аспекты». М.: МИФИ, 2003. С.106-120.
36. *Бурцев М.С., Гусарев Р.В., Редько В.Г.* Исследование механизмов целенаправленного адаптивного управления // *Изв. РАН «Теория и системы управления»* 2002. N.6. С.55-62.
37. *Бурцев М.С.* Модель эволюционного возникновения целенаправленного адаптивного поведения. 2. Исследование развития иерархии целей // *Препринт ИПМ РАН*, 2002, N. 69.
38. *Redьko V.G., Mosalov O.P., Prokhorov D.V.* A model of evolution and learning // *Neural Networks*, 2005. V. 18. PP. 738-745.
39. *Redьko V.G., Prokhorov D.V., Burtsev M.S.* Theory of functional systems, adaptive critics and neural networks // *International Joint Conference on Neural Networks*, Budapest, 2004. PP. 1787-1792.
40. *Анохин К.В., Бурцев М.С., Зарайская И.Ю., Лукашев А.О., Редько В.Г.* Проект «Мозг анимата»: разработка модели адаптивного поведения на основе теории функциональных систем // *Восьмая национальная конференция по искусственному интеллекту с международным участием. Труды конференции*. М.: Физматлит, 2002. Т.2. С.781-789.
41. *Анохин П.К.* Системные механизмы высшей нервной деятельности. М.: 1979.
42. *Rumelhart D.E., Hinton G.E., Williams R.G.* Learning representation by back-propagating error // *Nature*. 1986. V.323. N.6088. PP. 533-536.
43. *Edelman G. M.* Neural Darwinism: The Theory of Neuronal Group Selection. Oxford: Oxford University Press, 1989.

---

# МОДЕЛИРОВАНИЕ ТВОРЧЕСКОГО ПРОЦЕССА

*В.А. Глазунов, Л.И. Чистоходова*

В.А. Глазунов, Л.И. Чистоходова

Рассматриваются модели, призванные в какой-то степени описать процесс творческого озарения. Это могут быть моменты, связанные с творчеством в области науки или искусства, а также с герменевтическим актом понимания текста.

Творческий акт – это едва ли не самое захватывающее и мало исследованное явление человеческой жизни. Он не поддается логическому осмыслению и не может быть вызван строгим алгоритмом действий субъекта. Как возникает творческое озарение? Над этим вопросом размышляли Пушкин и Станиславский, Эйнштейн и Рассел. В данной работе мы попытаемся дать свою интерпретацию перехода саморазвивающейся системы, одной из которых является человеческая личность, на некий новый и более высокий уровень развития. Речь может идти, например, о том, как ученик «вдруг» начинает понимать недоступную прежде для него теорему, или о том, как ученый формулирует новую революционную гипотезу происхождения Вселенной. Представляемые здесь модели<sup>1</sup>, понятно, не могут научить, как достигать озарения, но они, быть может, помогут представить этот процесс и тем самым внесут, пусть маленький, вклад в познание процессов творчества.

Мы исследуем процесс, когда некая саморазвивающаяся система<sup>2</sup> в своем развитии доходит до бифуркации, при которой возможны состояния, чреватые переходом в иную размерность более высокого порядка. Нашей целью является наглядное представление механизмов подобного процесса, и в данном случае слово «механизм» употребляется не в переносном, а в прямом смысле – это действительно должны быть механические модели упомянутого перехода. В связи со сказанным укажем на важность мысленных экспериментов, которые можно было бы поставить с помощью таких моделей. В литературе неоднократно упоминалась роль таких экспериментов. Но для того чтобы поставить их, исследователь должен обладать моделью, с одной стороны, весьма наглядной и простой, а с другой стороны, достаточно полно отражающей сущность исследуемого явления.

---

<sup>1</sup> Речь идет о механической модели, но это новый «постнеклассический механицизм», по выражению В.И. Аршинова.

<sup>2</sup> *Пригожин И., Стенгерс И.* Порядок из хаоса. Новый диалог человека с природой. М.: Прогресс, 1986, 432 с.

В данной связи упомянем примеры, приведенные В.С. Степиным<sup>3</sup> и касающиеся мысленных экспериментов на идеализированных моделях, применявшихся при выводе законов электромагнитного взаимодействия или при установлении характеристик квантованного поля. Подчеркнем, что Бору и Розенфельду (согласно принципу наблюдаемости) понадобился именно мысленный эксперимент для определения упомянутых характеристик – при этом рассматривались возможные взаимодействия между макросистемой, включающей в себя измерительные устройства, и весьма тонкими микрообъектами.

Саморазвивающаяся система, представляющая собой творческую личность, на определенном этапе развития может функционировать в пределах какой-то размерности, она может ставить перед собой и решать определенные задачи, например, школьник может использовать операцию сложения, а актер показывать этюд, имитирующий какую-то ситуацию. Данное положение дел можно сравнить с нормальным состоянием существующей научной парадигмы, исследованной Т. Куном.<sup>4</sup>

Однако наступает момент, когда школьнику требуется осознать сущность таблицы умножения. Это серьезный герменевтический акт, связанный с переходом на новый, более высокий уровень обобщения. Начинающему актеру, быть может, предстоит сыграть гораздо более значимую, чем ранее, роль, например, прохожего в «Вишневом саде». В данном случае самоорганизующаяся система должна выстроить новую цепочку взаимосвязанных элементов, которой становится «тесно» в теле прежней структуры. Примерно об этом писал Станиславский, когда указывал, что новую роль рождает система, состоящая из драматурга, режиссера и актера, должностующего осознать себя в новых предлагаемых обстоятельствах.<sup>5</sup>

После некоторых усилий в системе возникает новая цепочка взаимосвязанных звеньев. Ученик, изучающий таблицу умножения, начинает осознавать, что не обязательно последовательно складывать одинаковые слагаемые, а можно иметь в виду лишь их величину и количество. Актер, занимающийся ролью, нашел «приспособления» и «маленькие правды» (выражения Станиславского), могущие помочь испытать вдохновение к семи часам вечера, когда начнется спектакль.

В сложившейся ситуации самоорганизующаяся система готова выйти на более высокий уровень развития, но внутренние ее ресурсы не способны это сделать. В то же время «...небольшое энергетическое воздействие – укол в нужном пространственно-временном ло-

---

<sup>3</sup> *Степин В.С.* Теоретическое знание. М.: Прогресс-традиция, 2000, 574с.

<sup>4</sup> *Кун Т.* Структура научных революций. М.: Прогресс 1976, 288 с.

<sup>5</sup> *Станиславский К.С.* Моя жизнь в искусстве. М.: Искусство, 1983, 424 с.

кусе оказывается достаточным, чтобы система перестроилась и возник новый тип структур».<sup>6</sup> Для постижения таблицы умножения, по-видимому, таким «уколом» будет непосредственное сравнение результатов при последовательном сложении и при умножении:  $2+2+2=2 \times 3$ . Для актера, досконально изучившего эпоху, в которой живет его герой, а также все нюансы сквозного действия и сверхзадачи, это может быть напутственное слово режиссера перед спектаклем. Для научной парадигмы, переживающей период противоречий между теорией и экспериментом, это может быть парадигмальная прививка (В.С. Степин), появившаяся в результате междисциплинарного взаимодействия.

Перейдем к рассмотрению моделирования творческого процесса. Одна группа моделей может послужить для представления перехода системы из одного функционального состояния в другое без изменения ее размерности. Система может содержать два линейных двигателя, связанных с основанием и друг с другом. Эти двигатели способны охватить какую-то область в плоскости, и если внешняя сила действует через пружину, то линейные двигатели могут данную силу уравновесить. Эта робототехническая система, работающая в плоскости, может выполнять какие-то операции лишь до того момента, пока оси двигателей не совпадут по направлению – это будет так называемое особое положение или положение бифуркации.

Рассматриваемая система с двумя степенями свободы может представлять своего рода машину катастроф. Известна машина катастроф Зимана<sup>7</sup>, имеющая одну степень свободы, данная же система, обладая двумя степенями свободы, может описать гораздо более сложные процессы, чем машина Зимана. При приближении к поверхности (в данном случае линии) бифуркации происходит срыв – катастрофа. Если мы учтем жесткости внутренних и внешних пружин, то станет ясно, что процесс бифуркации будет сопровождаться первоначальным усилением сопротивления, затем ослаблением этого сопротивления и далее, после пересечения линии бифуркации система сама себя стремительно будет выводить в новое функциональное состояние.

Представляемая модель, может характеризовать локальную научную революцию, происходящую в ситуациях, когда одна из семантически различаемых теорий (по Е.А. Мамчур<sup>8</sup>) выигрывает конкуренцию. При этом другая научная школа, исповедовавшая иную теорию, вынуждена под действием внешних, по отношению к ней, сил претерпеть катастрофу, бифуркацию.

<sup>6</sup> *Степин В.С.* Теоретическое знание. М.: Прогресс-традиция, 2000, с. 696.

<sup>7</sup> *Постон Т., Стюарт И.* Теория катастроф и ее приложения. М.: Мир, 1980, 642 с., Арнольд В.И. Теория катастроф. М.: Наука, 1990, 128 с.

<sup>8</sup> *Мамчур Е.А.* Проблема выбора теории. М.: Наука, 1975, 232 с.

Подобный процесс может происходить при «гештальт-переключении», когда одно представление об объекте сменяется другим. Школьник может «механически» запомнить строки таблицы умножения. Актер, следуя принципам «театра представления», а не «театра переживания», может научиться копировать приемы своего коллеги или режиссера, не достигнув «истины страстей и правдоподобия чувств» в предполагаемых (Пушкин) или предлагаемых (Станиславский) обстоятельствах. Иными словами, рассматриваемая модель не может описать перехода на более высокий уровень развития системы, связанный с увеличением ее размерности.

Другая модель, характеризующая переход на более высокую размерность системы, отличается от первой тем, что в ней имеется дополнительный линейный двигатель. Представим себе, что в саморазвивающейся системе появились новые элементы, выстроившиеся в связанную цепочку. Эта ситуация может характеризоваться появлением нового привода, и система в этом состоянии подвержена действию внутренних напряжений – она переопределена.

Предположим, наступит момент, когда на систему подействует сила, перпендикулярная плоскости, в которой лежат все рассматриваемые двигатели. Структура, качественные свойства системы могут при этом существенно измениться – система может выйти за пределы плоскости, мы получим уже не плоскую, а пространственную модель, и система обретет новую степень свободы. Именно этот переход, на наш взгляд, является воплощением творческого акта, хотя переход к новой размерности обуславливает некоторую ломку старых структурных элементов.

Прежнее состояние, по отношению к новому состоянию, теперь будет характеризовать поверхность (в данном случае двумерную плоскость) бифуркации. Отметим еще раз, что имеет место не точка, а поверхность бифуркации, при нахождении на которой система функционирует в пространстве на единицу меньшей размерности. Укажем, что идея творческого перехода на иную размерность в произведениях искусства была во многом разработана А.А. Кобляковым<sup>9</sup>, в данной же статье представлена механическая или робототехническая модель данного процесса.

Переход к новой размерности, иному числу степеней свободы связан с наличием некоторой силы, действующей ортогонально исходной структуре низшего ранга. Саморазвивающаяся человекоразмерная система, представляющая в нашем примере школьника или ак-

---

<sup>9</sup> Кобляков А.А. Синергетика и творчество: универсальная модель устранения противоречий как основа новой стратегии исследований. /Синергетическая парадигма. Многообразие поисков и подходов. М.: Прогресс-традиция, 2000, с.305-325.

тера, как бы интуитивно ищет «ортогональные» воздействия путем некоего «рыскания», «сканирования» плоскости. После получения такого импульса система сама себя может довести до нового относительно устойчивого состояния. Школьник, усвоив суть какой-то теоремы, сможет осознанно и с успехом ее применять для разных задач, а актер, почувствовав «правду» в исполнении какой-то роли, сможет вдохновенно импровизировать «внутри заданной композиции».

Поверхность бифуркации по отношению к нормальному функциональному состоянию представляет собой некоторое состояние деградации: если система попадает в область бифуркации и затем остается в ней, то это означает уменьшение размерности, числа степеней свободы. Длительное пребывание на этой поверхности может быть чревато нарушением одной из цепочек взаимосвязи – тогда система (быть может, на время) теряет возможность возврата к более высокой размерности.

Представленная модель иллюстрирует и некоторые положения герменевтики, поскольку для понимания текста, например, параграфа учебника также необходимо выстроить структуру некоторых сведений, которые в дальнейшем переходят к «новой размерности». Примерно так же, на наш взгляд, можно трактовать положение А. Маслоу о моменте «инсайта» – творческого озарения при решении научных или художественных задач. После того, как школьник освоил таблицу умножения, он когда-нибудь постигнет и следующую ступень – возведение в степень. Актер, научившийся творчески переживать роль прохожего из «Вишневого сада», когда-нибудь сможет сыграть и Фирса в этой пьесе.

В заключение сделаем некоторые выводы.

Творческий акт, характеризуемый переходом системы на более высокий уровень функционирования, может быть смоделирован появлением новой подсистемы, которая внутри старой структуры вызывает «напряжения». Для выхода на новую размерность необходимо воздействие, направленное «ортогонально» исходному состоянию, старой парадигме, чтобы «взломать» имеющуюся структуру. После этого система может сама себя достаточно быстро переводить в новое относительно устойчивое функциональное состояние. Гиперповерхности бифуркации самоорганизующихся систем представляют состояния, характеризуемые меньшей их размерностью.



---

# МИР КАК МОДЕЛЬ ВНУТРИ СВЕРХМАШИНЫ И ВИРТУАЛЬНЫЕ МИРЫ

*М.Б.Игнатьев*

Картина мира в каждую из эпох формируется на основе синтеза всех достижений во всех сферах человеческой деятельности. В эпоху расцвета механики формировался механицизм, в эпоху развития компьютерной техники естественно ожидать формирования компьютеризма. Неклассические науки привнесли в картину мира наблюдателя. Постнеклассические науки привнесли в картину мира управителя.

Успехи современной науки со времен Ньютона неоспоримы, но чем энергичнее внедряются ее результаты в виде различных машин и технологий во все сферы жизни, тем явственнее проступают ее недостатки. Один из главных недостатков классической и неклассической науки заключается в том, что современные технологии рассчитаны на использование больших количеств энергии и материалов, на использование больших давлений, напряжений, усилий, температур и т.д., что приводит к загрязнению окружающей среды, истощению источников энергии и материалов, гибели живой природы – то есть к тому, что называют экологическим кризисом.

Истоки этих недостатков лежат в самих парадигмах классической и неклассической науки, ее деятели слишком часто пользовались бритвой Оккама, срезая как бы все лишнее и слишком упрощая проблемы. В итоге сложилось стремление к «гениальной» простоте, физика заполнилась формулами из трех букв вроде закона Ома. И если это было простительно в докомпьютерный век, то с появлением мощных компьютеров, которые буквально входят в каждый дом, неоправданное упрощение недопустимо, недопустимо пренебрежение тонкими сущностями. Информатика имеет дело со слабыми сигналами, которые могут управлять большими процессами. Информатизация всех отраслей человеческой деятельности – это прежде всего выявление возможностей управления с помощью слабых сигналов, слабых по мощности, температуре, напряжению. Но для того, чтобы управлять системами, необходимо иметь новые модели различных процессов, в сами эти модели должна быть заложена возможность информационного управления. В этом сущность процесса информатизации. Постнеклассическая наука должна иметь свои модели.

Ниже рассматривается новый класс таких моделей – лингвокомбинаторные модели и вопросы реализации их с помощью рекурсивных вычислительных систем для построения виртуальных миров.

## РЕКУРСИВНЫЕ ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Важным достижением на пути развития вычислительной техники стало развитие рекурсивных вычислительных систем с не-фон-Неймановской архитектурой, создание которых было провозглашено в 1974г на конгрессе ИФИП в Стокгольме [8] В результате большой работы в Ленинградском институте авиационного приборостроения к 1979г были изготовлены многие блоки машины и осенью 1979 г экспериментальный образец рекурсивной машины был предъявлен государственной комиссии во главе в академиком А.А.Дородницыным. В специальном Постановлении ГКНТ СССР и Комиссии Президиума Совета Министров СССР от 14.09.1979г за № 472/276 отмечалось, что запуск первого в мире экспериментального образца многопроцессорной рекурсивной машины высокой производительности и надежности является достижением мирового уровня

В математике существует большой раздел – рекурсивные функции [2]. Долгое время термин «рекурсия» употреблялся математиками, не будучи четко определенным. Его приблизительный интуитивный смысл можно описать следующим образом. Значение искомой функции  $\Phi$  в произвольной точке  $X$  (од точкой подразумевается набор значений аргументов) определяется, вообще говоря, через значения этой же функции в других точках  $H$ , которые в каком-то смысле предшествуют  $X$ . Само слово «рекурсия» означает возвращение [7,9]. Рекурсивные функции – это вычислимые функции. По сути дела все вычислимые на компьютерах функции – это рекурсивные функции, но разные компьютерные архитектуры поразному ведут вычислительные процессы. Чем лучше соответствует структура компьютера структуре задач, тем меньше затраты памяти и времени. Так что когда мы говорим о рекурсивных машинах, мы говорим о соответствии структур машины и задач, а так как задачи бывают разные, то структура машин должна гибко подстраиваться к структурам задач. Математика в настоящее время погружена в программирование, и в программировании рекурсивные операции распространены.

ЭВМ выступает как средство материализации логико-математических преобразований. ЭВМ являет собой иллюстрацию концепции потенциальной осуществимости, поскольку при отсутствии ограничений на время работы и емкость памяти любая ЭВМ в состоянии провести любые вычисления. Конкретное же протекание процессов вычисления проявляется лишь на уровне организации преобразований информации (задействуются конкретные регистры, коммутаторы, процессоры, линии передачи данных в определенном порядке и сочетании и т.д.). С этой точки зрения «архитектура ЭВМ» – это ее структура в состоянии (процессе) реализации алгоритма, то есть как бы ожившая структура. такого представления является возможность отображени категорий и явлений одной природы (числа, алгоритмы)

на объекты другой природы (физические элементы, сигналы). При чем это отображение взаимно неоднозначно – алгоритму  $a_j$  может соответствовать множество архитектур  $\{A\}$  и обратно – архитектуре  $A_j$  непосредственно не соответствует какой-либо алгоритм  $a_j$ . Специфика взаимодействия  $\{a\}$  и  $\{A\}$  раскрывает глубинные свойства процесса развития математики и вычислительной техники. Как отмечает С.А Яновская, «лицо машинной математики все более зависит от развития философских и логических оснований математики» [9]. Не представляется возможным непротиворечивая формализация отображения  $\{a\} \rightarrow \{A\}$  из-за его неоднозначности. Поэтому построить соответствующую аксиоматическую теорию проектирования ЭВМ не представляется возможным [10].

Когда мы формулировали принципы организации рекурсивных машин, мы исходили из потребностей развития вычислительных машин и систем, получили множество авторских свидетельств, это был интересный творческий процесс и с точки зрения достоверности сделанного тогда, в 1974-1979 годах, стоило бы обратиться к нашему докладу на конгрессе ИФИП в Стокгольме [8]. Этот доклад содержал анализ недостатков машин традиционной архитектуры, ревизию принципов фон Неймана, принципы архитектуры рекурсивных машин, основные особенности языка рекурсивных машин, фрагментарное описание рекурсивной машины. В качестве иллюстрации рекурсивной структуры можно привести систему 3М – модульную микропроцессорную систему. Система 3М строится из модулей трех типов – операционных, коммуникационных и интерфейсных. Операционные модули выполняют основную работу по обработке данных, реализации объектов математической памяти, процессов определения готовности и выполнения операторов программы на внутреннем языке. Коммуникационный модуль предназначен для реализации коммуникационной системы – установления логического соединения между модулями, обмена информацией между модулями поиска в системе ресурсов запрошенного типа. Интерфейсные модули подключаются к внешним устройствам своими блоками ввода-вывода.

Вопросы организации обмена информацией с внешним миром имеют большое значение для существенно многопроцессорных систем, оказывают значительное влияние на их фактические характеристики. Различные классы задач требуют различной интенсивности обмена с внешними устройствами. Вычислительная система должна обеспечивать построение таких ее конфигураций для каждого конкретного применения, которые бы обладали оптимальными для этого применения характеристиками по вводу-выводу. Система 3М обеспечивает инкрементное наращивание вычислительной мощности до любого необходимого значения путем подключения дополнительных блоков без внесения изменений в имеющуюся систему и ее про-

граммное обеспечение как на этапе разработки системы, так и в ходе ее эксплуатации. Методология проектирования и реализации системы 3М базируется на рассмотрении вычислительной системы как иерархии виртуальных машин. Система 3М имеет рекурсивно-организованную многоуровневую структуры. Рекурсивность структуры состоит в том, что структура всякой модификации системы задается рекурсивным определением. Динамически меняющиеся в ходе вычислений виртуальные процессы требуют постоянной динамической реконфигурации связей между модулями. Сейчас реализуются системы, содержащие тысячи и миллионы процессоров.

В связи с изложенным хотелось рассмотреть проблемы развития вычислительной техники. Вычислительные машины предназначены для решения задач. Общая схема решения задач имеет вид:

Ячел –>Я ос –>Я пр –>Ямаш —>Я рез

где Ячел – формулировка задачи на естественном языке, Яос – формулировка задачи на языке основных соотношений, Япр – формулировка задачи на языке программирования, Ямаш – формулировка задачи на машинном языке, Ярез – формулировка задачи на языке результата в виде графиков, таблиц, изображений, текстов, звуков и т.п. К сожалению, для большинства задач имеется только формулировка на естественном языке, большинство задач плохо формализованы. Поэтому актуальным является переход от описания на естественном языке на язык основных соотношений, лингво-комбинаторное моделирование является одним из способов такой формализации [4,16,21]. В результате такой формализации порождаются рекурсивные структуры со структурированной неопределенностью. Таким образом рекурсивная структура должна включать три составляющих – явления, смыслы и структурированную неопределенность, которые наличествуют в любой задаче. В рекурсивных структурах многократно реализуется принцип обратной связи как для воспроизведения структур так и для порождения новых структур. Рекурсивные структуры являются основой самоорганизующихся структур.

### **Лингво-комбинаторное моделирование планет солнечной системы**

Человечество издревле наблюдало и изучало космос, который условно влиял на развитие человечества. В процессе своего развития человечество создавало самые различные артефакты – здания, дороги, машины и т.д., но самым значительным артефактом, созданным человечеством, является естественный язык, который выбрал и выбирает все знания, навыки, учения, созданные людьми на сознательном и подсознательном уровне. Поэтому очевидно стремление обратиться к естественному языку для того, чтобы глубже понять, как устроена вселенная [17,23].

Если обратиться к моделированию солнечной системы, то в качестве ключевых слов можно взять Солнце, Меркурий, Венеру, Землю, Марс, Юпитер, Сатурн, Уран, Нептун, Плутон – 10 переменных, в структуре эквивалентных уравнений этой системы будет содержаться 45 произвольных коэффициентов, наличие которых определяет возможность управления характеристиками. Выявление этой новой возможности управления важна для человечества ввиду астероидной опасности. Наша планета хранит свидетельства разрушительного воздействия астероидов, падение которых вызывало глобальные катастрофы. В настоящее время налаживается мониторинг околоземного пространства и выявляются все новые и новые аспекты астероидной опасности. Для борьбы с астероидной опасностью планируется использовать всю мощь накопленного ядерного оружия землян, но в случае если размеры астероида будут превышать 10 километров в диаметре, то и это средство не поможет. Единственная надежда – на открытие новых способов управления планетарными процессами. Аналогичным образом возможно моделирование галактик и их взаимодействия. Обратимся к анализу предложенной модели.

Во-первых, когда мы говорим о системе, это значит, что из всего мы выделяем часть – систему и рассматриваем ее взаимодействие с оставшимся, с окружающей средой. Также и с солнечной системой – мы выделили планеты и можем рассматривать как взаимодействие между ними, так и воздействие остального космоса на всю солнечную систему. Солнечная система существует в потоке переменных воздействий остального космоса и ее устойчивость зависит от ее адаптационных возможностей, которые определяются числом произвольных коэффициентов. В данном случае это число 45, в общем случае оно определяется числом сочетаний из  $n$  по  $m+1$ , где  $n$  – число переменных системы,  $m$  – число ограничений, наложенных на систему. Как очевидно из этой формулы, в зависимости от числа наложенных ограничений для числа переменных больше шести количество произвольных коэффициентов будет сначала возрастать, достигнет максимума и потом будет уменьшаться. Это явление в теории систем называется феноменом адаптационного максимума [12,13], в зоне адаптационного максимума система обладает максимальными адаптационными возможностями. Можно предположить, что в процессе эволюции адаптационные возможности солнечной системы изменяются в соответствии с феноменом адаптационного максимума, что можно подтвердить или опровергнуть соответствующими исследованиями. Этот феномен может быть основой для объяснения ритмов развития как солнечной системы в целом, так и солнца в частности, и галактических систем.

Во-вторых, в качестве ключевых слов для построения модели мы взяли сами планеты, которые можно наблюдать, то есть это феноме-

нологическая модель. Люди издревле наблюдали планеты и для объяснения их движения создали целую систему понятий, в частности для объяснения их движения были привлечены законы Ньютона, которые опираются на понятия «сила», «масса», «ускорение», «скорость». Опираясь на эти понятия как на ключевые слова можно нашим методом построить другую лингво-комбинаторную модель, которая бы определила взаимодействие между этими понятиями. Таким образом можно рассматривать два языка – язык феноменологического описания и язык научных теорий и попытаться осуществить перевод с одного языка на другой язык, а можно замешать в общую структуру как сами явления, так и научные понятия, что и определит взаимодействие между ними.

В-третьих, о прямых и обратных задачах. Прямая задача связана с моделированием следствия по заданной причине. В обратной задаче мы хотим восстановить причину по известному следствию. Прямая задача имеет единственное и устойчивое к малым возмущениям решение. Для обратной задачи единственность решения нарушается, т.к. различные причины могут вызвать одно и тоже следствие. Если обратиться к движению планет, которое наблюдается людьми много тысяч лет, то этот феномен может быть по-разному объяснен. Существовала система Птолемея, потом были открыты законы Кеплера, потом законы Ньютона. Если считать, что движение планет объясняют законы Ньютона, то исходя из них можно рассчитать траектории движения планет. Но исходя из других соображений тоже можно получить такое же движение планет, такой же феномен. Эта обратная задача, которая имеет множество решений [25]. Лингво-комбинаторное моделирование позволяет построить множество различных генераторов таких же движений.

В-четвертых, в современной науке и в обществе в настоящее время получили большое распространение понятия «управление», «информация», и они пронизывают биологические науки (от генетики до высшей нервной деятельности), социально-экономические науки, технические науки, но эти понятия в большинстве случаев исключены из физики и астрономии. Конечно, и астрономия, и физика сформировались довольно давно, когда понятия «управление» и «информация» не были разработаны, но в настоящее время вызывает недоумение отсутствие этих понятий при объяснении многих астрономических и физических явлений. Если исходить из этой точки зрения, то насущной задачей астрофизики должен быть поиск центров управления, систем связи и самой возможности управления малыми воздействиями, вызывающими большие последствия в планетарных и галактических системах. В 1949 г. Норберт Винер возродил кибернетику, опубликовав книгу «Управление и связь в животном и машине, но он не стал распространять принципы кибернетики на фи-

зику, остановился перед ней как перед священной коровой. В настоящее время рассматривается вопрос о создании кибернетической физики, в которой вопросы управления займут достойное место. Уже существует гипотеза Гея, в которой рассматривается вся планета Земля как живая система. Можно высказать гипотезу о живой вселенной со своей нервной системой.

В-пятых, еще в Древней Греции Анаксимандром была высказана гипотеза о существовании апейрона – субстанции, которая пронизывает вся и все. В свете развиваемой нами теории апейрон – физическая символическая структура, которая реализует структурированную неопределенность в виде произвольных коэффициентов  $U$ , субстанции  $U$ . Структурированной неопределенности противостоит хаос – неструктурированная неопределенность. Структурированная неопределенность – основа жизни во вселенной.

Лингво-комбинаторное моделирование – это универсальный метод моделирования плохо формализованных систем в самых различных областях науки, техники, в различных областях человеческой деятельности. В каждом конкретном применении этого метода необходимо осуществлять верификацию модели, проверять ее на соответствие поведению реального объекта. Наличие произвольных коэффициентов и возможность расширения модели, возможность включения новых переменных, новых ключевых слов, позволяют настраивать модель для моделирования сложных реальных объектов. Лингво-комбинаторное моделирование позволяет сформировать новую картину мира, которая опирается на все достижения современной науки. Лингво-комбинаторная картина мира состоит из трех групп переменных, во-первых, это явления (Appearances), во-вторых, это смыслы (Essences), в третьих – это структурированная неопределенность (Structural Uncertainty), из которых состоят все неживые и живые системы.

### **Миры реальные и виртуальные**

Уже давно существует разделение на мир реальный и мир воображаемый. Человеческая психика с развитой способностью к воображению позволяла и позволяет осуществлять мысленное моделирование разнообразных ситуаций и тем самым предвидеть последствия от тех или иных действий и предостерегать от опасности. Человек при рождении получает от своих родителей генетическую информацию, которая содержит некоторые сведения об окружающем мире, эти сведения в процессе жизни пополняются за счет взаимодействия с другими людьми, с окружающей обстановкой, через систему образования и средства массовой информации, за счет случайности биографии – места и времени рождения, семьи, окружающей среды и т.д.. В результате у каждого человека формируется свое представление о

мире, которое проявляется в процессе деятельности этого человека и умирает вместе с ним.

В свое время Альберт Камю сказал, что «Мышление – это новое сотворение мира». Каждый человек является носителем своего воображаемого, виртуального мира, который формируется в зависимости от случайностей биографии этого человека. Таким образом, человечество – это множество виртуальных миров. По мере развития общества источники для формирования этих миров изменялись. Сначала рядом с человеком появилась книга, которая необычайно расширила его кругозор, потом появился телефон и телевизор. В настоящее время рядом с человеком появился компьютер, подключенный ко всемирной сети, и каждый человек стал частью гибридного интеллекта человечества. Граница между реальными и виртуальными мирами размыта.

При обсуждении вопросов о виртуальных мирах нельзя обойти проблему сна и сновидений. Это большая и сложная проблема. В свое время Фрейд предложил психоаналитическое представление о сне и сновидениях. Фрейд считал, что источником сновидений служат скрытые сексуальные и агрессивные влечения человека, которые он и удовлетворяет в такой символической форме. Карл Юнг рассматривал их функции гораздо шире – по его мнению, во сне, помимо символов «индивидуального подсознания, наблюдается и коллективное подсознание», а потому сновидения нужно рассматривать как момент жизни, над которым властвует бессознательное, погружая человека в параллельную реальность. В настоящее время складывается научное направление по моделированию происхождения и функционирования сознания человека на основе процессов бодрствования и сна.

Каждый человек в той или иной степени является архитектором виртуальных миров. Появление компьютера дало в руки человека инструмент для создания различных виртуальных миров. Из этой возможности возникла гипотеза о том, что наш реальный мир – это один из миров внутри сверхкомпьютера [29]. Параллельно с этим миром могут существовать и другие миры. Взаимодействие между мирами возможно через апейрон. Каждый из миров является самоорганизующейся системой.

Испокон веков искусство было средством создания виртуальных миров с опорой на человеческое воображение. Некоторые выстраивают отношения с искусством исходя из математики, но искусство гораздо богаче методами, приемами, образами, чем математика, и именно искусство является источником новых идей для информатики. Также и с бионикой – именно биологические структуры являются источником новых идей для техники. В последнее время оформилось новое научное направление – артоника, основанное на изуче-



нии структур и методов искусства и использовании их в различных компьютерных программах и системах.

У каждого человека есть свое представление о мире, которое сложилось на основе невербальной и вербальной информации. Это представление о мире и есть виртуальный мир конкретного человека, который непрерывно дополняется и модифицируется. У другого человека складывается другое представление о мире, складывается другой виртуальный мир. Эти миры частично пересекаются, но никогда полностью не совпадают, так как люди отличаются друг от друга и имеют различный жизненный опыт. В компьютерах вычислительной системы тоже хранятся изображения, звуки, слова, которые составляют компьютерный виртуальный мир. Компьютерные виртуальные миры пересекаются по содержанию с человеческими виртуальными мирами, это пересечение позволяет им совместно общаться и работать [27]. Для того, чтобы человек мог жить и работать в реальном мире, его виртуальный мир должен отражать существенные черты мира реального, иначе он бы не смог ориентироваться в мире реальном, то есть виртуальный мир должен быть таким, чтобы в нем были представлены все существенные элементы мира реального. В психологии используется понятие субъективного мира человека, мы не хотим использовать этот термин, хотя конечно понятия виртуального мира и субъективного мира пересекаются. Могут быть различные предложения по структуре эгоцентричного мира человека.

Во-первых, это население. Ребенок после рождения знакомится прежде всего со своей матерью, потом с другими людьми, которые характеризуются своим здоровьем, генетикой, образованием, занятостью и т.д. В мире формируется информационное общество, островки компьютеризации объединяются в виртуальные миры. Возникли виртуальные миры финансовых банков, здравоохранения, различных производств, возникли виртуальные предприятия в торговле, на транспорте, в образовании. Важной составляющей этих миров является их население – агенты, которые являются аппаратно-программными комплексами и которые выполняют задания своих принципалов – людей, являясь их помощниками и представителями в различных структурах. Агенты объединяются в многоагентные системы. Агенты (иногда их называют аватарами) могут иметь тот или иной облик на дисплее и голос, в зависимости от желания принципала, но самое главное – их функции в том или ином виртуальном мире. Набор этих функций должен быть таким, чтобы агент мог взаимодействовать как с принципалом, так и с другими агентами в конкретном виртуальном мире, добиваясь выполнения задач, поставленных принципалом. Например, в финансовом мире принципал может поручить своему агенту управлять деньгами на своем банковском счете таким образом, чтобы увеличить прибыль, рискуя в заданных рамках.

Агент должен поддерживать образовательный процесс, начиная со школы, и каждый ученик, заканчивая школу, должен уносить диск со своим агентом, который должен помнить все, чему учили в школе, с тем, чтобы помочь выпускнику школы в дальнейшей жизни. Внедрение многоагентной технологии и формирование виртуальных миров существенным образом скажется на социальной структуре общества.

Таким образом, население виртуального мира – это не только образы людей, но и агенты, и аватары.

Во-вторых, пассионарность, устремления людей. Мать стремится защитить своего ребенка. В обществе складываются социальные группы, имеющие разные интересы. Основой пассионарности является свобода воли людей, свобода выбора тех или иных решений конкретных проблем. При этом необходимо учитывать наличие различных ограничений. Социологические опросы позволяют выявлять и изучать устремления социальных групп.

В-третьих, территория. По мере развития ребенка эта территория расширяется – от детской кроватки через освоение комнаты, квартиры, двора, школы, города вплоть до освоения территории всей планеты. Генноинформационные системы подробно описывают территорию в различных масштабах.

В-четвертых, производство, это то, что может делать человек – сначала по обслуживанию самого себя (личная гигиена), а потом других людей – проектировать и производить машины, учить, лечить и т.п. Для производства человек использует различные предметы, артефакты – от бутылочки с молоком и горшочка, через компьютер до автомобиля и ракетоплана. Самый главный артефакт-- естественный язык.

В-пятых, экология и безопасность, которая опирается на врожденные инстинкты самосохранения и включает борьбу с компьютерными вирусами.

В-шестых, финансы, средства эквивалентного обмена.

В-седьмых, внешние связи. Деление на своих и чужих, обмен и обман, торговля и т.д. Внешние связи характеризуются входящими и выходящими потоками людей, информации, энергии и других ресурсов.

Такова предлагаемая нами структура виртуального мира человека, каждый из этих блоков может быть расширен. Аналогичной должна быть структура компьютерных виртуальных миров. Виртуальные миры – это системы со структурированной неопределенностью, что позволяет реализовать иерархию управления как внутреннего, так и внешнего [24].

ЛИТЕРАТУРА

1. А.А.Воронов, А.Р.Гарбузов, Б.Л.Ермилов, М.Б.Игнатьев, Г.Н.Соколов, Ян Си Зен «Цифровые аналоги для систем автоматического управления» Изд. АН СССР, 1960.
2. А.И.Мальцев «Алгоритмы и рекурсивные функции» М., 1965.
3. М.Б.Игнатьев «О совместном использовании принципов введения избыточности и обратной связи для построения ультраустойчивых систем» Труды III Всесоюзного совещания по автоматическому управлению, том 1, изд. АН СССР, 1968г.
4. М.Б.Игнатьев «О лингвистическом подходе к анализу и синтезу сложных систем» Тезисы Межвузовской научно-технической конференции «Техническая кибернетика» изд. МВТУ, М., 1969.
5. М.Б.Игнатьев «Избыточность в многоцелевых системах» Труды IV симпозиума по проблеме избыточности, Л., 1970.
6. М.Б.Игнатьев, Ф.М.Кулаков, А.М.Покровский «Алгоритмы управления роботами-манипуляторами» изд.Машиностроение, 1972г, издание в США Вирджиния пресс, 1973г, третье издание 1977.
7. С. Бритов, М.Б.Игнатьев, Л.А.Мироновский, Ю.М.Смирнов «Управление вычислительными процессами» изд. ЛГУ, 1973.
8. V.Glushkov, M.Ignatyev, V.Miasnikov, V.Torgashev «Recursive machines and computing technology» Proceedings IFIP-74, computer hardware and architecture, p. 65-70, Stockholm, August 5-10, 1974.
9. С.А. Яновская «Методологические проблемы науки» М, 1972.
10. П.Г.Суворова «Диалектика абстрактного и конкретного в понятии «архитектура ЭВМ» Сб. «Новые идеи в философии науки и научном познании». Ред.Ю.И.Мирошников, Екатеринбург, 2002.
11. Моль А. «Социодинамика культуры» изд. Прогресс, Москва, 1973.
12. Игнатьев М.Б. «Голономные автоматические системы» М. – Л., изд. АН СССР, 1963, 204 с.
13. Игнатьев М.Б. Диплом №25-S на открытие в области кибернетики и системного анализа «Закономерность изменения адаптационных возможностей в жизненном цикле сложных развивающихся систем» с приоритетом от 28.06.1963г.
14. Ignatiev M. B. «Simulation of Adaptational Maximim Phenomenon in Developing Systems» Proceedings of The SIMTEC»93 – 1993 International Simulation Technology Conference, San Francisco, USA, 1993, P.41-42.
15. Ignatyev M.B., D.M. Makina, N.N. Petrishev, I.V. Poliakov, E.V. Ulrich, A.V. Gubin «Global model of organism for decision making support» Proceedings of the High Performance Computing Symposium – HPC 2000, Ed. A. Tentner, 2000 Advanced Simulation Technologies Conference, Washington D.C. USA, 2000, P.66-71.
16. Ignatyev M. B. «Linguo-combinatorial method for complex systems simulation» Proceedings of the 6<sup>th</sup> World Multiconference on Systemics, Cybernetics and Informatics, vol. XI, Computer science II, Orlando, USA, 2002, P.224-227.
17. Ignatyev M. B., Pinigin G. I. «Linguo-combinatorial simulation of universe» XXV General Assembly of International Astronomical Union, Sydney, Australia, 2003 [www.astronomy2003.com](http://www.astronomy2003.com)

18. Бейдер Р. «Атомы в молекулах» М, изд. Мир, 2001, 450 с.
19. Игнатъев М.Б. «Самоорганизующиеся робототехнические системы и игра в футбол» Сборник трудов Первой международной конференции по мехатронике и робототехнике», том 2, Санкт-Петербург, 2000, С. 127-131.
20. Игнатъев М.Б. «Семиблочная модель города для поддержки принятия решений» Труды семинара «Компьютерные модели развития города» изд. Наука, СПб, 2003, С. 40-45 .
21. Игнатъев М.Б. «Лингво-комбинаторное моделирование плохо формализованных систем», журнал «Информационно-управляющие системы», №6, 2003, С. 34-37.
22. Игнатъев М.Б. «Новая модель атома с блоком управления» Тезисы докладов Второй международной конференции «Устойчивость и управление для нелинейных трансформируемых систем» Москва, 2000.
23. Игнатъев М.Б. «Лингво-комбинаторная картина мира и познание реальности» Труды Конгресса-2002 «Фундаментальные проблемы естествознания и техники», Серия «Проблемы исследования Вселенной» вып.25, Санкт-Петербург, 2002, стр. 117- 128.
24. Игнатъев М.Б., Тихомиров М.Е. «Внешнее управление самолетами и проблемы измерения» Вестник Северо-Западного филиала Метрологической Академии, вып.12, СПб, 2004, стр. 51- 66.
25. Дмитриев В.И. «О методах решения обратных задач» Вестник МГУ, серия 15 «Вычислительная математика и кибернетика» №4, 2001, стр.3-7.
26. М.Б.Игнатъев «Роботы, аватары и люди как системы со структурированной неопределенностью» Сб. «Новое в искусственном интеллекте» М., 2005.
27. М.Б.Игнатъев, А.А.Никитин, А.В.Никитин, Н.Н.Решетникова «Архитектура виртуальных миров» Монография, СПб, 2005.
28. M.Ignatyev «The study of the adaptational maximum phenomenon in complex systems» Seven International Conference on Computing Anticipatory Systems. Abstract book. Ed.Daniel M.Dubois. HEC – Ulg, Liege, Belgium, August 8-13, 2005, Simposium 2, p.18.
29. М.Б.Игнатъев «Философские вопросы компьютеризации и моделирования» Тезисы докладов межреспубликанского симпозиума «XXXУП съезд а актуальные задачи совершенствования работы философских(методологических) семинаров» Изд. АН СССР, Ленинград, 1987.

---

# ТРАНСЦЕНДЕНТАЛЬНАЯ (КАНТОВСКАЯ) МОДЕЛЬ СОЗНАНИЯ КАК НОВАЯ ПАРАДИГМА «ИСКУССТВЕННОГО РАЗУМА»\*

*С.Л. Камречко*

Предлагаемый подход направлен на преодоление существующего в настоящее время разрыва между прикладными исследованиями в computer science (artificial intelligence) и философскими концепциями сознания (Платон, Аристотель, Кант, Гуссерль). Современная парадигма искусственного интеллекта (ИИ) страдает рядом недостатков, одним из которых является заведомо узкое понимание человеческого Ума, сведение его к дискурсивно-рассудочному интеллекту: именно поэтому мы отказались в названии от узнаваемой кальки «искусственный интеллект». Для дальнейшего развития ИИ необходим переход к более богатой модели человеческого сознания (resp. «естественного интеллекта»). Ее основой могла бы послужить кантовская концепция сознания, где сознание рассматривается как система «познавательных способностей» человека, в которой одно из центральных мест принадлежит *воображению*, этой «слепой, но необходимой функции души» (Кант). В развитой модели сознания также должна быть учтена кантовская *рефлектирующая способность суждения*, которая лежит в основе *метафорического мышления* (*мышления по аналогии*). Все это позволит осуществить синтез «нелогических» механизмов сознания с хорошо развитыми в ИИ средствами логического вывода.

## **Введение. Основные тезисы**

Нас интересует вопрос о том, как протекает познавательный акт, постулирование каких трансцендентальных условий (механизмов работы сознания) для этого необходимы, т.е. тот набор «действий чистого мышления» (Кант), который этот акт фундируют. В качестве общей методологии нашего исследования выбран трансцендентальный подход (Платон, Аристотель, Декарт, Кант, Бергсон, Гуссерль). При этом и вместе с тем необходимо дать не *декларативную*, а, по возможности, *процедурную* интерпретацию этих механизмов с целью последующего создания программных продуктов.

Основой нашего подхода является восходящая к Канту концепция *эпистемологического гилеоморфизма*, в которой познавательный

---

\* Данное исследование поддержано грантом РГНФ № 06-03-00197а

процесс рассматривается как иерархическая система актов *анализа* и *синтеза*: при этом результат предыдущего акта выступает как *материал* для последующего *оформления* синтезом. Сознание же рассматривается здесь как определенная конфигурация познавательных способностей, основополагающими из которых являются *чувственность, воображение и рассудок* (мышление). Эта, восходящая еще к Аристотелю, модель *естественного интеллекта*, достаточно простая с точки зрения современной эпистемологии, не получила процедурного (программного) воплощения в исследованиях по ИИ, которые вплоть до сегодняшнего дня ориентированы в основном на моделирование рассудочной — *интеллектуальной* — деятельности, что нашло выражение и в его названии как *искусственного интеллекта*. Основная трудность, возникающая при искусственном воссоздании такой модели *естественного интеллекта*, состоит в том, что необходимо найти опосредующие «механизмы» (типа кантовского *схематизма*) между *разнородными* познавательными способностями, которыми, по Канту, являются чувственность и рассудок.

Предваряя дальнейшее изложение, отметим следующее. Аристотель и неоплатоники (Прокл), а впоследствии и мыслители Нового времени и Кант, рассматривают *воображение* («общее чувство», фантазию) как такой модус чувственности, который является опосредующим звеном между *чувственностью* и *рассудком* как «двумя основными стволами» человеческого познания [см., например, трактат Аристотеля «О душе»]. Мы же рассматриваем воображение как *самостоятельную фундаментальную способность* сознания, точнее как тот «общий корень», из которого вырастают «...крайние звенья чувственности и рассудка» [1, 513]. Понятно, что такая трактовка воображения несколько усложняет задачу, т.к. теперь нужно выявить опосредующие механизмы между 1. Чувственностью и воображением и 2. Воображением и рассудком.

Исходной моделью сознания//познания нами выбрана кантовская модель познавательного акта, который представляет собой следующую последовательность синтезов: 1. синтез схватывания; 2. синтез апперцепции; 3. фигурный (пространственный) синтез; 4. схематический (временной) синтез; 5. образно-понятийный синтез « $E_{to_j}$  —  $A_i$ » (здесь « $E_{to_j}$ » — образ, а « $A_i$ » — понятие); 6. пропозициональный синтез « $A_j$  —  $B_i$ » (здесь, « $A_j$ » и « $B_i$ » — понятия). С одной стороны, это более развитая по сравнению со схемой «чувственность — рассудок» модель, с другой — она сформулирована в более «программистском» стиле. В эту модель мы вносим две новации. Во-первых, мы обогащаем кантовский схематизм путем выделения в его составе как самостоятельного момента (3.1) способности к *рефлексивному переключению*. Во-вторых, мы вводим новый (4\*) *образно-образный синтез (метафорический синтез)*, который выступает как альтернатива

*образно-понятийному синтезу* рассудка. Именно эти два механизма и лежат, на наш взгляд, в основе творческой деятельности (способности) человека.

Первый из вводимых нами «механизмов» (3.1) осуществляется при переходе от 3 к 4 синтезу и связан с тем, что при «прорисовке» фигуры сознание *переключает* свое внимание на *способ* прорисовки и, тем самым, выявляет *схему* построения фигуры как то «общезначимое созерцание» (Кант), которое выступает посредником между единичным образом чувственности и общим (дискурсивным) *понятием* рассудка.

Второй из них (4\*) выступает либо как альтернатива 4 синтеза, либо как самостоятельный акт синтеза, который осуществляется между 4 и 5 синтезами. Он связан с тем, что при *узнавании* созданного воображением образа (3) наше мышление может осуществлять не только рассудочный синтез, связанный с соотносением образа «Это<sub>3</sub>» с понятием «А<sub>3</sub>», но и *метафорический синтез*, когда образ «Это<sub>3</sub>» соотносится с более общим образом «Это<sub>4</sub>». В определенном смысле этот акт представляет собой тот же кантовский схематизм, однако в данном случае наше внимание не «переключается» на способ (алгоритм) прорисовки фигуры, а, осуществляя абстрагирование, переходит к более общему образу, например от детального (фотографического) рисунка человека к его схематическому изображению как «крестика (=туловище с двумя руками и ногами) с кружочком (=голова) наверху». При этом происходит то, что М. Мамардашвили удачно назвал «мышление образами»: например, для Сезанна — это «мышление яблоками». В основе метафорического синтеза лежит кантовская *способность рефлектирующего суждения*, которая в своем развитом виде приводит к созданию «эстетических идей» [разума], не выразимых с помощью дискурсивных механизмов рассудка [2]. Одним из центральных механизмов этого синтеза является способность человека к изменению масштаба своего рассмотрения. Например, при рассматривании офорта Дюрера (Гуссерль) мы можем «одновременно» (на самом деле, последовательно) видеть (1) отдельные цветные пятна — модус чувственности; (2) *образы* картины (рыцаря на коне...) — модус собственно воображения; (3) картину как таковую — модус рассудка.

В логическом плане изменение масштаба рассмотрения может быть соотносено с переходом к метаисчислениям, который в рамках метафоры «правое — левое полушарие» связан с «правополушарной» деятельностью (см. [3]). Введение в этом метаисчислении новых *понятий*, невыразимых в более бедном языке исходного исчисления, позволяет повысить эффективность работы системы ИИ. Проиллюстрируем это на одном примере. Пусть нам дано исчисление с аксиомами:  $A \equiv A$ ,  $(A \equiv B) \equiv (B \equiv A)$ ,  $(A \equiv (B \equiv C)) \equiv ((A \equiv B) \equiv C)$  и правилом вывода:  $\{A[B], B \equiv C\} \Rightarrow A[C]$  и надо доказать формулу *W*:

$(P \equiv Q) \equiv ((Q \equiv R) \equiv (R \equiv P))$ ). Как утверждается в [4], вывод формулы  $W$  в исходном исчислении занимает около двух страниц, но вывод становится тривиальным (эффективным) с помощью метаправила «формула, содержащая четное число каждой из своих переменных является теоремой» (в данном метаправиле фигурирует понятие «четность», не выразимое в языке исходного исчисления). Теперь для вывода можно просто посчитать четность переменных искомым формул: в частности, выводимой является наша формула  $W$ , т.к. она содержит четное количество каждой из своих переменных.

Особо следует подчеркнуть, что приведенный выше набросок модели сознания является моделью именно *естественного*, т.е. человеческого, интеллекта. Его важной структурной особенностью является то, что человеческий рассудок, осуществляющий абстрактное мышление, связан с внешним миром не напрямую, а посредством чувственности: соответственно, здесь возникает проблема перехода от чувственных созерцаний к рассудочным понятиям. Тем самым в кантовской концепции отвергается возможность так называемой *интеллектуальной интуиции*, т.е. непосредственных интеллектуальных созерцаний, к которым относятся *понятия* рассудка и *идеи* разума. Заметим при этом, что мы, вслед за Кантом, не отрицаем принципиальную возможность другого типа Интеллекта, например божественного *интуитивного рассудка* (*intellectus archetypus*: см. § 77 «Критики способности суждения» Канта [2]), который способен к непосредственному, без помощи чувственности, созерцанию идей. Более того, вполне возможно, что *искусственный интеллект* и должен быть отличным от *естественного*, причем отличен в принципе, т.е. иметь другую конфигурацию познавательных способностей. Однако в данном тексте мы исходим из того, что описание структуры («конфигурации») *естественного интеллекта* кантовского типа может быть полезно при создании систем *искусственного интеллекта* и/или повышения их эффективности.

### **Кантовская концепция сознания как модель «искусственного интеллекта»**

«Новое время» европейской философии по праву начинается с фигуры Декарта, который открывает новый регион философского исследования — область *cogito*. Следующий после Декарта существенный шаг в этом направлении делает Кант, предложивший *структурную модель сознания*. Постулируемая *структурность сознания* снимает тезис о качественной простоте (однородности) сознания, который в той или иной мере разделялся его предшественниками: Декартом, Локком, Лейбницем, Юмом. Тем самым Кант предлагает модель *качественно разнородного сознания*, состоящего из пассивной *чувст-*



венности и активного рассудка<sup>1</sup>. При этом сознание рассматривается им как особая активность. Взяв за образец механику Ньютона, Кант выделяет как бы основные *силы* (способности) Ума и строит своеобразную *механику* сознания.

В общем виде под *способностью сознания* понимается то, что позволяет осуществлять взаимосвязь между субъектом и объектом, т.е. те *способы* активности субъекта, которые позволяют ему выходить за свои пределы. Соответственно, объект в данной (субъектно-объектной) взаимосвязи фиксируется как *пред-стоящее* субъекту, т.е. как *предмет*; а данность предмета сознанию — как *представление* (*repraesentatio, Vorstellung*). *Способности* мыслятся Кантом как *способы* образования разных *представлений*, поэтому между *способностями* и *представлениями* существует тесная связь<sup>2</sup>. Кант выделяет способности *чувственности/воображения* (*чувственность* в широком смысле) и *рассудка/суждения/разума* (*рассудок* в широком смысле). Разные конфигурации (сочленения и подчинения) элементарных способностей образуют три мета-способности: *теоретический разум* (мета-способность познания), *практический разум* (мета-способность воли) и *эстетический разум* (мета-способность чувства удовольствия/неудовольствия). Каждой из них посвящена соответствующая кантовская Критика. Нас здесь будет интересовать *познавательная мета-способность* под «законодательством рассудка».

<sup>1</sup> Точнее, Кант развивает, восходящее к Мендельсону, разделение на *рассудок, чувство и волю*. Но для (наших) целей познания определяющим является различие между *рассудком* и *чувственностью*.

<sup>2</sup> На наш взгляд, именно *классификация типов представлений* является ключом к пониманию кантовского различения *способностей познания*. Кант пишет:

«Ведь у нас нет недостатка в обозначениях, вполне соответствующих каждому виду представлений... Вот их градация. Представление вообще (*repraesentatio*) есть род. Ему подчинено сознательное представление (*perceptio*). Ощущение (*sensatio*) есть перцепция, имеющая отношение исключительно к субъекту как модификация его состояния; объективная перцепция есть познание (*cognitio*). Познание есть или созерцание, или понятие (*intuitus vel conceptus*). Созерцание имеет непосредственное отношение к предмету и всегда бывает единичным, а понятие имеет отношение к предмету опосредствованно, при посредстве признака, который может быть общим для нескольких вещей. Понятие бывает или эмпирическим, или чистым; чистое понятие, поскольку оно имеет свое начало исключительно в рассудке (а не в чистом образе чувственности), называется *notio*. Понятие, состоящее из *notiones* и выходящее за пределы возможного опыта, есть идея, или понятие разума. Для тех, кто привык к такому различению, невыносимо, когда представление о красном называют идеей. На самом деле это представление не есть даже *notio* (рассудочное понятие)» [1, 229].

Познание по Канту имеет двухступенчатый характер и представляет собой рассудочное *оформление* чувственной *материи*. В общем виде оно задается формулой: «*знание (опыт) = чувственная материя + рассудочная форма*», в которой чувственность составляет первую (низшую), а рассудок — вторую (высшую) ступени познания. В точном смысле, «[перво]материей чувственного знания» выступает *ощущение* [1, 70], которая оформляется априорными формами уже на ступени чувственности. Но данная формула передает важную интенцию Канта, который мыслит познание как последовательный ряд этапов *оформлений*, в которой главенствующую роль отведена рассудку, а материалом («материей») для последующих оформлений выступают результаты предшествующих этапов. Тем самым Кант развивает своеобразную концепцию *эпистемологического гилеоморфизма* (ср. с аристотелевским онтологическим гилеоморфизмом).

Перейдем теперь к более подробному описанию познавательного процесса. Собственно (элементарный) познавательный акт имеет по Канту сложную структуру и представляет собой последовательность (иерархию) *синтезов*<sup>3</sup>. Началом познания является акт чувственного восприятия, в результате которого *внешняя* для сознания «вещь в себе» *вос-принимается* вовнутрь, т.е. помещается на *внутренний экран сознания* и, тем самым, превращается в «вещь для нас». Вся дальнейшая наша познавательная активность разворачивается на этом экране. При этом *объект познания* одомашненный в *вещь для нас* лишается своей активности (до вос-приятия *вещь в себе* является если не активной, то, по крайней мере, «дикой» и непредсказуемой для нас) и превращается в пассивную *материю* последующих оформлений; а *субъект* познания — изменяет свою интенциональность: он как бы отворачивается от внешнего мира и обращается к работе со своими собственными представлениями<sup>4</sup>. В зависимости от пропускной способности органов восприятия и мощности нашей «оперативной памяти» на экране сознания располагаются вос-принятые пер-

<sup>3</sup> Основными процедурами познания по Канту выступают *анализ* и *синтез*. С помощью *анализа* возможно *разлагать* (преобразовывать) имеющееся знание, при этом не получая ничего принципиально *нового*. Более богатое — новое — знание, по Канту, образуется в результате *синтетических актов*. Именно поэтому кантовское внимание сосредоточено на синтезе как процедуре «приращения» знаний. В данном тексте мы также центрируем свое внимание на описании синтетических актов.

<sup>4</sup> Отметим, что мы воспроизводим кантовское описание познавательный процесса, где сознание после первоначального акта чувственного восприятия как бы помещает себя в клетку своих собственных построений. Именно к этой фазе познания относится феноменологическая максима «Всматривайся!» (resp. «Назад к самим вещам!»), которая призвана предотвратить подобное «отворачивание».

цепции, которые, с одной стороны, отличены друг от друга, а, с другой стороны, из-за их совместного расположения в этой однородной среде могут теперь соотноситься друг с другом и синтезироваться в различные типы *представлений*.

Заметим, что именно здесь проходит водораздел между *чувственностью* как пассивной восприимчивостью и последующей активностью сознания, к которому относится способность *воображения*: если чувственность направлена *вовне*, то воображение направлено *вовнутрь*, на обработку схваченного в некоторое единство созерцания *Это<sub>x</sub>*. В результате этого *Это<sub>x</sub>* *оформляется* сначала чувственными априорными формами, а затем — априорными формами рассудка. Однако обо всем по порядку.

Следующим после восприятия осуществляется один из важнейших в процессе познания *синтез схватывания* как тройственный акт (1) схватывания эмпирического содержания, (2) апперцептивного синтеза и (3) первоначально-синтетического единства апперцепции. Последовательность протекания этих актов в ходе конкретного (эмпирического) познавательного акта, на наш взгляд, должна быть именно такой, как мы перечислили выше. Однако в кантовском трансцендентальном анализе их порядок изменен. Логически первым по Канту является синтез *единства чистой* апперцепции (соответственно, именно его Кант и называет его *первоначальным*): на экране сознания выделяется особое представление «Я [мыслю]» как некоторый центр («начало координат»), относительно которого координируются все остальные — воспринятые извне — представления. Вместе с тем единство чистой апперцепции — *самосознание* — выступает как *наблюдатель* происходящего на экране сознания, который *темпорально* структурирован (ср. с гуссерлевским «временем-сознанием» [Zeitbewußtsein]), на котором располагаются *промаркированные* временными метками представления *Это<sub>1</sub>*, *Это<sub>2</sub>*, *Это<sub>3</sub>* (заметим, что *время* Кант характеризует как «чистый образ всех предметов чувств вообще» [1, 125])<sup>5</sup>. По этой логике, вторым выступает *акт чистой апперцепции* (resp. классическая апперцепция), который заключается в маркировке любой перцепции как моей, *в простом сопровождении всякого представления сознанием* (парафраз Канта): синтетичность этого акта заключается в соединении схваченного *Это<sub>x</sub>* с представлением «Я мыслю». Собственно *синтез схватывания* состоит в том, что воспринимаемое чувственное *многообразие* синте-

<sup>5</sup> В своих «Лекциях по метафизике» Кант особо выделяет способность *маркирования* (обозначения), которая позволяет «связывать представления не непосредственно с их предметами, но посредством некоего заместителя [друг с другом]» [5, 146]. Тем самым *маркирование* — это координация представлений, благодаря чему они могут быть соотнесены друг с другом в актах синтеза.

зируется в некоторый *единый* протообраз, который мы и обозначили как  $Это_x$ , т.е.  $Это_x$  выступает как *со-единение* отдельных точечных перцепций  $Это_1$ ,  $Это_2$ ,  $Это_3$ . Опишем акт схватывания как эмпирический акт познания. В ходе схватывания, воспринятые чувственные ощущения — *материя* познания, структурируются в *протообраз*  $Это_x$ ; при этом чистый синтез апперцепции соединяет  $Это_x$  с представлением «Я мыслю», а благодаря первоначально-синтетическому единству апперцепции, схватываемое (схваченное) в  $Это_x$  многообразии получает *форму* единства. Здесь *апперцепция* и *схватывание* выступают скорее как две стороны единого синтетического акта: схватывание выступает как *деятельный акт* синтеза, а апперцепция как *единающая форма* этого акта.

Результатом работы синтеза схватывания является образование на экране сознания пространственно-временного *образа*. В его формировании можно выделить две стадии. Началом формирования образа является собственно *синтез схватывания*, который соединяет чувственное *многообразие* в некоторый *единый* протообраз (см. его описание выше). Он представляет собой *бытийно-качественный* акт, задача которого схватить бытие некоторого предмета как  $Это_1$ , отличая его от других  $Это_x$ . Далее наступает черед *формальных синтезов*, т.е. тех синтезов, которые осуществляются с помощью имеющихся в сознании априорных форм. Построение собственно *образа* осуществляется с помощью *формального пространственно-временного синтеза*, который представляет собой последующий *количественный* акт, задача которого обеспечить условия для последующего — *математического* — изучения схваченного представления<sup>6</sup>.

Поскольку в «Критике чистого разума» внимание Канта приковано к обоснованию априорного характера пространства и времени, то протекание пространственно-временного синтеза описано лишь в общих чертах (заметим, что проясняет суть дела последующее описание *фигурного синтеза*). В результате этого синтеза схваченное  $Это_x$  превращается в пространственно-временной образ. Это достигается путем со-единения воспринятого *содержательного представления*  $Это_x$  и имеющихся в нашей душе *чистых форм//созерцаний* пространства и времени. При этом *пространство* является *формой* внешних явлений, а *время* — *формой* как внешних, так и внутренних явлений.

В этой связи обратим внимание на то, что возможно двоякое понимание априорности пространства и времени, причем в текстах са-

<sup>6</sup> Соответственно, задача последующих формально-понятийных синтезов заключается в раскрытии *сущностных* отношений схваченного предмета, главным из которых являются причинно-следственные связи (ср. с гегелевским категориальным рядом *бытие — качество — количество — сущность...*).

мого Канта нет однозначного ответа на этот вопрос. Речь идет о том, являются ли пространство/время актуальными априорными *созерцаниями* или же они являются всего лишь *формами созерцаний*, т.е. не содержатся в нашем сознании как самостоятельные представления. В своих подготовительных материалах к «Критике чистого разума» Кант, дистанцируясь от критикуемой им концепции *врожденных идей*, говорит о пространстве /времени лишь как о *формах*, которые сами по себе не являются *представлениями* и актуализируются лишь во время чувственно-созерцательного акта. Отголоски подобного решения можно найти и в тексте «Критики чистого разума»: «чистые пространство и время, которые, правда, как формы созерцания суть *нечто*, но сами они не предметы [представления. — К.С.], которые можно созерцать» [1, 213]. Вместе с тем в окончательном тексте этого произведения Кант скорее склоняется к постулированию актуально-созерцательного характера пространства и времени: «пространство и время, а priori представляются не только как *формы* чувственного созерцания, но и как сами *созерцания*...» [1, 115]. Понятно, что второе, более сильное, решение позволяет выстроить стройную концепцию синтеза как простой «суммы» (resp. *со-единения*) двух или нескольких представлений: «под синтезом в самом широком смысле я разумею присоединение различных представлений друг к другу и понимание их многообразия в едином акте познания» [1, 85].

Для процедурной интерпретации кантовской модели сознания, на наш взгляд, целесообразнее принять более слабое решение (понятно, что при этом мы также модифицируем (расширяем) кантовскую процедуру синтеза как *любое привнесение нового содержания*, а отнюдь не только как *со-единение*). В этом случае можно сказать, что пространственно-временной образ образуется в сознании путем трансформации *Это<sub>x</sub>*, или происходит *модификация* исходного созерцания *Это<sub>x</sub>* с помощью пространственно-временного фильтра: чувственное созерцание как бы проходит через надетые на нас пространственно-временные очки, которые модифицируют поступающие ощущения в пространственно-временной образ. Если воспользоваться идущей от Аристотеля метафорой познания как оставления отпечатков внешних восприятий на мягком воске нашей чувственной души, то пространство и время являются не свойствами вещей, а характеристиками *среды* экрана сознания, на котором отражаются внешние вещи. Хотя скорее у Канта, в точном смысле этого слова, *средой* является лишь пространственность, которая выполняет роль *интеллигибельной материи* для последующих синтезов схваченных чувственных содержаний, в то время как *время* рассматривается скорее не как фоновая среда (хотя этот момент у Канта присутствует), а в большей степени как *способ осуществления последовательных синтезов* конечным

субъектом познания<sup>7</sup>, благодаря которым в пространственной среде строятся *фигуры*.

Собственно говоря, мы вплотную подошли к описанию *фигурного синтеза* [synthesis speciosa] *продуктивной способности воображения*, который Кант называет также *трансцендентальным синтезом воображения* [1, с. 110—111]. Этот синтез можно рассматривать как частный случай пространственно-временного синтеза, точнее как *пространственный синтез*, который осуществляется при визуальных восприятиях предметов внешнего мира. С его помощью наше сознание может, например, нарисовать «очертания дома... в пространстве» [1, 115], т.е. некоторую *фигуру* как пространственный образ предмета. Здесь Кант опирается на классическую трактовку воображения, восходящую к Платону. В диалоге «Филеб» Платон соотносит познающую душу с книгой и проводит различие между рассудком и воображением: в душе помимо *писца*, который делает в записи книге, т.е. дискурсивного рассудка, «в то же самое время обретается и другой мастер» — *живописец*, «который вслед за писцом чертит в душе образы названного» [6, 38e—39c]<sup>8</sup>. Платоновское понимание воображения как создания *образов* [εἰχόνας] живописцем в настоящее время прочно закрепилась на уровне обыденного сознания. Например, в русском языке есть выражение «представьте себе», или «вообразите себе». Что делает человек, когда слышит эти слова? Платоновский живописец, обитающий в душе, начинает рисовать в сознании картину происходящего, т.е. осуществлять *фигурный синтез пространственного* оформление схваченного ранее протообраза *Этох*.

В составе *фигурного синтеза* Кант выделяет два модуса. Выше мы говорили об *эмпирическом модусе воображения*, продуцирующем *образы*. Вторым модусом *фигурного синтеза*, точнее «надстройкой» над *пространственным синтезом*, выступает *чистый фигурный синтез* (*трансцендентальный синтез воображения*), продуктом которого является «схема чувственных понятий» [с. 125]<sup>9</sup>. Суть *схематического синтеза* состоит в том, что при рисовании *образа*, происходит не детальное прорисовывание какого-либо единичного предмета, а

<sup>7</sup> Для Канта это положение принципиально: *конечный субъект*, каковым является человек, может осуществлять только *последовательный синтезы*, парадигмальным случаем которого выступает последовательное, точка за точкой, построение прямой линии или окружности: «мы не можем мыслить линию, не *проводя* ее мысленно, не можем мыслить окружность, не *описывая* ее...» [1, 112].

<sup>8</sup> Правда последующая традиция, начиная с Аристотеля, изменяет порядок расположения структур «писца» и «живописца»: сначала — воображение («живописец»), а уже потом — рассудок («писец»).

<sup>9</sup> По Канту, в ходе осуществления этого синтеза задействована также *способность суждения*.

создание его схематического наброска – *схемы*, которую «следует отличать от *образа*» [1, 124]. Другими словами, наше воображение рисует здесь «*форму* [Gestalt] четвероногого животного [собаки] в *общем виде*» [1, 125; В180, 20–25: перевод мой. — К.С.]

Пониманию кантовского перехода от *фигурного* к *схематическому* синтезу может помочь этимология греческого слова «σχήμα», которое помимо значения «наглядный вид» имеет также значение «математическая *фигура*». А математики (геометры), как это показал уже Платон, свои выводы делают, «когда они пользуются чертежами... [для тех идеальных] фигур, подобием которых чертеж служит» [6: Государство, 510d–e).

В этом акте чистое воображение синтезирует «*не единичное созерцание*», а *схему* как *возможный*, т.е. обще-схематический, *предмет созерцания* (парафраз Канта). Понятно, что эмпирические восприятия могут породить не совпадающие друг с другом «единичные частные облики» [1, 125], или образы одного и того же предмета. В частности, при восприятии дома я могу синтезировать его образ различными (эмпирически-случайными) *последовательностями*, например: сверху/вниз или справа/налево. Т.е. эмпирическое воображение (как связка: синтез схватывания — фигурный синтез) порождает лишь *субъективное*, зависящее от точки наблюдения, единство объекта в образе. *Чистое же воображение* порождает уже не частный образ, а *схему* (resp. «форму») *всех возможных* фигур в пространстве как «*объективное единство данных представлений*» [1, 105], совместимых с множеством частных созерцаний *Это<sub>1</sub>, Это<sub>2</sub>...*

Как говорит Кант, схема — «это скорее представление о методе [или общем способе]» [1, 124] *образ-ования* того или иного образа, или другими словами, схема — это *общий алгоритм* построения фигуры как *возможного образа* созерцания. Например, схема треугольника приложима «ко всем треугольникам — прямоугольным, остроугольным и т.п., [хотя сама схема] есть нечто такое, что нельзя привести к какому-либо [конкретному] образу» [1, 125]. Например, схема треугольника — это правило построения данной пространственной фигуры путем двойного замкнутого излома в одном направлении при проведении линии, благодаря чему мы можем отличить треугольник от, например, четырехугольника, который порождается тройным изломом (resp. схема собаки — это правило рисования четвероногого животного с хвостом, ушами и немного вытянутым носом)<sup>10</sup>.

<sup>10</sup> Схематический синтез у Канта четко не прописан, поэтому для его реконструкции мы использовали контекст §§ 24–26 [1, гл. «О схематизме чистых рассудочных понятий»]. Собственно с этой реконструкцией кантовского схематизма и связаны вносимые нами новации, о которых мы упоминали выше.

Каков механизм образования *схем*? В данном случае воспользуемся соображениями Канта из «Критики способности суждения»<sup>11</sup>, где, фактически, описывается процедура, которая у Гуссерля в «Логических исследованиях» называется *варьирующей*. Ее суть состоит в том, что воображение строит не столько «единичный частный облик» вещи, сколько *обобщенный* образ, путем наложения множество образов, совместимых с данным единичным созерцанием [2, 103]. При этом воображение *варьирует* характеристики созданного единичного образа–фигуры: например, варьируя величину углов и/или размеры созерцаемого треугольника, и создает образ–схему «*возможного* для нас [предмета] созерцания» [1, 110].

Из приведенного описания схематического синтеза можно извлечь два следствия.

Во-первых, схема выступает как общее созерцание (которое, в свою очередь, также можно обобщать и далее), что указывает на ее родство с рассудочными понятиями. Тем самым схема как *общий образ* может быть приложима ко многим *единичным* образам. А это означает, что возможен не описанный в кантовской иерархии *метафорический синтез*, когда единичный образ «*Это*» соотносится с более общим образом «*Это*».

Во-вторых, одним из существенных моментов кантовского схематизма является то, что в ходе порождения *схем* сознание осуществляет *рефлексивное переключение*, которое заключается в том, что при выделении общего «ядра» множества фигур путем их прорисовки сознание как бы переключает свое внимание с *содержания* образа на сам *акт* прорисовки, в результате чего и выделяется *схема* как *метод* их построения. В отличие от *пространственного* синтеза образования *схем* имеет временной характер, т.е. схематический синтез является *временным синтезом*.

Следующий этап познания связан с переходом от *чувственных созерцаний* к *дискурсивному рассудку*, который работает со своим типом представлений — *понятиями*. Рассудочные синтезы завершают процесс познания. Их первая задача — *узнать* в построенном *образе* соответствующее ему *понятие* и сформировать протосуждение вида

<sup>11</sup> В «Критике чистого разума» Кант предлагает другое решение проблемы генезиса *схем*: *схема* есть *процедурная* развертка *понятия*, которое вполне пригодно для решения проблемы связи *чистых рассудочных понятий* и эмпирических созерцаний. Нас же интересует здесь проблема генезиса *схем эмпирических* понятий, фигурирующих в составе нашего опытного знания. Поэтому указанный механизм «сверху вниз» здесь не работает. Подход, предложенный Кантом в «Критике способности суждения», позволяет объяснить происхождение не только *схем*, но и образование эмпирических понятий, «снизу вверх», т.е. путем обобщения–оформления схваченных на первом этапе познания чувственных *образов*.



« $Это_j — A_i$ », где  $Это_j$  — образ (например, стола), а  $A_i$  — понятие (resp. понятие стола). В ходе *понятийного* синтеза в *образе* через *схему* рассудок должен узнать соответствующее этому образу *понятие*: например, сформировать протосуждение «Это — дом». В ходе последующих — *пропозициональных* — синтезов уже из имеющихся *понятий* рассудок образует *предложения* субъектно-предикатного типа « $A_j (S) — B_i (P)$ », где  $A_i$  и  $B_i$  — понятия. Например, таким суждением является «Этот дом — кирпичный», которое и завершает собой элементарный познавательный акт.

Подведем итог. В элементарном (эмпирическом) познавательном акте, представляющем собой иерархию синтезов, можно выделить следующие структурные моменты:

***первоначальный синтез схватывания // синтез апперцепции;***

***фигурный (пространственный) синтез;***

***схематический (временной) синтез;***

***4\* обобщающий фигурный синтез:*** образование общих схем — образов (как основа для метафорического синтеза);

***образно-понятийный синтез « $Это_j — A_i$ »;***

***заключительный пропозициональный синтез « $A_j — B_i$ »***

#### ЛИТЕРАТУРА

1. *И. Кант*. Критика чистого разума. — М.: Мысль, 1994.
2. *И. Кант*. Критика способности суждения. — М.: Искусство, 1994 // [Соч. в 6 тт. — Т.5, с.434 — 438]
3. *Маслов С.Ю.* Теория дедуктивных систем и ее применения. — М.: Советское радио, 1986.
4. *Weyhrauch R.W.* Prolegomena to a theory of mechanized formal reasoning //Artificial Intelligence, 1980, Vol. 13
5. *И. Кант*. Рукописные материалы, 2001 // Opus postumum
6. *Платон*. Сочинения. Т. 3, с. 42–43

---

# V

## Логические и математические вопросы

---

---

### ЭПИСТЕМОЛОГИЧЕСКИЙ АНАЛИЗ КОГНИТИВНОЙ ПРИРОДЫ МАТЕМАТИЧЕСКИХ И ЛОГИЧЕСКИХ ФОРМАЛИЗМОВ\*

*И.П. Меркулов*

Открытие когнитивных типов мышления и интенсивное развития за последние десятилетия когнитивной науки (включая исследования в области искусственного интеллекта) создали достаточно надежные, экспериментально обоснованные теоретические предпосылки для разработки принципиально новых эпистемологических представлений, касающихся когнитивной природы формальных наук – математики и логики. Благодаря выявлению присущей человеческому мозгу информационной активности, проявляющейся в реализации его правым и левым полушариями различных доминирующих стратегий переработки когнитивной информации, стало ясно, что математика и логика имеют *непосредственное* отношение не к каким-то структурам внешнего мира, а к работе когнитивной системы человека – к структурам управляемого нашим символьным (вербальным) сознанием знаково-символического (логико-вербального) мышления (функционирующего, естественно, в кооперации с мышлением пространственно-образным), к его способности генерировать идеальные понятия и концептуальные системы, к аналитическим стратегиям этого мышления, которые мы можем конструктивно оптимизировать. С учетом этого математику, видимо, вполне оправданно рассматривать не как науку о числе, пространстве и т.д., а как *науку об идеальных математических структурах, специальных формальных структурах нашего (неречевого) знаково-символического мышления*, свойства которых она описывает с помощью аксиоматических теорий.

---

\* Исследование проведено при финансовой поддержке РФНФ грант № 03-03-00092а

В формальные структуры математики включены такие идеальные (обозначаемые символами и их последовательностями) концептуальные объекты, как, например, числа, множества, группы, функции, векторы, операторы, матрицы, интегралы и т.д., с которыми могут осуществляться формально заданные операции – сложение, умножение, преобразование, композиция, интегрирование и др. Независимо от вида символической репрезентации концептуальных объектов – будь то теоретико-множественные, алгебраические и пр. символы, либо визуально представляемые символичные изображения в виде геометрических фигур, графиков и т.д. – математика конструирует свои формальные структуры только с помощью своих собственных гипотез и правил преобразования (а также формальных и неформальных логических правил)<sup>1</sup>. Разумеется, это не означает, что математическое познание, являющееся продуктом эволюции знаково-символического мышления и символического сознания, совершенно не использует ресурсов пространственно-образного мышления. В силу межполушарной кооперации и «разделения труда» именно пространственно-образное мышление обеспечивает (кроме всего прочего) наше общее *целостное понимание* смысла математических и логических формализмов. Оно также способно манипулировать образными репрезентациями в воображаемом идеальном трехмерном математическом пространстве.

---

<sup>1</sup> Можно спорить с неинтуиционизмом по вопросу о том, является ли геометрия самостоятельной областью математики (если, конечно, не считать самого понятия континуума, вытекающего из их представления о «свободно становящихся последовательностях») и сводится ли она к анализу или нет. Но каким бы не был итог этого спора, он не может изменить эпистемологической природы положений математики. Использование наглядных символических репрезентаций в математике не означает, что манипулирование ими в воображаемом идеальном математическом пространстве подчинено генетически направляемым *неартикулированным* холистическим стратегиям пространственно-образного мышления. Воображаемое математическое пространство генерируется нашим пространственно-образным мышлением благодаря развившемуся в ходе когнитивной эволюции людей управлению со стороны символического (вербального) сознания и доминирующего знаково-символического мышления. В результате появляется возможность использовать аналитические стратегии, адаптированные к нуждам визуального оперирования с перцептивными мысленными репрезентациями – символическими изображениями, графиками, схемами и т.д., которые поддаются разложению на более простые элементы. Переход от «аналитических» к «геометрическим» репрезентациям и обратно дает огромные когнитивные преимущества в математике, но этот переход не равнозначен некоему «переходу» от знаково-символического мышления к мышлению пространственно-образному.

Как формальные системы математические теории непосредственно не приложимы к внешней, «внемыслительной» реальности, и ничего о ней не говорят. Но они применимы к этой реальности *опосредованно* – через применение к идеальным понятиям и концептуальным системам, создаваемым нашей когнитивной системой, в которых заключены наши эмпирически проверяемые знания, зафиксированы теоретические допущения и гипотезы эмпирических наук. Математические формализмы позволяют извлечь из идеальных объектов эмпирических дисциплин потенциально содержащуюся в них концептуальную информацию, т.е. новые знания о природных и социальных явлениях. Исследуя функционирование механизмов, технических устройств и т.п. с помощью математических моделей, мы можем вывести из них, вычислить ранее неизвестную концептуальную информацию, касающуюся их поведения в различных ситуациях, сделать соответствующие расчеты, позволяющие улучшить их конструкции, их производительность, эффективность, экономичность и т.д. Разумеется, научные и технические знания, полученные благодаря применению математических формализмов, подлежат эмпирическим (экспериментальным) проверкам, которые могут их подтвердить или опровергнуть. Но это не означает, что вместе с этими знаниями подобной эмпирической проверке подвергаются математические формализмы, обеспечивающие их выведение. В силу своей независимости от эмпирического опыта, относящегося к «внешней» реальности, они не могут быть с его помощью доказаны или опровергнуты.

Математические формализмы не являются частью физических, химических, астрономических и пр. гипотез, они «нейтральны» по отношению к их содержанию и сами по себе не обладают специально-научным эмпирическим смыслом (интерпретацией). Математические формализмы могут быть частью только математических теорий. Но почему мы тогда уверены, что математические формализмы действительно являются «описаниями природы» (например, уравнение Дирака) или «описаниями эволюции общества» (например, нелинейные уравнения)?

В нашей повседневной жизни мы широко используем математические вычисления. Мы считаем вещи и предметы, подсчитываем прибыль или убытки, вероятность получения дохода при покупке или продажи акций, сравниваем рыночные цены товаров, составляем бухгалтерские балансы, сметы расходов и т.д. Пересчитывая какое-то множество предметов, мы, однако, далеки от мысли, что число является внутренне присущим им признаком. Таким образом, в нашем обычном познании мы ограничиваемся лишь инструментальной функцией математических формализмов и не делаем далеко идущих выводов о внутренней «присущности» реальным физическим объек-

там (вещам) свойств идеальных концептуальных объектов и формальных структур математики.

Однако в развитом научном познании исследователи-теоретики имеют дело не с непосредственно, перцептивно воспринимаемыми физическими объектами, а с идеальными концептуальными системами (гипотезами, научными теориями, теоретическими моделями и т.д.), адекватность которых структурам физического мира может быть проверена только косвенным образом с помощью экспериментов. Поскольку и идеальные концептуальные системы научных знаний и математические (и логические) формализмы – это «однопорядковые» структуры нашего знаково-символического (логико-вербального) мышления, то в высокоабстрактных областях теоретического естествознания (например, в физике) математические теории (или их фрагменты) могут выступать не только как средство вычислений, но и как исключительно мощный когнитивный инструмент порождения новых научных понятий и идеальных концептуальных систем конкретных научных дисциплин (или даже нескольких областей одной дисциплины)<sup>2</sup>. Приписывая математическим понятиям физические интерпретации (смыслы), мы получаем возможность как бы «оседлать» формализм и путем его преобразования выявлять с его помощью ранее неизвестную концептуальную информацию. Формальная достоверность математических преобразований (выводов) лежит в основе нашей когнитивной уверенности в том, что полученные в результате таких преобразований математические понятия (формулы) также должны иметь какие-то физические смыслы. Благодаря способности генерировать новые научные понятия и концептуальные системы математические формализмы оказываются важнейшим

---

<sup>2</sup> Применение математических формализмов в качестве «порождающей грамматики» концептуальной системы конкретной дисциплины предполагает приписывание специально-научной интерпретации (смысла) математическим формализмам и их преобразований. Так, например, выражение  $dx/dt$  обладает сугубо математическим смыслом и может быть интерпретировано как полная производная некоторой функции  $x$ . В классической механике этому же выражению может быть приписано физическая интерпретация (смысл) – мгновенная скорость изменения положения в пространстве (обозначаемой  $x$ ) материальной точки. Аналогичным образом с помощью математических формализмов и их преобразований определяются и другие понятия классической механики, например, ускорение, импульс силы, количество движения тела, мощность и т.д. в механике формулируются с помощью Дифференциальные или операторные уравнения, на языке которых формулируются законы движения в механике, также порождают новые понятия. Так, например, стало возможным специфицировать понятие волны (как процесс распространения колебаний в среде) с помощью математической функцией, удовлетворяющей некоторому дифференциальному уравнению (волновому).

структурным элементом дедуктивной системы абстрактных научных теорий, они позволяют выводить следствия из их исходных и дополнительных гипотез, а также вычислять экспериментально проверяемые параметры (величины)<sup>3</sup>. Более того, в силу универсальности математических формализмов как идеальных структур знаково-символического мышления они могут порождать не только понятийные каркасы отдельных научных теорий, но и концептуальные системы целого класса теорий, выступая по отношению к этим теориям в качестве *базовых математических моделей*. Примером могут служить уравнение теплопроводности, волновое уравнение и уравнение Лапласа, составляющие основу математического аппарата классической физики<sup>4</sup>. И, наконец, функцию своего рода «синтаксиса», порождающей грамматики математические формализмы выполняют и по отношению к изолированным научным гипотезам и конкретным теоретическим моделям, а также в прикладных и технических дисциплинах, где широко применяются приближенные методы вычислений.

История физики дает множество примеров, свидетельствующих о том, что формированию каждой её новой области, возникновению каждой новой физической теории предшествовала разработка соответствующих разделов математики. Без изобретения математического анализа, интегрального и дифференциального исчисления было бы невозможно создание классической механики, без теории вероятности – молекулярно-кинетической теории и вообще классической статистической физики, без векторного анализа – классической электродинамики Максвелла, без тензорного анализа – теории относительности, без теории гильбертовых пространств, матричной алгебры, теории линейных дифференциальных уравнений – квантовой механики, без теории групп и обобщенных функций – теория элементарных частиц и т.д. Причем создание новых математических теорий, как правило, обуславливалось проблемами, совершенно независимыми от положения дел в физике.

В формальных аксиоматических теориях математики значения исходных терминов с самого начала не определяются, и они остаются

<sup>3</sup> По словам Л. де Бройля, «теория также должна иметь свои инструменты, для того чтобы получить возможность формулировать свои концепции в строгой форме и строго вывести из предположения, которые можно было бы точно сравнить с результатами эксперимента; но эти инструменты являются, главным образом, инструментами интеллектуального порядка, математическими инструментами, если можно так сказать, которые теория постепенно получила благодаря развитию арифметики, геометрии, анализа и которые не перестают множиться и совершенствоваться». (*Бройль Л. де. По тропам науки. М., 1962, С. 163.*)

<sup>4</sup> См. *Курдюмов С.П., Малинецкий Г.Г., Потапов А.Б. Синергетика – новые возможности. М., 1989. С. 4.*

неопределенными при выводе теорем из аксиом. Мы, поэтому, можем произвольно выбирать значения этих исходных терминов при одном только условии, что аксиомы останутся истинными. Но почему мы выбираем именно эти аксиомы, а не другие? Выбор диктуется прежде всего тем, приложима ли данная система к любым идеальным объектам, заданным извне в качестве интерпретации исходных терминов. Во многих случаях в качестве интерпретации какой-либо системы аксиом берется система концептуальных объектов из какой-нибудь другой аксиоматической теории. Тогда вопрос сводится к значению этой другой аксиоматической теории. Но когда мы утверждаем, что какое-то конкретное предложение данной формальной аксиоматической теории является теоремой, и что оно истинно, то это означает лишь, что это предложение вытекает из аксиом. Однако вопрос о том, что этому предложению соответствует в действительности и соответствует ли вообще, остается открытым, поскольку в формальной аксиоматике формальные выводы проводятся до какого бы то ни было приписывания значений исходным терминам. Неопределенность исходных терминов и нефиксированность операций составляют теоретическую основу *универсальности* математики и её языка, многообразия приложений математических формализмов как идеальных знаково-символических структур, обеспечивающих развертывание *потенциальной концептуальной информации* в понятийных системах самых различных дисциплин. Убедительным примером здесь может служить современная алгебра, в частности, абстрактная теория групп, система аксиом которой допускает существенно различные интерпретации – она обеспечивает извлечение новой когнитивной информации из самых разнообразных концептуальных объектов, будь то античастицы (ядерная физика) или брачные отношения (социология). Таким образом, в силу идеальной и сугубо «ментальной» природы математики, язык формальных математических структур знаково-символического мышления универсален. Благодаря этому открываются возможности для глубоких *аналогий* между различными областями математизированных дисциплин. Открытая, например, в свое время Гамильтоном оптико-механической аналогии, сыгравшей впоследствии важную роль в создании квантовой механики, была инициирована сугубо формальным подобием математического уравнения движения материальной точки в консервативном поле и уравнения лучевой оптики.

Разумеется, во многих областях естествознания и особенно в общественных и гуманитарных дисциплинах гипотезы формулируются не на языке математики, а вербально, в словесной форме. Соответственно, их дедуктивный аппарат включает главным образом логические формализмы (законы логики). Это, однако, не означает, что в этих дисциплинах в качестве полезных эвристических инструментов

вообще не привлекаются математические гипотезы, математические и компьютерные модели. Для их подключения здесь обычно вводятся упрощающие идеальные допущения, которые, правда, не всегда оказываются по результатам эмпирических проверок достаточно реалистичными. Для современных социогуманитарных дисциплин характерна все увеличивающаяся инструментальная значимость математических моделей. Речь, конечно, идет не только о прикладной социологии или экономических науках, широко использующих компьютерное моделирование (в том числе и динамических процессов). В антропологии, например, применение компьютерных моделей позволило установить, что в силу анатомических особенностей гортани все виды древних гоминид, включая неандертальцев, в принципе не могли обладать способностью к членораздельной речепродукции. Подсчеты вариаций митохондриальной и аутомсомной ДНК дали основание предполагать, что наш подвид *Homo sapiens sapiens* возник приблизительно 200 тыс. лет назад. Такого рода вычисления легли также в основу ряда исторических и этнографических гипотез о географическом расселении древних народов и эволюции современных человеческих популяций<sup>5</sup>. В истории и сравнительной социологии оказалось весьма продуктивным привлечение статистических и иных математических моделей, которые позволили обнаружить зависимости между появлением цивилизаций и плотностью населения, между резкими изменениями климата, экологическими катастрофами и социальной стабильностью, между возникновением сельскохозяйственного производства и ростом численности населения и т.д. Радиологический анализ позволяет довольно точно датировать обнаруженные археологами ископаемые останки людей, предметов древних культур, сооружений и т.д. Трудно даже представить, каким был облик современных общественных наук, если бы они полностью исключали привлечение прикладных математических гипотез.

Поскольку математические формализмы обеспечивают выведение новой информации, потенциально содержащейся в концептуальных объектах эмпирических наук, то в условиях архаического, преимущественно образного мышления возникает когнитивная уверенность в том, что эти формализмы суть структуры внешней реальности. Примером может служить «числовая парадигма» древних пифагорейцев, сформулированная в их известном тезисе: «Все есть чис-

<sup>5</sup> См.: Лимборская С.А., Хуснутдинова Э.К., Балановская Е.В. Этногеомика и геногеография народов Восточной Европы. М., 2002; Underhill P.A. et al. Y chromosome sequence variation and history of human population. // *Natural Genet.* 2000, № 26; Zhevetovsky L.A., Rosenberg N.A., Pritchard J.K. et al. A global picture of ancient expression of modern human revealed by a large set of autosomal STRs // *Amer. J. Human Genet.* 2002 (subm). Zhivetovsky L.A. et al. Genetic structure of human population // *Science.* 2002 (subm).



ло». Конечно, мы можем «превращать» наши идеальные интеллектуальные инструменты в некое «субстанциональное» знание о внешнем мире только в силу того, что действительно мыслим с их помощью и мыслим достаточно эффективно. Отождествляя цель и эффективное интеллектуальное средство её достижения («магия цели»), архаическое мышление распространяет эту когнитивную установку и на математические формализмы, ориентируя на поиск соответствующих коррелятов в структурах внешней реальности. Аналогичным образом дело обстояло и с логическими истинами, которым на протяжении многих веков, начиная с Аристотеля, приписывалась онтологическая интерпретация. Постепенная эволюция логико-вербального мышления (в кооперации с мышлением пространственно-образным) создала предпосылки для разрушения атавизмов древней «магии цели» и способствовала формированию всё более артикулированных эпистемологических представлений о наших интеллектуальных инструментах познания.

По-видимому, некоторые элементарные математические структуры врожденны в силу генетической детерминированности встроенных в нашу когнитивную систему относительно «низкоуровневых» аналитических стратегий знаково-символического мышления. В пользу этого также свидетельствуют результаты исследований интеллекта отдельных видов животных (в частности, птиц), обладающих зачатками неречевого знаково-символического мышления, – как оказалось, они способны считать предметы (естественно, в весьма ограниченных пределах, как правило, не более десяти). История прикладной математики берет своё начало с простейших арифметических вычислений и геометрических построений, отвечавших сакральным и сугубо прозаическим, практическим целям. Изобретение языка символов и формул позволило (по крайней мере уже древнеавилонским математикам) конструировать из простейших математических формализмов все более сложные и абстрактные, не забываясь о том, имеют ли они какое-либо прикладное значение. Развитие устной культуры и искусства аргументации в Древней Греции способствовало формированию теоретической математики, где впервые стали применяться неформальные доказательства (с помощью геометрических построений, силлогистического типа и др.). Критерии неформальных доказательств эволюционировали на протяжении всей последующей истории математики, оставаясь при этом достаточно неопределёнными. Даже интуиционистской математике, вопреки первоначальным планам Брауэра, так и не удалось разработать критерии абсолютного понятия конструктивного доказательства, а лишь его более слабые и более сильные версии. Только по отношению к формальным доказательствам, отвечающим жесткому набору требований, можно говорить о стандартах строгости, не зависящих от времени. Но ведь многие математические теории, в том числе такие «про-

стве», как, например, арифметика натуральных чисел, формально не аксиоматизируемы. По-видимому, любой перечень методов математического доказательства и принципов построения математических объектов всегда остаётся и должен оставаться неполным. К тому же эти методы и принципы подлежат пересмотру и уточнению, поскольку возможность дальнейшего расширения и углубления конструктивных способностей человеческого мышления ничем не ограничена (естественно, в пределах эволюции человека как биологического вида).

Таким образом, элементы гипотетичности, предположительности обязательно присущи математическим формализмам как идеальным структурам знаково-символического мышления. Эпистемологический статус утверждений формальных наук ничем принципиально не отличается от статуса гипотез наук эмпирических, хотя они и имеют прямое отношение к аналитическим стратегиям переработки информации, которыми мы (в отличие от природных и социальных процессов) можем непосредственно сознательно управлять (хотя бы частично) и которые мы в состоянии конструктивно оптимизировать. Конечно, появление современной вычислительной техники, её бурное развитие за последние десятилетия постепенно размывает функциональные границы между искусственным и естественным интеллектом. В этой связи возникает много вопросов эпистемологического характера. Можно ли утверждать, например, что эпистемологический статус (а точнее, степень достоверности) математических утверждений, доказательство которых проведено с использованием современной вычислительной техники, не уступает утверждениям, полученным традиционным способом?

В качестве формальных систем математические теории не обязательно должны быть наглядными или «интуитивно истинными» – наглядность и интуитивная достоверность не являются их критериями истинности. Достаточно лишь, чтобы эти теории были бы формально правильными, свободными от внутренних противоречий. В рамках формальной аксиоматики система аксиом может быть также исследована на предмет наличия таких свойств, как независимость какой-либо аксиомы от других, полнота, категоричность, и т.д. Традиционным способом проверки истинности, формальной правильности математических доказательств является их перепроверка другими математиками, сообщество которых выступает в роли окончательных «верховных» судей. Однако математики тоже люди, и они могут ошибаться – такие случаи широко известны из истории научного познания. Формализация доказательств автоматизировала вычисления и тем самым создала предпосылки для конструирования современной вычислительной техники. Создание такой техники, в свою очередь, открыла новые возможности для проверки правильности математических доказательств и их поиска компьютером. Первые попытки ис-

пользовать ЭВМ для проверки и получения логико-математических доказательств были предприняты еще в 50-х гг. XX в. (программа «Логический Теоретик»). К настоящему времени эта область применения ЭВМ значительно расширилась, причем перечень проблем, решаемых только компьютером, постоянно пополняется.

Одной из таких проблем, представляющей особый интерес для эпистемологии, является доказательство математических гипотез и решение задач, относящихся к категории необозримых. В таких случаях традиционные методы вычислений и проверки полностью исключаются – никакой исследователь не в состоянии ни провести требуемые вычисления «вручную», ни шаг за шагом повторить и перепроверить весь процесс доказательств или решений. Поиск необозримых доказательств и их верификация может быть осуществлен исключительно «искусственным интеллектом», компьютером. Поскольку доказательство в формальных дедуктивных системах является *эффективным*, т.е. существует некоторая механическая процедура, задающая последовательность выполнения тех или иных действий (алгоритм), которая позволяет проверить, будет ли полученная последовательность формул правильно построенным выводом или нет, то задача проверки правильности формального доказательства, представленного в виде текста, оказывается *алгоритмически разрешимой* и может быть реализована на компьютере. Перепроверку формальных доказательств в принципе может выполнить любой компьютер, удовлетворяющий соответствующим системным требованиям и обладающий нужными вычислительными характеристиками. Таким образом, в случае формального вывода полученные результаты оказываются *независимыми от конкретного типа компьютера*. Более того, обнаружилось, что эти результаты также независимы и от программы, обеспечивающей получение формального вывода, и даже от языка программирования, который был использован для написания исходной программы. Сложнее дело обстоит с проверкой правильности программ.

Компьютер представляет собой *физическое устройство*, состоящее из электронных микросхем, материальных накопителей информации, оборудования для её ввода и вывода и т.д., которое обеспечивает его функционирование как *логического устройства*, позволяющего представить доказательства в виде формальных выводов в некоторой дедуктивной системе и зафиксировать эти выводы на материальных носителях информации. Работоспособность аппаратных средств, отсутствие у оборудования ошибок и сбоев в работе можно проверить с помощью соответствующих тестирующих программ, перепроверки результатов тестовых вычислений на других компьютерах и т.д. Если работоспособность компьютера как физического и логического устройств сомнений не вызывает, то для проверки пра-

вильности необозримых доказательств и решений остаётся лишь убедиться в правильности соответствующих программ. На практике правильность программы (алгоритма) выявляется на этапе её прогона и проверки с помощью специальных тестов, подбор которых во многих случаях представляет собой далеко не тривиальную задачу. Такое тестирование, по-видимому, носит сугубо теоретический характер, так как в ходе него соотносятся два теоретических продукта – программа и полученные с её помощью результаты вычислений. Однако положительные тесты все же не могут гарантировать отсутствие логических ошибок в программе и служить доказательством её правильности. Проблема сдвигается в плоскость традиционных методов, поскольку удостовериться в формальном выполнении алгоритма и тем самым доказать формальную правильность программы может только человек. Рассматривая текст программы как статический математический объект, на который распространяются аксиомы и логико-математические правила вывода, разработчик должен «вручную» воспроизвести действия, строго соответствующие имеющемуся алгоритму. Если для создания программы был использован язык с обозримым алфавитом, то доказательство правильности программы оказывается *обозримой процедурой* и осуществить его не представляет особого труда. Обозримая программа в принципе способна совершить необозримое число шагов, проверить и вывести сколь угодно длинные формулы. Если же текст программы *необозримый*, то доказательство правильности такой программы оказывается задачей *алгоритмически неразрешимой* (так как вопрос о завершении, остановке произвольного вычисления остается открытым), которая не может быть реализована на компьютере. Понятно, что только в случае обозримости используемых программ, а также при наличии формального вывода в некоторой формальной дедуктивной системе степень строгости и эпистемологический статус математических утверждений, доказательство которых проведено с использованием современной вычислительной техники, несколько не уступает результатам, полученным традиционным способом<sup>6</sup>.

### **Эпистемологические особенности логики как формальной науки**

Несмотря на то что логика возникла еще в IV в. до Р.Х., и её методы на протяжении столь длительного исторического периода менялись, основная задача этой науки в общем и целом существенно не претерпела изменений – она всегда исследовала и продолжает исследовать то, как из одних утверждений логически выводятся другие утверждения. Логика исходит из предположения, что логический вывод

---

<sup>6</sup> Более подробно см. : *Знатнов С.Ю.* Эпистемологический статус компьютерных теорем. Кандидатская диссертация. М., 2000.

зависит только от «формы», т.е. от способа связи входящих в него утверждений и их структуры (строения), а не от конкретного содержания этих утверждений. Современная неклассическая логика не внесла в этот подход каких-либо принципиальных изменений – все её разделы и направления также игнорируют конкретное концептуальное содержание высказываний (умозаключений) и оперируют только с их логической формой, структурой. Логика, как и математика, не является опирающейся на эксперимент эмпирической наукой (наподобие физики) и в силу этого не может быть отображением каких-то «наиболее общих отношений реального мира» или «реального положения дел в физическом мире». Она также имеет непосредственное отношение лишь к нашему знаково-символическому мышлению, а точнее к его разновидности – *мышлению логико-вербальному*. Логика – это наука об идеальных, формальных логических структурах вербального мышления, принудительная сила которых вытекает из их тавтологичности. Формулы, представляющие любой логический закон, всегда истинны, независимо от каких-либо интерпретаций переменных. Все законы логики являются логическими тавтологиями, которые нельзя подтвердить или опровергнуть никаким опытом. Конечно, логические тавтологии «пусты» в том смысле, что они не содержат когнитивной информации о внешнем мире. Но эти формализмы содержат информацию об идеальном, «правильном» вербально репрезентируемом мышлении, о правилах этого мышления, которые обеспечивают трансляцию истинностных значений от посылок к заключениям. Благодаря этому они позволяют выявить скрытую, потенциально содержащуюся в концептуальных объектах, когнитивную информацию.

Мы начинаем пользоваться логикой как инструментом нашего знаково-символического мышления практически с того момента, как начинаем говорить. Нетрудно, однако, заметить, что многие люди способны мыслить и рассуждать логически правильно (по крайней мере в простых случаях) сугубо интуитивно, не обладая какими-либо детальными знаниями о законах логики. И это неудивительно, так как в человеческих популяциях с относительным доминированием знаково-символического мышления имеется генетически врожденная предрасположенность к использованию преимущественно аналитических стратегий переработки когнитивной информации. Такие логические правила, как *modus ponens* и *modus tollens*, по-видимому, являются алгоритмами мыслительных программ и метапрограмм, которые оказались включенными в арсенал некоторых «встроенных» в нашу когнитивную систему аналитических стратегий. Разумеется, эти мыслительные программы (также как и достаточно развитый естественный язык) не возникают в окончательном и готовом виде одновременно с появлением подвидов *Homo sapiens* и в этом смысле не

являются нашим древнейшим филогенетическим наследием, как, например, перцептивное восприятие или пространственно-образное мышление. Они формируются вместе с соответствующими когнитивными структурами в ходе биологической, когнитивной и культурной эволюции человеческих популяций, эволюции естественного языка, символьного (вербального) сознания, памяти и т.л. в процессе постепенной смены доминирующего пространственно-образного мышления мышлением преимущественно знаково-символическим (логико-вербальным). С этой, эволюционно-когнитивной точки зрения, нет и не может быть никакой «общечеловеческой логики», также как и нет «человека вообще», независимо от той или иной стадии биологической, когнитивной и культурной эволюции отдельных человеческих популяций.

Эволюция логико-вербального мышления в особенности тесно сопряжена с эволюцией символьного (вербального) сознания, которое позволило людям управлять «встроенными» аналитическими стратегиями и соответствующими мыслительными схемами, оперировать ментальными репрезентациями (в словесном формате) независимо от содержащейся в них конкретной концептуальной информации. Благодаря этому у человечества открылись принципиально новые когнитивные возможности конструировать идеальные логические схемы рассуждений (формализмы) и использовать их для извлечения из концептуальных объектов новой информации. Характерно, что само возникновение логики как науки неразрывно связано с определенным этапом когнитивной эволюции не всех без исключения, а только некоторых человеческих популяций. Оно стало реально возможным лишь благодаря широкому распространению вербальной формы культурно-информационного контроля окружающей среды и архаической магии слова. Если с помощью «истинного» слова нельзя «овладеть» окружающим миром, то незачем тогда исследовать, при каких условиях одни высказывания «безошибочно» выводятся из других высказываний!

Итак, логика изучает идеальные, формальные структуры логико-вербального мышления (как разновидности мышления знаково-символического), которое (в кооперации с мышлением пространственно-образным) подлежит когнитивной эволюции. Создание идеальных мыслительных схем, а затем использование для обозначения соответствующих структурных «формальных» единиц общепринятого в математике языка символов и формул, а также математических методов и т.д. — все это позволило конструировать абстрактные логические исчисления. Но отсюда ясно, что по отношению к нашему целостному мышлению, функционирование которого основано на тесной кооперации и взаимодействии правого и левого полушарий, все таутологии любых логических исчислений с эпистемологической точки

зрения могут рассматриваться только как идеальные схемы, гипотезы. Именно поэтому логические выводы, осуществляемые в соответствии с формальными схемами, далеко не всегда согласуются с нашим интуитивным пониманием и влекут за собой появление парадоксов. Непонимание того обстоятельства, что логика – это не эмпирическая «наука о мышлении», а наука об идеальных, формальных структурах нашего логико-вербального мышления, которая в силу идеального, а следовательно, и гипотетического характера своих допущений не всегда обязана следовать за нашей интуицией, за нашим интуитивным пониманием «правильного мышления» или «правильного умозаключения», нередко влечет за собой незаслуженную критику логических формализмов. В недалеком прошлом оно приводило к сугубо схоластическим попыткам разработать какие-то «логики» (типа «диалектической логики»<sup>7</sup>), «законами» которых были бы отрицания логических законов (например, закона противоречия или закона исключенного третьего).

Конечно, наше мышление в целом не охватывается логикой, так как мы мыслим не только в соответствии с логическими правилами и формальными схемами, но и с помощью множества идеальных схем нелогического характера – причинно-следственные отношения, математические правила, эмпирические обобщения, извлеченные из обыденных знаний, гипотезы и допущения эмпирических наук и т.д. Кроме того, мы активно привлекаем ресурсы нашего пространственно-образного мышления, которое использует неаналитические стратегии переработки когнитивной информации, – оно обеспечивает целостное понимание (а это – важнейший аспект мышления), дарит нам творческое озарение, позволяет открыть новые знания, и т.п. Наш естественно развивающийся язык хорошо адаптирован к взаимодействию систем правого и левого полушарий, и от него было бы нелепо ожидать свойств, приемлемых только для искусственных, формализованных языков. Всё это, однако, не умаляет значения логики. Символьная формулировка логических формализмов помогает нам явным образом и с гораздо большей точностью использовать их в качестве составной части нашего интеллектуального, мыслительного арсенала. Изучение логики как теории формального мышления помогает укрепить и расширить наши врожденные аналитические способности. Хотя наши врожденные способности к аналитическому мышлению, позволяющие «интуитивно» обнаруживать верные аргументы, скорее всего не возрастают в результате изучения логики, все же нельзя отрицать, что в ходе такого изучения увеличивается воз-

---

<sup>7</sup> См., например: *Бирюков Б.В.* Борьба вокруг логики в Московском государственном университете в первое послесталинское десятилетие (1954-1966). В: *Логика и В.Е.К. М.*, 2003.

возможность проверить правильность рассуждений. Люди могут сознательно (или неосознанно, по неведению) нарушать законы логики (как и правила грамматики), которые имеют силу независимо от нас. Но если в случае грамматических ошибок мы рискуем оказаться непонятыми собеседником (что, конечно, приводит к недоразумениям), то незнание законов логики может повлечь за собой ошибочные выводы и поступки, т.е. неадекватное, неадаптированное поведение.

Итак, математические и логические формализмы имеют непосредственное, прямое отношение только к нашему левополушарному знаково-символическому (логико-вербальному) мышлению. В отличие от мышления пространственно-образного это мышление людей по своей когнитивной природе является филогенетически «вторичным», оно использует «вторичное», символическое кодирование мысли и порождает идеальные понятия и концептуальные системы, оперируя своими репрезентациями с помощью аналитических стратегий. Как идеальные и формальные структуры знаково-символического мышления математические и логические теории применимы к «внемыслительной» реальности только косвенным образом, опосредовано — через применение к нашим эмпирическим знаниям, теоретическим допущениям, гипотезам и теориям. Математические (и логические) формализмы позволяют выявить, развернуть огромный массив скрытой, потенциально содержащейся в теоретических объектах эмпирических наук концептуальной информации, они дают возможность вычислить эмпирически проверяемые параметры и величины, получить новые следствия из научных теории и гипотез и т.д., а в абстрактных, математизированных областях естествознания выступают и как инструмент порождения новых научных понятий и концептуальных систем.

Отталкиваясь от результатов исследований когнитивных типов мышления, эволюционно-информационная эпистемология дает ясное понимание несостоятельности как наивно реалистических, так и инструменталистских взглядов на когнитивную природу математических и логических формализмов. Наивный реализм постулирует наличие истоков математических и логических формализмов в структурах и корреляциях окружающего нас внешнего мира. Он апеллирует к перцептивно воспринимаемым признакам реальных объектов, относительно которых мы должны заранее обладать знанием, позволяющим нам решить, от каких из них нам следует абстрагироваться и какими идеальными свойствами и отношениями эти объекты необходимо «наделить». Понятно, что без точного выбора и спецификации требуемых идеальных свойств и отношений мы не знаем и в принципе *не можем знать* границ абстрагирования от *бесконечной совокупности всех признаков реальных объектов*. Таким образом, процедуры абстрагирования и идеализации должны обязательно базиро-



ваться на каких-то (в том числе и неявно принимаемых) *гипотезах* – математических или специально-научных. В математизированном научном познании выбор идеальных признаков определяется подключаемыми к концептуальной структуре теорий математическими формализмами, математическими понятиями и моделями, которые порождают соответствующие идеальные объекты и концептуальные системы. Именно в результате такого подключения в структурах физических теорий генерируются такие понятия, как, например, материальная точка, идеальный газ, абсолютно черное тело и т.д. Об их правомерности и адекватности можно судить лишь на основании результатов эмпирических (экспериментальных) проверок научных теорий.

В отличие от наивного реализма инструментализм базируется на предпосылке, что математические и логические формализмы представляют собой лишь сугубо интеллектуальные орудия для вычислений и выведения следствий, и что им ничего «не отвечает» в окружающей нас реальной действительности. Конечно, математические и логические формализмы непосредственно не имеют прямого отношения к внешнему миру, а только к работе нашей когнитивной системы, к нашему левополушарному, знаково-символическому мышлению. Но необходимо учитывать, что и наша когнитивная система в целом и наше знаково-символическое мышление (также как и работающее в тесной кооперации с ним наше правополушарное пространственно-образное мышление) относятся к тому же типу объективной реальности, как и окружающая нас действительность. Когнитивные способности людей (включая мышление, сознание, память, а также их интегральная характеристика – интеллект) имеют информационную природу, они являются продуктом продолжающейся биологической (когнитивной) и культурной эволюции человеческих популяций.

---

# ФОРМАЛЬНАЯ ОНТОЛОГИЯ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

*В.Л. Васюков*

## Э. Гуссерль (возникновение термина)

Мы обязаны идеей (и термином) формальной онтологии Эдмунду Гуссерлю, который в своих «Логических исследованиях» (1900/01) различает *формальную логику*, с одной стороны, и *формальную онтологию*, с другой. Формальная логика имеет дело с взаимосвязями истин (или пропозициональных значений в общем случае) – с отношением выводимости, с непротиворечивостью и общезначимостью. Формальная онтология имеет дело с взаимосвязями вещей, с объектами и свойствами, частями и целым, отношениями и совокупностями. Как формальная логика имеет дело с отношениями выводимости, которые формальны в том смысле, что они применимы к выводам в силу лишь одной своей формы, так и формальная онтология имеет дело со структурами и отношениями, которые формальны в том смысле, что они экземплифицированы, в принципе, всей материей, или, говоря другими словами, объектами всех материальных сфер или областей реальности.

Гуссерлевская формальная онтология основывается на мереологии, на теории зависимости и топологии. Его Третьи Логические Исследования озаглавлены «О теории целого и частей» и состоят из двух глав: «Разница между независимыми и зависимыми объектами» и «Мысли в отношении теории чистых форм целого и частей». В отличие от более известных «экстенциональных» теорий частей и целого, таких, например, как системы Лесьневского, Леонарда и Гудмена, гуссерлевская теория не затрагивает того, что мы могли бы назвать вертикальными отношениями между частями и целым, когда первые охватывают, включают в себя вторые. Скорее гуссерлевская теория касается горизонтальных отношений между сосуществующими частями, отношения, которые придают единство или интегральность рассматриваемым целостностям. Проще говоря, некоторые части целого существуют просто рядом друг с другом, они могут быть разрушены или удалены из целого, не нанося ему никакого ущерба. Целое, все части которого связаны исключительно подобными отношениями рядоположенности, называется массой или агре-

гатом, или, выражаясь более технически, чисто совокупным целым. Во многих целостностях, можно даже сказать во всех целостностях, демонстрирующих любую разновидность единства, некоторые части находятся друг с другом в формальных отношениях, которые Гуссерль называет *необходимой зависимостью* (иногда, но не всегда, он говорит о *необходимой взаимозависимости*). Подобные части, например, отдельные случаи оттенков цвета, насыщенности и яркости, характерные для данного цвета, не могут с необходимостью существовать в целостности данного типа, без ассоциированности с их дополнительными частями. Имеется огромное разнообразие подобных побочных отношений зависимости, порождающих огромное многообразие типов целостности, которые неразличимы при стандартном подходе экстенциональной мереологии.

Следует подчеркнуть, что Гуссерль в своих работах не использует **какой-либо формальный аппарат в современном понимании смысла этого слова**. Его исследования представляют собой скорее анализ с целью выяснения интуитивных оснований и понятий для разработки систем формальной онтологии. И не удивительно, что первый раздел работы П. Саймонса «Три эссе по формальной онтологии» носит название «Формализация гуссерлевской теории части и целого» [Simons 1982].

Гуссерлевская формальная онтология становится более понятной, если обратиться к идеям его учителя Ф. Brentano.

## **2. Ф. Brentano (дескриптивная онтология)**

Brentano пропагандировал идею того, что он называл «дескриптивной психологией», дисциплины, которая должна, с одной стороны, дать точное знание структур и категорий ментальной жизни, а с другой стороны, обеспечить эпистемологически достоверные основания других разделов философии.

Эта Brentановская дескриптивная психология является, в сущности, **дескриптивной онтологией сознания** — онтологией ментальных структур. Это влечет тезис о том, что подходящая для каждого рассматриваемого случая форма описания подразумевает нечто вроде таксономии различных видов конституэнт в данной области и различных форм отношений между этими конституэнтами. Возникает нечто вроде установки дескриптивного, или таксономического реализма. Следующий шаг, сделанный учениками и последователями Brentано, в частности Штумпфом, Марти, Твардовским, Мейнонгом, Эренфельсом и Гуссерлем, заключается в перенесении подобной установки в иные области философии. Тем самым мы оказываемся за

пределами чистой ментальности и имеем дело уже с некоторым видом дескриптивной общей онтологии.

Внутреннее восприятие — согласно Brentano — является источником нашего знания природы бытия, так же как оно является источником знания природы истины и природы добра и зла. И что-то, что может быть сказано о бытии вещей, которые не воспринимаются во внутреннем восприятии, можно понимать только по аналогии с тем, что мы способны сказать о нас самих как мыслящих субъектах. Отсюда мы получаем также соответствующее трехчленное деление дескриптивной онтологии. Это (1) онтология вещей, (2) онтология состояний дел и (3) онтология оценок.

*Онтология вещей* возникает, если мы переходим от психологии представления к исследованию непсихологических коррелятов актов представления. Поэтому вещь теперь рассматривается как возможный коррелят представления, включающий простые и сложные смысловые данные. Поздний Brentano в своей онтологии вещей исследует коллективы и пространственные и временные континуумы. Тогда же он приходит и к реизму — доктрине, утверждающей, что существуют только лишь вещи, позднее поддержанному и развитому сначала Tвардовским, а затем Т. Котарбинским.

*Онтология состояний дел* возникает сходным образом в случае перехода от психологии суждения к исследованию онтологических коррелятов актов суждения. Последние, согласно Brentановской экзистенциальной теории суждения, будут иметь следующую первичную форму: существование А и не существование А. (а) «А существует» означает, что А является таким, что каждый, кто судит о нем, очевидным образом должен принять его; (б) «А не существует» означает, что А является таким, что каждый, судящий о нем, очевидным образом должен отвергнуть его;

Наконец, *онтология оценок* возникает, когда мы переходим от психологии чувств, воли и предпочтения к исследованию онтологических коррелятов соответствующих актов. Еще в своей диссертации, относящейся к 1862 г., Brentano анализирует Аристотелевское различие «быть в смысле категорий» и «быть в смысле быть истинным» таким образом, что это можно рассматривать как принадлежащее сфере общей онтологии. Однако без интерпретации взглядов других философов Brentano крайне неохотно формулирует свои собственные онтологические тезисы. Как следствие, по-видимому, гораздо больший вклад в онтологию оценок сделали его ученики, которые использовали психологический анализ Brentano в качестве основы для своих исследований.

### **3. К.Твардовский (развивая Brentano)**

Что касается онтологии вещей, то здесь среди других учеников Brentano большой вклад был сделан К. Твардовским. Твардовский был, вероятно, первым из Brentанистов, кто рассматривал общую теорию предметов, все еще называя ее метафизикой. Понятие предмета у Твардовского дается совершенно Brentановским образом:

«Предмет можно описать приблизительно так: все, что представлено посредством представления, признано либо отброшено суждением, желаемо либо отвергаемо эмоциональной деятельностью, мы называем предметом. Предметы бывают реальными и нереальными, возможными и невозможными. Общим для них является то, что они могут быть или же являются объектом (не интенциональным) психического акта, что их языковым обозначением является название (в вышеизложенном смысле) и что они, рассматриваемые как род ... находят свое языковое выражение в названии «нечто». Все, что в самом широком смысле есть «нечто», называется, сперва в отношении к представляющему субъекту, а потом также и независимо от него – предметом» [Твардовский 1997, с.78]

### **4. Н. Коккьярелла (не-гуссерлевская концепция формальной онтологии)**

Однако существует и не-гуссерлевская традиция в формальной онтологии, не основывающаяся на мереологии. Следуя аналитической онтологии, формальная онтология определяется как теория бытия с точки зрения формальной логики, т.е. теория бытия в рамках и на языке элементарных формальных теорий. Главным сторонником данной позиции может считаться Нино Коккьярелла. Принимая, в частности, что каждая наука рассматривает свой специфический «способ существования», Коккьярелла утверждает, что формальная онтология изучает различные формализации, относящиеся к систематической классификации всех «способов» или категорий существования в самом общем виде. Обычно каждый «способ существования» подходит для некоторой специфической формальной онтологии и представляет некоторый тип переменных, для которых синтаксическое правило их применения отражает онтологическое правило для данного «способа» существования [Cocchiarella 1974, pp. 29-30]. Тем самым формальная онтология изучает логические характеристики предикации, квантификации по переменным и различные теории универсалий.

Несмотря на отличие от гуссерлевской формальной онтологии, обе разновидности формальной онтологии по большинству вопросов занимают сходные позиции. Само по себе это следует из того факта, что и теория множеств и мереология, используемые для caracteriza-

ции онтологических понятий, представляют собой конкурирующие системы в основаниях математики. Спектр вопросов и понятий, на которых они основываются, в значительной степени обусловлен интересами математики.

Во всяком случае, сегодня принято считать, что вторая из них больше занята систематикой категорий и страт, образующих формальную онтологию, в то время как первая систематически анализирует темы противопоставления вариантов формальной онтологии.

## 5. Лесьневский

Особое место среди различных систем формальной онтологии занимает система Онтологии Лесьневского, удовлетворяющая определению формальной онтологии в смысле Коккьяреллы, в то время как другая егос система – Мереология, надстраиваемая над Онтологией, в точности подпадает под определение Гуссерля. И, фактически, это разделение принимается как само собой разумеющееся во всех исследованиях, касающихся онтологических проблем.

Тонкость, которую следует учитывать в подобного рода рассуждениях, заключается в том, что Онтология Лесьневского представляет собой как логическую систему, так и систему формальной онтологии. Уникальность ситуации здесь связана с онтологическим смыслом связки «есть», подразумеваемым Лесьневским. По сути дела, теоремы Онтологии Лесьневского — это некоторые онтологические положения, но в то же время это и чисто логические утверждения.

Стандартная семантика обычной логики предикатов задается с помощью понятия модели, представляющей собой некоторое множество с заданной на нем системой отношений и функций. Все что мы можем сказать непосредственно об элементах этого множества, исчерпывается отношением равенства, т.е. мы можем констатировать лишь совпадение некоторых элементов, и ничего больше. Все остальное определяется «внешним» образом, с помощью отношений и функций.

В семантике Онтологии Лесьневского, предложенной З. Стахняком [Stachniak 1981], множество-носитель модели представляет собой булеву алгебру, отношение порядка которой интерпретирует связку «есть». Все отношения и функции должны быть согласованы с этой базисной булевой структурой. Отсюда предметная область семантики обладает как «внешней», так и некоторой «внутренней» структурой, не проясняемой с помощью отношений и функций. Именно ее и можно рассматривать как некоторую формальную онтологию, с одной стороны, предшествующую всякому последующему прояснению свойств объектов, определяемых отношениями, а с другой стороны,

эта структура имеет чисто логическую природу, поскольку детерминирована совместной интерпретацией логических связей и связи «есть».

Таким образом, Онтологию Лесьневского можно рассматривать как некий вид систем формальной онтологии, в которой характеристика онтологических понятий предполагает не теоретико-множественные и не мереологические термины, но лишь термины исчисления имен (в терминологии Е. Слупецкого). Преимущество такой точки зрения сказывается при расширении диапазона понятий онтологических объектов и отношений между ними.

«Гибридный» характер системы Онтологии Лесьневского, являющейся одновременно и логической системой и системой формальной онтологии, заставляет уделить больше внимания сфере взаимоотношения логики и онтологии, удельному весу логических методов в построении систем формальной онтологии.

### **5. Онтологика и формальная онтология**

В своей вводной статье, открывающей публикацию эссе по формальной онтологии в журнале «Логика и логическая философия», Е.Пежановский определяет онтологию в ее наиболее общей и традиционной форме как теорию того, что есть, теорию бытия. «Она рассматривает полный онтологический универсум, включая все предметы, являющиеся возможными. Два основных вопроса направляют логическое исследование: что возможно и почему? Или в более общем и глубоком виде: каким образом возможно возможное?» И далее: «Ввиду природы этих вопросов онтология является наиболее общей дискурсивной дисциплиной. Фактически, она представляет собой общую теорию возможности. С другой точки зрения, она может рассматриваться как общая теория отношений, общая теория вещей и свойств или теория ситуаций, событий и процессов» [Perzanowski 1994, p. 4]

Так определяемая онтология состоит из трех частей: онтики, онтометодологии и онтологики. «Онтика посвящена выбору онтологических проблем и понятий, их дифференциации, классификации и анализу; конструированию концептуальной сети данной онтологической теории и формулировке разумных онтологических гипотез. Онтометодология занимается способами разработки онтологии и их принципами, наряду с методами и типами онтологических конструкций». Согласно Пежановскому, онтологика – это «логика царства онтики. Она касается организации онтологического универсума и пытается описать его механизмы. Онтологика является дисциплиной, ис-

следующей онтологической связи, в частности, логические отношения между онтическими положениями».

Рецепт подобной онтологии выглядит следующим образом: берем достаточно интересную (и реальную) онтологическую проблему и пытаемся решить ее теоретически, используя некоторую теорию. С этой целью проводим концептуальный анализ (принадлежащий сфере онтики), определяем подходящие примитивные понятия и проясняем их до такой степени, чтобы быть в состоянии обнаружить разумные аксиомы, затем используем логическую дедукцию и проводим соответствующие семантические исследования. Если метод срывает, то мы получаем некоторые теоремы, способные пролить свет на поставленные проблемы.

Итак, резюмирует Пежановский, «онтология — это просто онтология, полученная путем ответа на онтологические вопросы с использованием логических методов и процедур. Вкратце,... онтология = онтология / логика».

При подобном определении онтологии возникает возможность не только логического исследования связей между онтическими положениями, но в более широком плане — формального исследования. В сущности уже даже первопорядковое исчисление с равенством получается путем добавления нелогического символа равенства и соответствующих аксиом, описывающих его поведение, к логическому исчислению<sup>1</sup>. То же самое относится и к теории множеств, к теории порядка и т.д. По-видимому, следует различать чистую онтологику и формальную онтологию, определяя последнюю как (перефразируя Пежановского) онтологию по модулю формальных систем.

## 6. Онтологические Типологии

Диапазон формальной онтологии гораздо более широк, чем у онтологии. Достаточно заметить, что согласно приведенным выше определениям онтология является составной частью формальной онтологии, подобно тому как логической частью любой теории неизменно является классическая логика предикатов (если только не рас-

<sup>1</sup> Существует некоторое разногласие по вопросу о том, является ли символ равенства нелогическим знаком. Согласно [Расёва Сикорский 1972, с. 222], «... формализованные теории первого порядка содержат в своем языке некоторый бинарный предикат, который соответствует отношению равенства и называется *знаком равенства*». С другой стороны, в [Шенфилд 1975, с. 42] находим следующую формулировку: «Функциональный символ или предикатный символ, отличный от =, называется *нелогическим символом*; остальные символы называются *логическими*»



смагивать неклассические формальные системы, для которых это ограничение не имеет места). С другой стороны, многие онтологические типологии (классификации по онтическим положениям) явно подразумевают нелогические аспекты анализа. Возьмем, например, типы онтологии, приводимые Р. Поли в его книге «Формальная онтология» [Polí 1992, p. 42-43]. Он определяет их следующим образом.

*Онтология объектов и свойств.* Основывается на номинальной предикации. Восходит от Аристотеля к Мейнонгу как наиболее радикальному ее представителю. Предполагает соответствие между лингвистической и онтической формами (наложение языковой решетки на мир). Лингвистические варианты в перспективе моделирования представлены в работах Карнапа и Робинсона.

*Стратифицированная онтология.* Описывает мир в категориях зависимости. Первичная внутренняя дифференциация определяется по вопросу принятия или опровержения различия между общей и локальной онтологиями. Случай опровержения различия очень близок к математике. Если зависимость между слоями (стратами) описывается с помощью отношений, то получаем семантику Тарского, если с помощью функций — то Фреге и Чёрча. В случае принятия различия между общей и локальной онтологиями мы имеем дело с правилами зависимости и независимости между различными областями онтологии. Здесь среди наиболее разработанных версий привлекают внимание теория систем, теория катастроф и термодинамика. С философской точки зрения подобной онтологией является феноменологическая онтология, представленная в трудах Гуссерля, Шелера, Инггардена и Гартмана.

*Онтология событий.* Основывается на вербальной и невербальной предикации (Иван говорит, Иван выходит). Ее предметами являются динамические сущности. Можно соотнести ее с современной физикой, в частности, с теорией относительности и квантовой механикой. Среди философов наиболее видные представители такого рода онтологии — Уайтхед и Гегель.

*Комбинаторная онтология.* Описывает универсум как композицию элементов и комбинацию элементов. Если комбинация является следствием внешних факторов и все элементы комбинируемы, то получаем комбинаторную логику Чёрча и Карри. Если же комбинации детерминируются внутренними факторами, то получаем онтологию, близкую к стратифицированной онтологии. С точки зрения мышления она является плюралистической онтологией. С лингвистической — категориальной онтологией в смысле семантических категорий (иначе интерпретируется как функциональная онтология). Онтология

комбинаторного типа была чужда классическому греческому и средневековому европейскому мышлению, но близка арабо-индийской традиции диалектической теологии Калама и мутаазилитов. В Европе классический представитель комбинаторной онтологии — Лейбниц.

*Трансформационная онтология.* Монистична. Существуют внутренние и внешние трансформации. Плюралистичность достигается только как совокупность состояний или модификаций. Это онтология Спинозы. Подобна геометрической интерпретации релятивистской физики. Не имеет специальных лингвистических коррелятов.

*Онтология ментальных конструкций.* Описывает ментальную активность непосредственного созерцания, не рассматривает «факты» внешнего мира. Вероятно, наиболее значительным ее представителем является математический интуиционизм Брауэра. Наиболее радикальные ее формы не допускают достоверных лингвистических выражений.

Это несомненно пример далеко не исчерпывающей типологии, поскольку сам Р. Поли далее говорит о том, что возможны и другие типологии на основе иных принципов. Например, мы получаем следующий список оппозиций:

*Статическая онтология — динамическая онтология.*

*Однородная онтология — стратифицированная онтология.*

*Модальная онтология — немодальная онтология.*

*Дескриптивная онтология — конструктивная онтология.*

*Лингвистическая онтология — экстралингвистическая онтология.*

Ясно, что учет всех особенностей и принципов, положенных в основу онтологических типологий, требует рассмотрения, выходящего за рамки чистой онтологии, что и приводит к необходимости формально-онтологического анализа.

## **7. Онтологические обязательства логических языков**

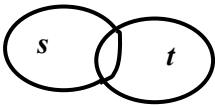
Возникающий здесь вопрос связан с взаимоотношениями структуры языка и онтологических допущений. Известно, что структура языка и мышления связана с допущениями о познаваемом. Это справедливо не только для естественных языков, но также и для искусственных, в частности, для языков логики, несмотря на то, что они более просты, что их структура более прозрачна, и, принимая их, мы заведомо абстрагируемся от ряда моментов. Как пишет В.А.Смирнов, еще Кант «... показал, что онтология как самостоятельная наука о бытии невозможна. Философия не может делать обоснованные утверждения о внешнем мире самом по себе. Означает ли это, что вся философия сводится к теории познания и логике, что все онтологические про-

блемы философии являются псевдопроблемами. На этот вопрос мы отвечаем отрицательно. Онтологические проблемы, несомненно, являются правомерными. Однако решаются они не в рамках натурфилософии и не методами, подобными естественнонаучным, а путем анализа познавательных процедур и категориальной структуры мышления» [Смирнов 1987, с. 132]. И далее: «Мы исходим из допущения, что принимаемый язык, используемые познавательные процедуры не безразличны к познаваемому; принятие того или иного языка, той или иной логики вынуждает нас сделать определенные допущения о познаваемых объектах. Одна из задач философии и состоит в том, чтобы установить связь между принимаемыми средствами выражения и рассуждения, с одной стороны, и допущениями об объектах рассуждения — с другой. И не только описать, но и четко сформулировать и обосновать. Конструирование искусственных языков и выяснение содержащихся в них онтологических допущений является хорошим средством изучения проблем онтологии» [Смирнов 1987, с. 132].

### 8. Формальные языки формальной онтологии

Подобным конструированием формальных языков, требуемых для нужд формальной онтологии, занимаются Б. Смит и К.Маллиган, стремясь при этом избежать «логизации», т.е. стремясь строить языки на совершенно иных основаниях, в отличие от того, как это делается в логике, соблюдая тем самым различие между формальной логикой и формальной онтологией, о котором говорил Гуссерль. Так, в работе [Smith Mulligan 1983] они разрабатывают онтологический язык типа диаграмм Венна, который можно представить следующим образом.

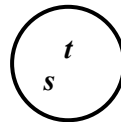
(1) Два объекта  $s$  и  $t$  существуют и они материально перекрываются:



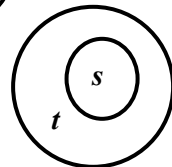
(2) Два объекта  $s$  и  $t$  существуют и они разделены:



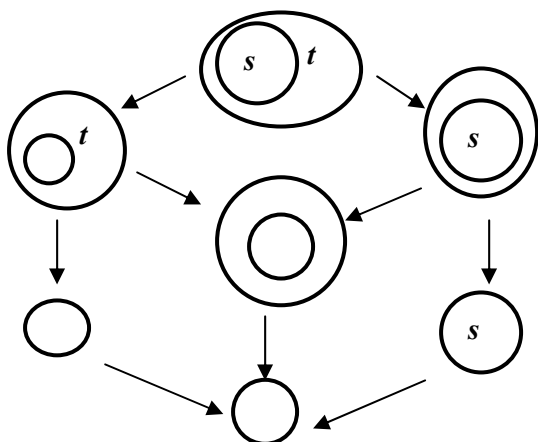
(3) Две целостности тождественны:



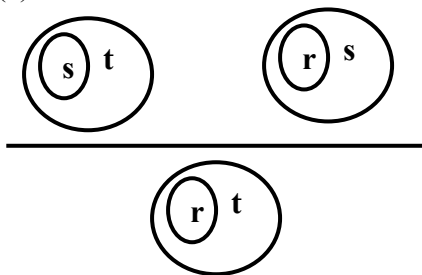
(4) Целостность  $s$  является собственной частью второй целостности  $t$ :



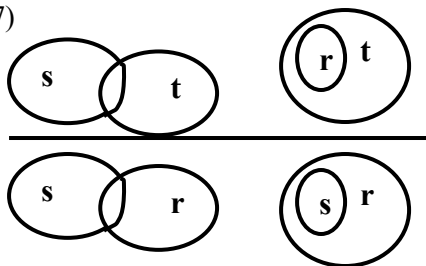
(5) Онтологический вывод



(6) Онтологический силлогизм



(7)

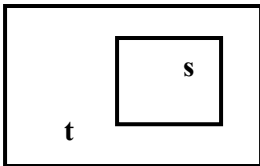


Во второй части своей работы они предлагают еще одну версию языка для формальной онтологии, так называемый онтологический язык зависимостей, основные моменты которого можно представить следующим образом.

**Онтологический язык зависимостей**

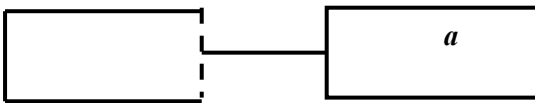
Момент = абстрактно выделяемые, но материально неотделимые части материальной вещи (от немецкого *das Moment*, обычно переводимого как «элемент», или «фактор», и контрастирующего с *der Moment*, обозначающего момент или мгновение времени).

(8)



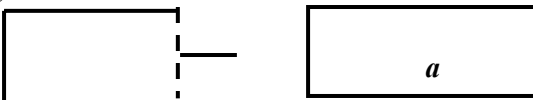
Интегральное целое *s* является частью интегрального целого *t*.

(9) *Односторонняя зависимость:*



Специфический синяк, синева или баронский титул основывается или зависит от его носителя, Альфредо.

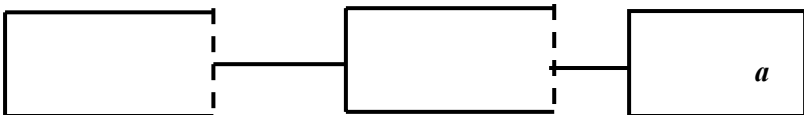
(10) *b*



Мереологическая разделенность

(11) *j*

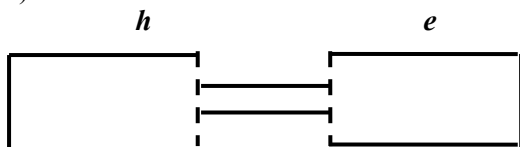
*c*



Суждение *j* (специфический ментальный акт суждения, осуществленного Альфредо в некоторый момент) основывается на определенной

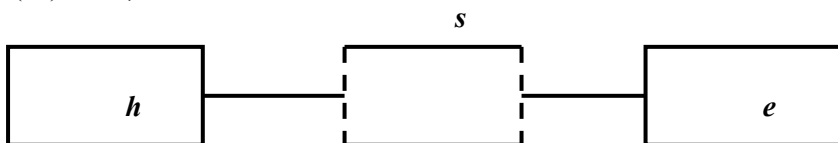
компетенции (включающей знание языка, в котором  $j$  формализовано), которая в свою очередь основывается на Альфредо.

(12) *Взаимная зависимость*



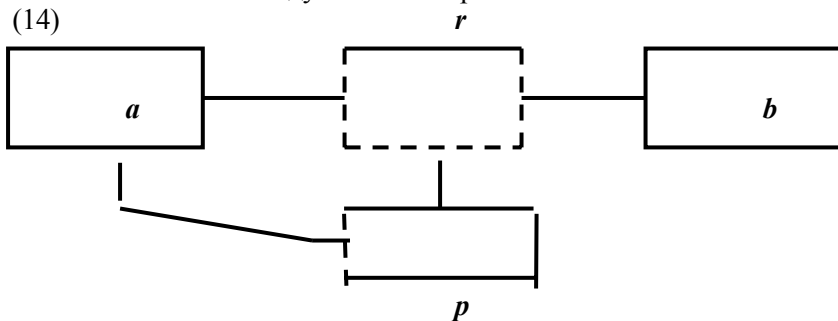
Муж Ганс и жена Эрна взаимно основываются друг на друге.  
Взаимная зависимость северного и южного полюсов магнита.

(13) *Реляционная зависимость*



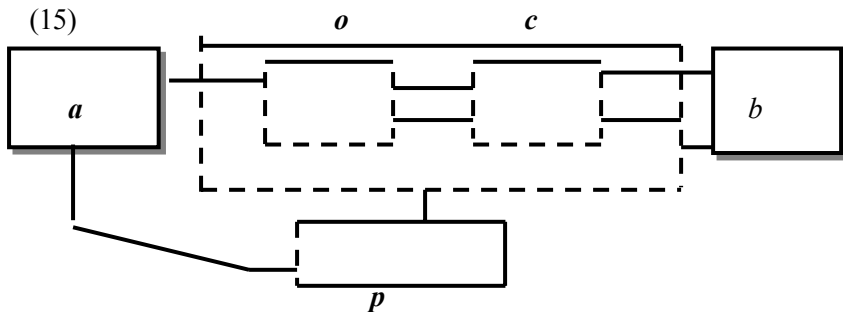
Фехтование на мечах между Гансом и Эрной

(14)



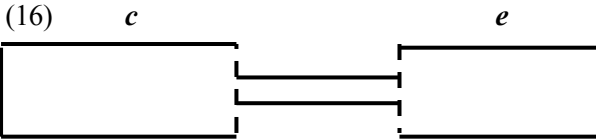
Альфредо и Бернадетто связаны материальным отношением  $r$ , которое существует в силу обещания  $p$ , чье содержание все еще не реализовано.

(15)



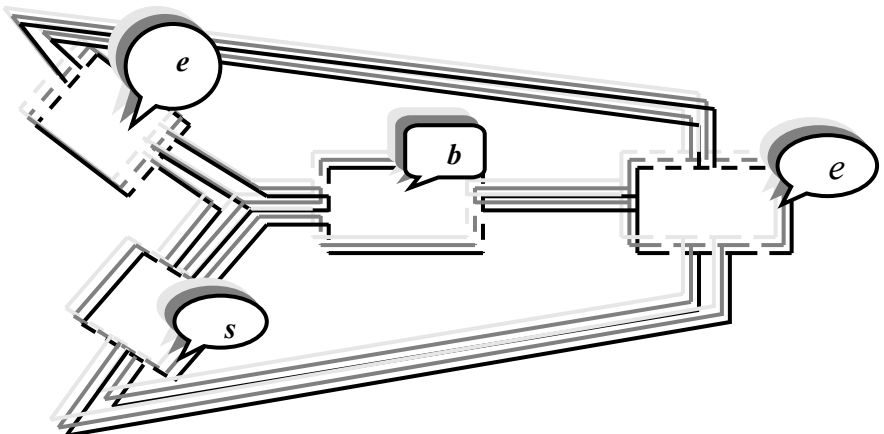
Материальное отношение состоит, в сущности, из требования (*c*) Бернадетто к Альфредо, взаимно основанное на обязательстве (*o*) со стороны Альфредо по отношению к Бернадетто.

*Моменты протяженности*



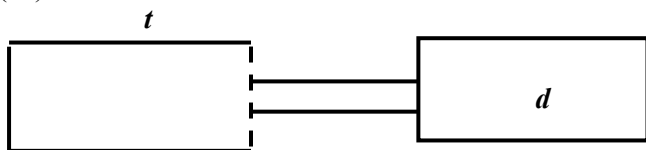
Момент цвета не может существовать иначе как цвет некоторого момента визуального диапазона, который в свою очередь не может существовать иначе как носитель некоторого цвета.

(17)



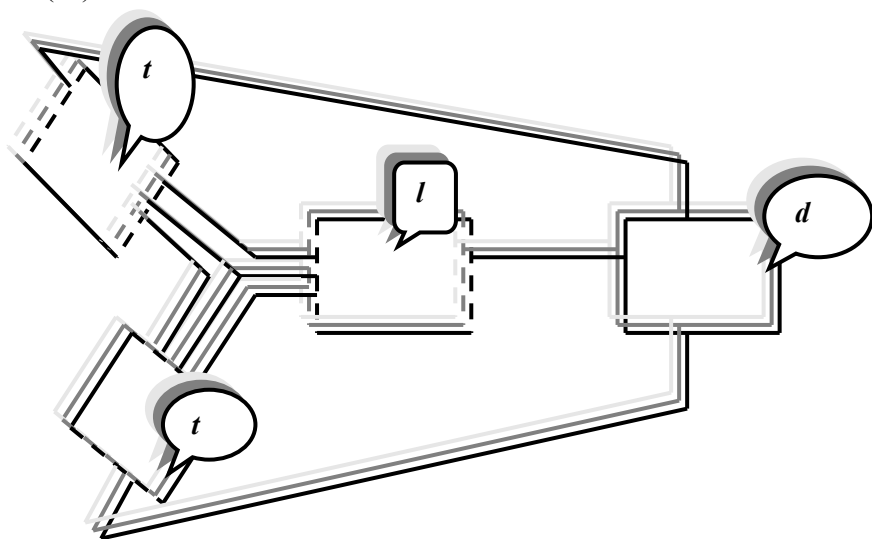
Каждый момент цвета выражает конститутивный момент оттенка, яркости и насыщенности. Оттенок цвета не может по своей природе существовать иначе как ограниченный некоторой яркостью и насыщенностью; яркость и насыщенность не могут существовать иначе как ограниченные некоторым оттенком.

(18)



Момент тона зависит от некоторого момента темпорального диапазона (каждый актуально существующий тон имеет некоторую длительность).

(19)



Тон также проявляет конститутивные моменты (высота, тембр и громкость), взаимно зависящие друг от друга.

## 9. Формальные онтологии и языки для искусственного интеллекта

Очевидным образом нет никакой необходимости использовать лишь графические языки типа диаграмм Венна для разработки более гибких языков, пригодных для целей формальной онтологии. Одним из принципиально иных подходов заключается в разработке или расширении языков программирования с помощью чисто онтологических понятий. Потребность в подобных языках давно уже начала



ощущаться разработчиками систем искусственного интеллекта. Приведем краткое описание некоторых из существующих языков формальной онтологии в рамках систем искусственного интеллекта.

#### а) GOL

Достаточно сложный язык был представлен Б. Смитом, Б. Хеллер, Х. Херре и В. Дегеном в проекте GOL (General Ontological Language – Общий онтологический язык) в 2001 г. [Degen et al 2001]. Этот язык по замыслу разработчиков призван был служить целям медицинской диагностики.

Каждая база данных компьютера, отражающая некоторую часть предметной области, обязана, по их мнению, использовать специальную онтологию верхнего уровня, которая описывает наиболее общие, независимые от особенностей работы компьютера, категории реальности. Предлагаемая Смитом и другими авторами онтология верхнего уровня служила основанием для базы знаний моделирующего языка GOL (предназначенного для нужд медицинской диагностики). В отличие от других проектов подобного рода онтология GOL не является теоретико-множественной. Недостатком использования такого мощного средства как теория множеств, является то, что множества чересчур абстрактны, они существуют вне времени, пространства и причинности. Это заставляет дополнять теоретико-множественный аппарат различными средствами, удобными для компьютерных исследований.

В данном проекте исходные сущности реального мира разбиваются на множества и урэлементы, а затем вводится несколько новых онтологических отношений между этими урэлементами. Урэлементы разделяются на индивиды и универсалии, и таким образом всего возникает три базисные категории: индивид, универсалии и множества. Индивиды, в свою очередь, разделяются на *моменты*, *субстанции*, *хроноиды*, *топоиды* и *ситуиды*. Предикаты  $Mom(x)$ ,  $Subst(x)$ ,  $Chron(x)$ ,  $Top(x)$ ,  $Sit(x)$  определяются очевидным образом.

Субстанции представляет собой то, что может существовать само по себе, или не нуждается в другой сущности для своего существования. Примеры субстанции: я и вы, луна, теннисный мяч. Моменты, по контрасту, представляют собой сущности, которые могут существовать в других сущностях (например, как электрический заряд в проводнике). Моменты включают в себя действия и страдания, рукопожатие, мысль и т.д. Некоторые моменты являются одноместными *качествами*, например, цвета или температуры. Но существуют также относительные моменты, например, поцелуи или беседы.

Ситуоиды интуитивно означают часть мира, которая может пониматься как когерентное целое и не нуждается в других сущностях для своего существования. Например: поцелуй Джоном Мэри в некотором окружении. Этот ситуоид содержит субстанции «Джон» и «Мэри» и относительный момент «поцелуй», который связывает их. Сами по себе эти сущности в изоляции не образуют ситуоид, мы должны добавить некоторое окружение, чтобы получить некоторое целое.

Специальным типом ситуоидов являются ситуации. Это ситуоиды во времени, так что они представляются собой моментальный снимок некоторого фрагмента мира.

Хроноиды и топоиды являются примером универсалий *Время* и *Пространство*. Хроноиды можно понимать как темпоральные длительности, а топоиды как пространственные регионы, имеющие некоторую меретопологическую структуру. Согласно одно из версий данной теории хроноиды и топоиды не имеют независимого существования, в каждый момент своего существования они зависят от ситуации, в рамках которой они оформлены.

Отношения являются сущностями, которые склеивают вместе вещи реального мира. Каждое отношение имеет определенное количество аргументов, которые служат для связывания. Допускаются отношения с неопределенным количеством аргументов. Отношения делятся на классы, называемые *материальными* и *формальными* соответственно. Важное формальное отношение называется поддерживающим отношением. Тот факт, что отношение поддерживается непосредственно достаточен для предотвращения регресса: поддержание поддерживается непосредственно. Среди других отношений можно выделить базисные отношения: неотделимость, тернарное отношения бытия чьей-то частью, конкретизация, обрамление, отношения содержания, ассоциации и т.д.

Наряду с этим водятся также процессы, представляющие собой переход от одной конфигурации к другой внутри некоторого ситуоида. Проблему составляет введение отношений эквивалентности между ситуоидами.

## **6) KIF**

Другим известным проектом является KIF (Knowledge Interchange Format – Формат обмена знаниями), представляющим собой формальный язык для обмена знаниями между компьютерными программами, написанными разными программистами в разное время и на различных языках. Онтология этого проекта [Genesereth Fikes 1992] принимает в качестве самой широкой категории категорию *объекта*. Это понятие достаточно широко: объекты могут быть кон-

кретными (напр., кусок скалы, Ницше, молекула) или абстрактными (понятие справедливости, число два), объекты могут быть простыми или сложными, и даже фиктивными (напр., единорог). Принимается деление на индивиды и множества. Множество представляет собой совокупность объектов, индивидом является объект, не являющийся множеством. Функции и отношения вводятся как множества конечных списков. Онтологический базис KIF слабее базиса GOL. KIF можно понимать как теоретико-множественную часть GOL.

### **в) высокоуровневая онтология Рассела и Норвига**

Наиболее общими категориями здесь [Russell Norvig 1995] являются категории *абстрактных объектов* и *событий*. Абстрактные объекты разделяются на множества, числа и репрезентирующие объекты.

События можно классифицировать как *интервалы*, *места*, *физические объекты* и *процессы*. К сожалению, нет четкого различия между множествами, универсалиями и индивидами. Нет также категории формальных отношений. Класс универсалий (называемых здесь «категориями») является подклассом класса множеств. *Событием* в онтологии Рассела-Норвига является то, что они называют «куском» частного универсума темпоральным и пространственным измерением. Интервал – это событие, включающее в себя как подсобытия все события, произошедшие в данный период времени. Подобные интервалы можно, в некотором смысле, понимать как ситуиды. Но разница в том, что GOL-ситуиды являются частью реального мира, которые рассматривается как некоторые целостности.

### **г) высокоуровневая онтология Соувы**

В онтологии Дж. Соувы [Sowa 2000] также нет четкого разделения между множествами, универсалиями и индивидами. Его интересует главным образом различие между классами и единичными сущностями. Учитывая, что эти понятия интерпретированы в KIF, их можно понимать как соответствующие тому, что в GOL названы множествами и урэлементами. Имеются следующие двуместные примитивные отношения: иметь, быть примером чего-л., быть подклассом чего-л., быть темпоральной частью чего-л., быть пространственной частью чего-л. Отношение «быть примером» интерпретируется отношением принадлежности, а онтологический статус отношения «иметь» совершенно неясен. Онтология Соувы использует два эпистемических оператора *nes* и *poss*, которых нет в KIF. Однако, к сожалению, вновь неясен онтологический характер этих операторов.

### **д) LADSEB**

В [Gangemi et al 2001] и [Guarino 98] описываются некоторые принципы высокоуровневой онтологии проекта LADSEB. Формальные от-

ношения в рамках LADSEB рассматриваются как отношения, которые могут возникать между сущностями во всех материальных сферах. Примеры формальных отношений, рассматриваемых в [Gangemi et al 2001] включают в себя конкретизацию, принадлежность, частичность, связь, местоположение и расширение, а также зависимость. Формальные свойства включают конкретность, абстрактность, экстенциональность, единство, множественность, зависимость и независимость. Конкретность, принадлежность и частичность являются базисными отношениями в смысле GOL.

#### е) SUO

SUO представляет собой проект, спонсируемый IEEE с целью дальнейшей разработки «Стандартной высокоуровневой онтологии», основанной на KIF [SUO 2001]. Этот проект разрабатывается для обеспечения определениями от 1000 до 2000 терминов общего назначения таким образом, чтобы создать общую структуру для онтологии низкоуровневой области достаточно большого размера и более специфического диапазона. В сущности SUO представляет собой консервативное расширение высокоуровневой онтологий Дж. Соувы и Рассела-Норвига, получаемого путем добавления большого количества новых понятий.

\* \* \*

Развитие высокоуровневых онтологий, хорошо обоснованных и аксиоматизированных, является несомненно важным шагом на пути к обоснованию формальной онтологии в информационных системах. Каждая уровнево-специализированная онтология в качестве своих рамок должна иметь некоторую онтологию высшего уровня, которая описывала бы наиболее общие категории реальности, не зависящие от уровня рассмотрения.

ЛИТЕРАТУРА

- [Смирнов 1987] *Смирнов В.А.* Логические методы анализа научного знания. М.: Наука, 1987.
- [Твардовский 1997] *Твардовский К.* Логико-философские и психологические исследования. М.: РОССПЭН. 1997.
- [Cocchiarella 1974] *Cocchiarella N.* Formal Ontology and the Foundations of Mathematical // *Bertran Russell Philosophy / G Nakhikian (ed).* Duckworth, London. 1974. pp. 29-46.
- [Degen et al 2001] *Degen W., Heller B., Herre H., Smith B.* GOL: A General Ontological Language // <http://www.ontology.uni-leipzig.de>[Gangemi et al 2001] *Gangemi A., Guarino N., Masolo C., Otramari A.* Understanding Top-level Ontological Distinctions // Technical Report 04/2001, LADSEB-CNR.
- [Genesereth Fikes 1992] *Genesereth V.R., Fikes R..E.* Knowledge Interchange Format, Version 3.0, Reference Manual. Logic Group Report Logic-92-1, Computer Science Department, Stanford University.
- [Guarino 98] *Guarino N.* Formal Ontology and Information Systems // *Ajhvfl Ontology in Information Systems. Proceedings of FOIS»98, Trento. Italy / N. Guarino (ed.), Amsterdam, IOS-Press.*
- [Perzanowski 1994] *Perzanowski J.* Ontologic // *Logic and Logical Philosophy.* No 2. 1994. P. 4.
- [Poli 1992] *Poli R.* *Ontologia Formale.* Genova. Marietti. 1992.
- [Russell Norvig 1995] *Russell S., Norvig P.* Artificial Intelligence. Prentice Hall. 1995.
- [Simons 1982] *Simons P.B.* The Formalization of Husserl»s Theory of Wholes and Parts // *Parts and Moments. Studies in Logic and Formal Ontology/B. Smith (ed.), Munich: Philosophia, 1982. P. 113-159.*
- [Smith Mulligan 1983] *Smith B. and Mulligan K.* Framework for Formal Ontology // *Topoi, v.2. 1983. P. 73-85.*
- [Sowa 2000] *Sowa J.* // <http://bestweb.net/sowa/ontology/toplevel.htm>.
- [Stachniak 1981] *Stachniak Z.* Introduction to model theory for Leśniewski»s Ontology. Wrocław, 1981. [SUO2001] <http://suo.icee.org>.

---

# О ЛОГИКЕ СИМВОЛЬНЫХ ВЫРАЖЕНИЙ ДЛЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*С.А. Павлов*

С.А. Павлов

Обработка информации компьютерными системами включает в себя преобразования и различные операции с символьными выражениями языка. Целью этой работы является обсуждение языка логики, которая допускает логические операции в области символьных выражений языка. Символьным выражением некоторого языка  $L$  называется любая конечная линейная последовательность (упорядоченная  $n$ -ка) символов из алфавита этого языка  $L$ . Синонимом *символьного выражения* являются *слово*, *выражение* или *строка* в алфавите [5]. Отметим, что в языках программирования имеется несколько типов переменных: числовые, логические (булевы), строковые и др.

Обычно понятия истинности и ложности применяют к высказываниям и предложениям. А.Тарский пишет об этом так: «Предикат «истинно» ... относят к определенным физическим объектам – языковым выражениям, в частности, к предложениям» [6]. В то же время имеются трудности, связанные с определением того, что есть высказывание и предложение. А.Тарский пишет об этом: «Мы не знаем в точности, какие выражения являются предложениями» [6]. Там же А.Тарский говорит о новых возможностях: «тот факт, что нас прежде всего интересует понятие истины для предложений, не исключает возможности последующего расширения сферы применимости этого понятия на другие виды объектов».

У Г.Фреге в его статье «Функция и понятие» [7] имеется пример такого расширения. «Теперь можно рассмотреть некоторые функции, которые для нас важны именно тогда, когда их аргументом является истинностное значение». В качестве такой функции он вводит функцию, изображаемую в виде горизонтальной черты, —  $x$ , устанавливая, что «значением этой функции должна быть истина, когда в качестве аргумента берется истина, во всех же остальных случаях ее значение есть ложь – стало быть и тогда, когда он вообще не является значением истинности. В соответствии с этим, например,

—  $1+3 = 4$ ,

есть истина, тогда как

—  $1+3 = 5$ ,

есть ложь, так же как

—  $4$ » (см. [7]).

Выражение «4» не есть предложение, в отличие от предыдущих аргументов функции, но, тем не менее, Г.Фреге не затрудняется определить значение функции —  $x$  с таким аргументом.

Подобными свойствами в языке трехзначной логики Бочвара [2] обладает оператор внешнего утверждения «верно», а именно: формула « $x$  верно» принимает значение «истина», если вместо  $x$  подставляется высказывание, имеющее значение «истина», формула « $x$  верно» принимает значение «ложь», если вместо  $x$  подставляется высказывание, имеющее значение «ложь», и, наконец, формула « $x$  верно» принимает значение «ложь», если вместо  $x$  подставляется высказывание, имеющее значение «бесмыслица», то есть ни истинное, ни ложное высказывание.

Язык логики, которая допускает логические операции в области символьных выражений языка, строится следующим образом. К языку сентенциальной логики добавляем операторы истинности и ложности. Расширяем применение понятия истинности и ложности на класс выражений языка, которые определяются индуктивно. Такое расширения сферы применимости понятий истинности и ложности на универсум символьных выражений языка не ведет к трудностям или неясностям, так как все выражения, которые не являются предложениями (для любого определения предложения), заведомо ни истинны, ни ложны. Язык получаемой логики FSL4 (см. ее формулировку в [4]) двухуровневый, подобно языку логики Бочвара.

В ней доказывается тетралемма истинности и ложности, говорящая, что каждое символьное выражение языка либо истинно и неложно, либо ложно и неистинно, либо истинно и ложно, либо ни истинно, ни ложно.

Логика FSL4 для искусственного интеллекта и компьютерных рассуждений, предназначенная обрабатывать, в дополнение к двузначным высказываниям, высказывания, содержащие противоречивую или неполную информацию (см. также [10]) имеет четырехзначную интерпретацию со следующими истинностными значениями:

T – «строгая истинность», то есть «истинность и неложность»,

F – «строгая ложность», то есть «ложность и неистинность»,

V – «противоречивость», то есть «истинность и ложность»,

N – «индифферентность», то есть «ни истинность, ни ложность».

Эти значения близки по смыслу значениям четырехзначной логики Белнапа, предложенной им в статье «Как нужно рассуждать компьютеру» [1]:

- T – «говорит только Истину»,  
F – «говорит только Ложь»,  
B – «говорит и Истину и Ложь»,  
N – «не говорит ни Истины, ни Лжи».

Отметим, что логика FSL4 является обогащением и обобщением логики Белнапа (см. [4]) и в ее языке можно формализовать ряд дополнительных соотношений между высказываниями, которые были содержательно приняты Н.Белнапом, но их нельзя формально выразить в языке логики Белнапа.

Данн получил четыре значения, рассматривая подмножества двухэлементного множества  $\{T\}$ ,  $\{F\}$ ,  $\{T,F\}$ ,  $\{\}$ . Он также приводит в [9] пример тетралеммы (чатушкотти, fourcorner) индийского логика Санджая (6 век до нашей эры):

- a. S есть P.
- b. S есть не-P.
- c. S есть P и не-P.
- d. S есть ни P, ни не-P.

Аналогичные конструкции использовали в своих рассуждениях буддийские философы.

Значения истинности, подобные вышеприведенным, предложил фон Бригг в [3] для одной из логик истины: «Истинно и ложно» («true and false»), «истинно, но не ложно» («true but not false», «univocally true»), «ложно, но не истинно» («false but not true», «univocally false»), «ни истинно, ни ложно» («neither true nor false»), которые обозначаются им как «1», «+», «-», «0» соответственно.

Для логики FSL4 можно построить и другие семантики. Н.Белнап предложил автору рассмотреть для интерпретации логики FL4 бирешетки, введенные Фиттингом. Также представляет интерес построить семантику в собственном смысле слова, исходя из идеи Фреге о том, что предложения являются именами денотатов, являющихся абстрактными предметами «истина» и «ложь». При этом все истинные предложения обозначают истинностное значение истину, а все ложные предложения – истинностное значение ложь. Точка зрения Фреге, согласно Черчу [8], может быть передана утверждением, что ситуация указывает на существование таких двух предметов, как истина и ложь (или, по Н.Белнапу [1], онтологических значений «истина» и «ложь»).

Можно показать, что достаточно принять существование только одного абстрактного предмета «истина», чтобы построить необходимую для наших задач семантику. Положение предыдущего абзаца модифицируется следующим образом: все истинные предложения обозначают истину, а все ложные предложения не обозначают истину.



Говоря другими словами, нет необходимости в допущении существования такого абстрактного предмета как «ложь». Поэтому, будем исходить из того, что истина существует, а ложь не существует.

В классическом случае, если предложение  $A$  обозначает истину, то  $\sim A$  не обозначает истину или, если предложение  $\sim B$  обозначает истину, то  $B$  не обозначает истину.

В неклассическом случае, соответствующем четырехзначной логике Белнапа и логике FSL4, предложению  $A$  поставим в соответствие упорядоченную пару предложений  $\langle A, \sim A \rangle$ , каждое из которых независимо одно от другого обозначает, либо не обозначает истину. Тем самым для различных пар предложений имеем четыре возможных варианта денотации:

$\langle A_1, \sim A_1 \rangle$   $A_1$  обозначает истину, а  $\sim A_1$  не обозначает истину.

$\langle A_2, \sim A_2 \rangle$   $A_2$  не обозначает истину, а  $\sim A_2$  обозначает истину.

$\langle A_3, \sim A_3 \rangle$   $A_3$  обозначает истину и  $\sim A_3$  обозначает истину.

$\langle A_4, \sim A_4 \rangle$   $A_4$  не обозначает истину и  $\sim A_4$  не обозначает истину.

В этой бисентенциальной семантике выразимы все соотношения, необходимые для интерпретации логик, предназначенных для компьютерных рассуждений, то есть логики Белнапа и логики FSL4 (в последнем случае область предложений расширяется до универсума символьных выражений).

Отметим, что такая бисентенциальная семантика позволяет выделить классы логик, семантически основанных только на истине (одном денотате «истина»), тем самым исходя из утверждения Фреге, что «логика есть наука о наиболее общих законах бытия истины». При этом необходимо использовать различные возможные зависимости или их отсутствие между высказываниями о денотации для предложения  $A_k$  и его отрицания  $\sim A_k$ , имеющих следующий вид:

если  $A_k$  обозначает истину, то  $\sim A_k$  не обозначает истину;

если  $\sim A_k$  не обозначает истину, то  $A_k$  обозначает истину.

При этом семантика с единственным денотатом «истина» может быть согласована с двух-, трех- и четырехзначными математическими интерпретациями этих логик.

В заключение приведем схему, в которой приведены различные характеристики интеллекта, соответствующие упоминаемым в данной статье логическим системам, имеющим отношение к обсуждаемой логике символьных выражений для искусственного интеллекта FSL4.

	искусственный	
	естественный	
современный,		
западный	интеллект,	
	мышление	
		восточный,
		древний

#### ЛИТЕРАТУРА

1. *Белнап Н.* Как нужно рассуждать компьютеру // Белнап Н., Стил Т. Логика вопросов и ответов, М., 1981.
2. *Бочвар Д.А.* Об одном трехзначном исчислении // Математический сборник. 1938. Т.4. N2.
3. *Вригт Г.Х. фон* Логика истины // Вригт Г.Х. Логико-философские исследования. М., 1986.
4. *Павлов С.А.* Логика с операторами истинности и ложности. М., 2004.
5. *Смальян Р.* Теория формальных систем. М., 1981.
6. *Тарский А.* Семантическая концепция истины // Аналитическая философия: Становление и развитие. М., 1998.
7. *Фреге Г.* Функция и понятие // Готтлоб Фреге Логика и логическая семантика, М., 2000.
8. *Чёрч А.* Введение в математическую логику. М., 1960.
9. *Dunn J.M.* Partiality and its Dual // *Studia Logica*. Vol. 65. 2000. P.5-40.
10. *Pavlov S.A.* Logic For Computer Reasoning // International Conference on Informatics and Control, St. Petersburg, 1997. P.496-499.

---

# ПРОГРАММА КОЛМОГорова, ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕОРЕМЫ ГЕДЕЛЯ О НЕПОЛНОТЕ

*А. С. Кузичев*

Значительное место в современных исследованиях в области построения систем знаний занимают *нелогические* исчисления комбинаторов Шейнфинкеля–Карри и  $\lambda$ -конверсий Чёрча, на базе которых автором построены двухъярусные *логические* секвенциальные (без правила сечения) доказуемо непротиворечивые интеллектуальные системы, образующие, в частности, доказуемо непротиворечивые (как абсолютно, так и относительно отрицания  $\perp$ ) основания КМ (классической *теоретико-множественной* математики).

При этом одновременно решаются две хорошо известные проблемы оснований наук: проблема введения логических операторов в алгоритмические комбинаторно полные ( $\lambda$ -полные) неразрешимые исчисления Шейнфинкеля–Карри–Чёрча и Центральная проблема Гильберта построения доказуемо полных и доказуемо непротиворечивых оснований КМ.

Впервые формулируется программа Колмогорова по основаниям КМ. Обсуждаются историко-методологические вопросы её становления.

Работа развивает идеи и результаты, содержащиеся, в частности, в двух моих публикациях: «Паранепротиворечивость интеллектуальных систем компьютерных логик» (1989) [1] и «О роли теорем Гёделя о неполноте в основаниях наук» (2005) [2].

## **I. Центральная проблема Гильберта в основаниях математики**

Впервые предлагается вариант решения Центральной проблемы Давида Гильберта (1862–1943) построения доказуемо полных и доказуемо непротиворечивых оснований КМ в виде одного (но не аксиоматического, а *двухъярусного секвенциального без постулируемого правила сечения*) исчисления (теории).

Эта проблема Гильберта долго не поддавалась математикам. Многие даже были уверены, что решить её невозможно. При этом часто ссылаются на теоремы Гёделя 1931 года о неполноте богатых (по выразительным возможностям) формульных аксиоматических теорий первого порядка.

Видимо, первым был известный ученик Гильберта Герман Вейль (1885–1955), высказавшийся, что Гёдель нанес планам Гильберта ре-

шения Центральной проблемы «сокрушительный удар», от которого математика «не оправилась до сих пор» [3, с. 339–340]. Эта точка зрения Г. Вейля ничем не подкреплена, но тем не менее получила широкую известность и стала господствующей в научном мире (см., например, [4, 5]).

Как хорошо известно, на формульном аксиоматическом пути Готлоба Фреге (1848–1920) Центральная проблема Гильберта не решается: неограниченное теоретико-множественное свёртывание Георга Кантора (1845–1918), представленное аксиоматически в виде одной или нескольких формульных схем (задающих собственные аксиомы теории), вместе с соответствующей аксиоматически заданной логикой предикатов первого порядка, *ведет к противоречию*.

Наиболее важным и простым по изложению является широко известный парадокс Бертрانا Рассела (1872–1970), найденный в 1902 году в опубликованной системе Г. Фреге.

Из общих соображений очевидно, что возможны различные способы преодоления противоречий. Один – наложить ограничения на аксиомы свёртывания (этот способ почему-то без объяснений считается единственным!). Другой, излагаемый ниже в данной работе, связан с двухъярусными секвенциальными системами, порождаемыми программой Колмогорова, охватывающей все основания КМ *в её целостности*.

Различные же ограничения на схемы аксиом свёртывания привели к построению известных аксиоматических теорий первого порядка, выражающих на пути Фреге (о путях Фреге и Колмогорова см. в [6]) неполным (по Гёделю) образом определенные разделы *современной* математики. Само разбиение математики на разделы (математический анализ, алгебра, теория вероятностей и т.д.) обусловлено только различными ограничениями на аксиомы свёртывания. Тем самым современная математика (с её разбиением на различные разделы) отличается от целостной классической теоретико-множественной математики.

На фрегевском пути все постулаты известных теорий (1-го порядка), естественно, выбирались и выбираются так, чтобы эти теории были непротиворечивыми. Но доказать непротиворечивость каждой известной теории  $G$  – трудная проблема. Здесь используются термины «доказуемая непротиворечивость (теории)  $S$ » и «непротиворечивость  $S$ »; первый – точный, а второй – нет. Если в  $S$  выведено противоречие, то  $S$  перестает считаться «известной» и далее не исследуется.

К числу известных аксиоматических теорий относятся формульные исчисления гильбертовского типа, например, арифметики  $FA$  Пеано, теории множеств  $ZF$  Цермело–Френкеля, теории множеств  $NBG$  Неймана–Бернайса–Гёделя, теории множеств  $NF$  Куайна. Для большинства известных теорий доказуемая непротиворечивость на фрегевском пути пока не найдена и не опровергнута, а для таких, как исчисление

ФА арифметики Пеано, доказательства непротиворечивости весьма громоздки, используют теоретико-множественную индукцию.

Только в 2004 году в [6] впервые опубликовано негёделевское построение теорий первого порядка, дано доказательство непротиворечивости всех известных формульных аксиоматических теорий с соответствующим историко-методологическим обоснованием. Предварительно в [6] каждая такая теория  $K$  перестраивается теоретико-множественно по Колмогорову и для неё доказывается теорема 1 о редукции в логику высказываний. (См. публикацию [7] 1999 г. и замечания к ней в [6, с. 31–32]).

Доказательство непротиворечивости в [6, 7] известной теории  $K$  осуществлено в два этапа хорошо знакомыми комбинаторными средствами: на 1-ом этапе доказана теорема 1 (о редукции  $K$  в логику высказываний); на 2-ом этапе доказана теорема 2 (о непротиворечивости  $K$ ).

Доказательство теоремы 2 для  $K$  полностью зависит от теоремы 1 для  $K$  и не может быть перенесено на каждую теорию, например, на теории с правилом  $MP^*$  (см. [6, 7]) – в частности, не может быть получена *нелепость*: «доказательство непротиворечивости противоречивой теории».

Предыдущий абзац очевидно следует из двух этапов доказательства непротиворечивости известной теории  $K$  и поэтому в силу его очевидности часто не указывается, например, в [6, 7], что, к сожалению, иногда ведет к трудностям. Дело в том, что на колмогоровском пути теоретико-множественной общности некоторые факты доказываются весьма просто и неожиданно понятно!

Подчеркнем значение леммы 1, названной в [6] критерием непротиворечивости известной (но не любой!) теории  $K$ . Лемма 1 впервые сформулирована в [6, 7] и утверждает (доказывает), что все аксиомы теории  $K$ , по их виду, не являются ни  $W$ -формулами (см. в [6, 7] определение 1), ни Выделенными формулами (см. определение 2).

Доказательство леммы 1 проведено в [6, 7] комбинаторными средствами сравнения слов в алфавитах.

На основе леммы 1 получено в [6, 7] доказательство главной теоремы 1 о редукции множества  $M$  всех выводов теории  $K$  в логику высказываний.

Следствием теоремы 1 является в [6, 7] теорема 2 о непротиворечивости теории  $K$ , доказываемая методом от противного.

Критерий непротиворечивости (точнее, доказательства непротиворечивости) известных аксиоматических теорий первого порядка в виде леммы 1 найден (получен) на колмогоровском пути теоретико-множественной общности в основаниях математики, когда выбор всех аксиом и правил вывода каждой теории  $K$  осуществлен классически по Э. Мендельсону.

Как работает этот критерий, фактически показано в [6], например, на стр. 13 на различных, но эквивалентных исчислениях арифметики.

## II. Ещё раз о теоремах Гёделя

В 1931 году Курт Гёдель (1906–1978) на фрегевском пути опубликовал теоремы о неполноте многих известных теорий. Эти теоремы полностью обоснованы (доказаны). Никаких сомнений, казалось, нет. Более того, у многих сомневающихся находились конкретные ошибки.

Я был потрясен, когда узнал, что А.Н. Колмогоров относит себя к сомневающимся в теоремах Гёделя о неполноте. Нет, он не оспаривал результаты Гёделя, относящиеся к конкретным исследуемым теориям, но он не верил в распространение этих теорем без доказательства на все известные теории при любых их построениях. Он так и говорил мне: «А где доказательство?»

Действительно, нет доказательства, что теоремы Гёделя распространяются всеобъемлющим образом на все основания наук. А без доказательства Колмогоров не мог признать истинным обобщение этих теорем на все теории (при любых их построениях).

Надо сказать, что и сам Гёдель выражал некоторое сомнение в величии и универсальности своих результатов о неполноте, особенно следствий из них; см., например, [5].

Биограф Гёделя Г. Крайзель пишет, что «вопреки усилиям... представить результаты Гёделя как сенсацию, эти результаты не оказали революционизирующего влияния ни на представление большинства работающих математиков о своей науке, ни тем более на их практическую деятельность. Во всяком случае, их влияние намного меньше, чем влияние внутреннего развития самой математики.» [5, вып. 2 (260), с. 175]; *подчеркнуто мною – А.С.К.*

А как мы преподаем основания не только математики, но и всех наук, особенно теоретических? Принято почти в самом начале соответствующих курсов или семинаров ссылаться на теоремы Гёделя о неполноте (часто даже не формулируя их) как на ограничительные – запрещающие многое сделать в рассматриваемой области знания (как будто эти запреты в них доказаны или доказуемо следуют из них).

Взгляды Колмогорова на теоремы Гёделя о неполноте перевернули всю мою жизнь, особенно учитывая догматическую веру в эти теоремы моих ближайших родственников, коллег, учеников и подавляющего большинства математиков и философов.

В нашей педагогической практике такая догматическая вера в теоремы Гёделя и следствия из них по существу навязывается всем учащимся; доказательства, весьма громоздкие и сложные, разбираются лишь на узкоспециальных занятиях с небольшим числом заинтересованных студентов, да и в основном, как я сейчас глубоко убежден, занятия эти проводятся фактически с целью не разобрать все воз-

можные случаи (что малореально), а только усилить веру в результаты Гёделя и их обобщения.

Высказывания Колмогорова относительно теорем Гёделя были мало известны. Под руководством Колмогорова я работал с января 1980 по октябрь 1987 года, когда Андрей Николаевич заведовал кафедрой математической логики механико-математического факультета МГУ (я был сотрудником этой кафедры).

По поводу теорем Гёделя я спорил с Андреем Николаевичем, однако, следуя его рекомендациям, изучал различные теории, прежде всего теоретико-множественные.

В результате мною впервые были найдены доказательства непротиворечивости многих известных теорий — доказательство непротиворечивости каждой теории строится секвенциально по Генцену на основе интеллектуальной системы – неразрешимого алгоритмического (но не логического!) аппарата одного из комбинаторно полных исчислений Шейнфинкеля – Карри – Чёрча, представляющего бестиповым образом неограниченное теоретико-множественное свёртывание Кантора (1845–1918).

Многим такие доказательства кажутся (без предъявления серьезных математических или философских обоснований), особенно в силу их длиннот и использования исчислений (без учета их специфики), очевидно противоречащими теоремам Гёделя о неполноте.

Мои результаты после обсуждений А.Н. Колмогоров представлял в печать – работы опубликованы в ДАН и других изданиях ([8–17]).

Андрей Николаевич подчеркивал значимость полученных результатов. Он не только отмечал при этом, что его редукция 1925 года (см. [18]) позволяет, используя теоретико-множественную общность, значительно упростить построения и доказательства, сделав их общепонятными и общедоступными, но и впервые обратил внимание на правила вывода теорий, два этажа (посылки и заключение) которых могут быть основой упрощений. Важно при этом выбрать среди всех логически эквивалентных выводимых формул подходящие аксиомы для каждой теории.

Только к концу XX века я понял, что А.Н. Колмогоров был прав в своих сомнениях относительно роли теорем Гёделя о неполноте в основаниях современных наук.

В работах автора (см. [8–17] и предшествующие труды автора) все исследования ведутся на фрегевском пути. Однако полученные в них результаты находятся в полном соответствии с теоремами Гёделя о неполноте, поскольку предложенные автором системы и их метатеория основываются на *неразрешимом* бестиповом нелогическом *аппарате* теории алгоритмов (алгорифмов) в форме *исчислений* чистой комбинаторной логики Шейнфинкеля–Карри или  $\lambda$ -конверсии Чёрча. В данном случае теоремы Гёделя о неполноте для предложенных

систем даже не формулируются (подробнее об этом сказано в [6], особенно см. гл. III). В [6] же впервые рассказано о действительном месте этих теорем Гёделя в основаниях современной науки при исследовании всех аксиоматических теорий первого порядка.

Таким образом, все известные теории первого порядка доказуемо непротиворечивы (см. [6, 7]) и достаточно богатые из них доказуемо неполны (см. теоремы Гёделя о неполноте). Тем самым Центральная проблема Гильберта не может быть решена на фрегевском пути формализации различных разделов современной науки.

В силу теорем Гёделя о неполноте некоторых аксиоматических теорий 1-го порядка и в силу уже упомянутых парадоксов типа парадокса Рассела, доказуемо полные и доказуемо непротиворечивые интеллектуальные системы (такие, например, как интеллектуальные системы компьютерных логик), в частности, реализующие программу Колмогорова по КМ, естественно строить в их основной логико-предикатной части не как всем известные аксиоматические формульные теории 1-го порядка с собственными нелогическими аксиомами, а как вторые, логические, ярусы секвенциальных двухъярусных теорий, базирующихся на знаменитых результатах Кантора, Чёрча и Генцена в основаниях наук и введенных автором по публикациям с 1970 года. Основное внимание сосредоточим на КМ, естественно предполагая распространение результатов на любые интеллектуальные конструкции в их целостности.

### **III. Программа Колмогорова по основаниям классической теоретико-множественной математики**

Только Андрей Николаевич Колмогоров (1903–1987) высказал принципиально новые идеи, приведшие к решению Центральной проблемы Гильберта.

*Программа Колмогорова по основаниям КМ* состоит в том, что при построении исчисления, решающего Центральную проблему Гильберта, надо учитывать две компоненты КМ (классической теоретико-множественной математики) – вычислительную (алгоритмическую) и дедуктивную (логическую) и одновременно отражать без ограничений два канторовских принципа теории множеств – неограниченное свертывание и неограниченную логику, применением которых строятся все выводы КМ.

Программа Колмогорова обусловлена его работой 1925 г. «О принципе Tertium non datur» [18, 19] и сформулирована Андреем Николаевичем автору в 1980–1987 гг. (см. статьи [20] «Вариант формализации канторовской теории множеств» (1999 г.), [21] «Решение проблемы Гильберта по Колмогорову» (2000 г.) и настоящую работу).

Автор осознал существование и величие программы Колмогорова только к 1994 году, когда познакомился с анализом Константина



Алексеевича Рыбникова (1913–2004) в [22] широко известной программы 1666 года [23] Готфрида Вильгельма Лейбница (1646–1716). Анализ в [22] проведен без ссылок на известную специфическую интерпретацию теорем Гёделя 1931 года о неполноте как ограничительных в рассматриваемой области знания.

На основе общности программ Лейбница и Колмогорова историко-методологический анализ ряда исследований, проведенных в XIX–XXI вв. от Д. Гильберта до наших дней (особенно школами: Гёттингенской – Гильберта и Московской – Колмогорова), показал:

- принципиальное отличие предложенных и опубликованных автором с 1970 г. секвенциальных исчислений, имеющих два яруса, от широко известных одноярусных Аристотелевских аксиоматических исчислений Гильбертовского типа (с ограничениями теоремами Гёделя о неполноте);

- необходимость изучения по Колмогорову канторовской («наивной») теории множеств с её двумя неограниченными принципами, на основе которых строятся все выводы единой КМ (без каких-либо ограничений).

В работе предлагается вариант реализации программы А.Н. Колмогорова по основаниям математики. Впервые показывается, что КМ (классическая теоретико-множественная математика) доказуемо полностью и доказуемо непротиворечиво (как абсолютно, так и относительно отрицания  $\neg$ ) представляется одним исчислением (теорией).

Теория строится ступенчато по Андрею Андреевичу Маркову (1903–1979) в виде двухъярусного секвенциального исчисления, названного автором **КЧГ** (исчислением Кантора–Чёрча–Генцена): первый ярус задаёт неограниченное теоретико-множественное свёртывание Г. Кантора в алгоритмической (вычислительной) форме исчисления  $\lambda$ -конверсии Алонзо Чёрча (1903–1995); второй ярус задаёт классическую логику (предикатов 1-го порядка) в секвенциальной (без постулируемого правила сечения) форме Герхарда Генцена (1909–1945); связь между ярусами обеспечивают правила  $*\lambda$  и  $\lambda*$  (пишем:  $*\lambda*$ ), введенные автором и названные *канторовскими* (см., например, [20]). Такие двухъярусные исчисления строятся, следуя идее ступенчатых конструкций А.Н. Колмогорова и А.А. Маркова, исследуются автором на механико-математическом факультете МГУ с 1968 г. и публикуются с 1970 г.

Тем самым впервые Центральная проблема Гильберта построения доказуемо полных и доказуемо непротиворечивых (как абсолютно, так и относительно отрицания  $\neg$ ) оснований классической теоретико-множественной математики КМ в виде одного (хотя и двухъярусного) исчисления решается автором по Колмогорову.

На основании сказанного теория КЧГ строится как двухъярусное секвенциальное исчисление  $\mathbf{M}$  (без постулируемого правила сечения) из [20, 21]. В отличие от [20, 21] дедуктивные секвенции как слова вида  $(\Sigma \Rightarrow \Phi)$  имеют наборы  $\Sigma$  и  $\Phi$ , в которые могут входить только выводимые в данной работе *оснащенные*  $\mathbf{M}$ -формулы (оснащение  $\mathbf{M}$ -формул в работе осуществляется с помощью *индикаторов*).

Результат получает завершение ниже формулируемой и доказанной Теоремой Cut (о допустимости в КЧГ правила сечения), из которой в дополнение к абсолютной непротиворечивости [20,21], вытекает непротиворечивость КЧГ относительно отрицания  $\neg$ : не существует оснащенной  $\mathbf{M}$ -формулы  $B^s$  такой, что в КЧГ выводимы две секвенции  $(\Rightarrow B^s)$  и  $(\Rightarrow (\neg B)^s)$ , где  $B$  есть  $\mathbf{M}$ -формула,  $s$  и  $s'$  суть индикаторы.

Формулировка и доказательство Теоремы Cut в случае КЧГ осуществляются точно так, как это делается в соответствующей секвенциальной (генценовской) логике предикатов (1-го порядка) без постулируемого правила сечения с тем *новшеством*, что и аналог ранга логических формул участвует в построении двухъярусного исчисления КЧГ.

Непротиворечивость относительно  $\neg$  следует из Теоремы Cut в случае КЧГ точно так же, как в случае соответствующей генценовской секвенциальной логики предикатов.

Построив исчисление КЧГ, показываем, что в нем известные «парадоксы» отражаются выводимыми дедуктивными секвенциями, не влияющими на непротиворечивость КЧГ (еще до доказательства Теоремы Cut). Так, например, «парадокс» Б. Рассела 1902 г. представляется выводимыми в КЧГ секвенциями  $(\Rightarrow D^r)$  и  $(\Rightarrow (\neg D)^k)$ , где  $r$  и  $k$  – различные индикаторы (ср. в исчислении  $\mathbf{M}$  из [20] две выводимые секвенции  $(\Rightarrow D)$  и  $(\Rightarrow \neg D)$ , но «пустая» секвенция  $\otimes \Rightarrow \otimes$ , где  $\otimes$  – пустой набор слов, в  $\mathbf{M}$  по его построению не выводима – как принято говорить, исчисление  $\mathbf{M}$  (абсолютно) непротиворечиво (см. [20, 21]); имея две указанные выводимые секвенции и невыводимость пустой секвенции, легко доказываем недопустимость в  $\mathbf{M}$  из [20, 21] правила сечения).

Подчеркнём, что здесь, вопреки сложившейся практике, слово «парадокс» взято в кавычки, так как подразумеваемого под словом «парадокс» противоречия в данном случае нет.

Всё богатство КМ сохраняется в КЧГ, а выводимые в КЧГ типа «расселовских» секвенции  $\Rightarrow D^r$  и  $\Rightarrow (\neg D)^k$  (с различными индикаторами  $r$  и  $k$ ) можно рассматривать, следуя Хаскеллу Карри (1900–1981), как «монстры», не влияющие на непротиворечивость КЧГ.

Отметим особую роль исчисления  $\lambda$ -конверсии А. Чёрча [20, IV, п. 6] не только в теории КЧГ (например, 1-ый ярус КЧГ образует само исчисление  $\lambda$ -конверсии), но и в метатеории этой теории КЧГ. Так,

конвертируемость [20, IV, п. 7] заменяет равенство объектов (обов) [20, IV, п. 3], являясь его обобщением (см. постулаты  $\lambda$ -конверсии [20, IV, п. 6] как обобщающие известные свойства равенства). Она (конвертируемость) используется и при образовании исходных элементов (**М**-термов и **М**-формул) исчисления КЧГ.

Вообще, можно говорить о внутренней замкнутости теории КЧГ на себя: всё в теории КЧГ и её метатеории определяется средствами теории КЧГ с учетом двух ярусов КЧГ, отражающих две составляющие (компоненты) КМ.

Отношение конвертируемости, как известно, неразрешимо (это один из фундаментальных алгоритмических результатов школы Шейнфинкеля–Карри–Чёрча). В работе используется закон «исключенного третьего» с рассмотрением всех соответствующих случаев.

#### IV. Оснащенные **М**-формулы и индикаторы

В связи с предлагаемым доказательством Теоремы Cut к пункту 9 определения **М**-формул и **М**-термов [20, IV, с. 741–742] добавим новый подпункт (замыка по определению указанные в нем классы объектов (обов) относительно подстановки):

Если  $A$  – атомарная формула, формула, **М**-формула или **М**-терм,  $b$  есть **М**-терм,  $x$  – переменная, то об  $[b/x] A$  (результат подстановки  $b$  вместо  $x$  в  $A$ , естественно определяемой средствами  $\lambda$ -конверсии) считаем по определению соответственно атомарной формулой, формулой, **М**-формулой и **М**-термом.

Теперь определим *индикаторы* как (конечные, возможно, пустые) слова в однобуквенном алфавите | (буква «палочка»). Для *пустого индикатора* введем обозначение 0.

Если  $\alpha$  есть индикатор, то  $\alpha |$  называем *индикатором*.

Для индикатора | используем обозначение 1 («единица»). Соответственно, для индикатора | | используем обозначение 2 («два») и так далее. Так, к примеру, 5 есть индикатор | | | | |. Таким образом, все (натуральные) числа в работе воспринимаем как индикаторы.

По индукции доказываем, что конкатенация  $\alpha\beta$  индикаторов  $\alpha$  и  $\beta$  является индикатором (иногда вместо  $\alpha\beta$  пишем  $\alpha + \beta$ ).

Ниже *равенство слов* в алфавите | означает *совпадение* слов.

Считается, что индикатор  $s$  не равен индикатору  $r$ , если существует непустой индикатор  $t$  такой, что или  $s = r t$ , или  $r = s t$ .

Аналогично вводится неравенство  $r < s$  и другие знаки для индикаторов.

Определим далее новые слова, называемые *М-формулами с индикаторами* следующим образом:

Если  $A$  есть **М**-формула и  $r$  – индикатор, то конкатенация  $A^r$  называется *М-формулой с индикатором  $r$* , иногда пишем  $ind(A) = r$ .

Наконец, **М**-формулы с индикаторами будем называть *оснащенными М-формулами*.

Далее большими греческими буквами будем обозначать конечные (возможно пустые) упорядоченные наборы (списки) оснащенных **М**-формул. Слово  $(\Gamma \Rightarrow \Delta)$  называем дедуктивной секвенцией (ср. [20]), в отличие от [20] в дедуктивных секвенциях наборы состоят из оснащенных **М**-формул, пустой набор иногда обозначается  $\Theta$ .

## V. Построение теории Кантора – Чёрча – Генцена (КЧГ)

Как известно, в схеме аксиом – основной секвенции Генцена  $(A \Rightarrow A)$  – секвенциального формализма логики предикатов, не уменьшая общности, можно ограничиться только случаем атомарной формулы  $A$ .

Этот результат естественно распространить на построение исчисления КЧГ по исчислению **М** из [20] следующим образом.

Первый ярус исчисления **М** в КЧГ сохраняется без изменений; во 2-ом ярусе вместо правила подъема  $*$  из [20], не уменьшая общности, возьмем схему аксиом

(i):  $(A^0 \Rightarrow A^0)$ , где  $A$  – атомарная формула (но в **М**, с термами и **М**-термами), имеющая по определению индикатор 0, то есть  $A$  является оснащенной атомарной **М**-формулой с индикатором 0;

правила  $K^*$  –  $*\lambda$  из [20] сохраняются, только в них все дедуктивные секвенции состоят из оснащенных **М**-формул; при этом правила  $\Pi^*$  –  $*\lambda$  определяют индуктивно по построению выводов соответствующие индикаторы главных членов (что отражает в КЧГ известное определение рангов (по Генцену 1934 г.) формул логики предикатов), правила  $\Pi^*$  –  $*\lambda$  имеют следующий вид.

$$\Pi^*: \frac{\Gamma \Rightarrow \Theta, (av)^r}{\Gamma \Rightarrow \Theta, (\Pi a)^r}; \quad * \Pi: \frac{(ac)^r, \Gamma \Rightarrow \Theta}{(\Pi a)^r, \Gamma \Rightarrow \Theta};$$

$$P^*: \frac{a^r, \Gamma \Rightarrow \Theta, \varepsilon^s}{\Gamma \Rightarrow \Theta, (Pa\varepsilon)^{rs}}; \quad *P: \frac{(\Delta \Rightarrow \Lambda, a^r)(\varepsilon^s, \Gamma \Rightarrow \Theta)}{(Pa\varepsilon)^{rs}, \Delta, \Gamma \Rightarrow \Lambda, \Theta};$$

$$\lceil^*: \frac{a^r, \Gamma \Rightarrow \Theta}{\Gamma \Rightarrow \Theta, (\lceil a)^r}; \quad * \lceil: \frac{\Gamma \Rightarrow \Theta, a^r}{(\lceil a)^r, \Gamma \Rightarrow \Theta};$$

$$\lambda^*: \frac{(\Delta \Rightarrow \Lambda, a^r)(a \rightarrow \varepsilon)}{\Delta \Rightarrow \Lambda, \varepsilon^r}; \quad * \lambda: \frac{(\varepsilon \rightarrow a)(a^r, \Gamma \Rightarrow \Theta)}{\varepsilon^r, \Gamma \Rightarrow \Theta}$$

(в правилах  $r$  и  $s$  – произвольные индуктивно определенные (по гипотезе индукции) индикаторы, вводимые по посылкам правил; ограничения на  $u$  и  $c$ , указанные в [20], естественно сохраняются).

Построение теории КЧГ закончено.

Исчисление, в котором все **М**-формулы дедуктивной части являются оснащенными, будем называть *оснащенным исчислением*. Аналогично КЧГ можно строить после известных результатов Г. Генцена 1934 г. *оснащенные секвенциальные исчисления* логики предикатов 1-го порядка.

## VI. Теорема Cut и непротиворечивость теории Кантора – Чёрча – Генцена

### Формулировка Теоремы Cut

Если в КЧГ выводимы секвенции  $(\Delta \Rightarrow \Lambda, A^r)$  и  $(A^r, \Gamma \Rightarrow \Theta)$ , то в КЧГ существует вывод секвенции  $(\Delta, \Gamma \Rightarrow \Lambda, \Theta)$ , где  $A$  есть **М**-формула с индикатором  $r$ ;  $\Delta, \Gamma, \Lambda, \Theta$  суть наборы оснащенных **М**-формул.

Доказательство Теоремы непосредственно следует из следующей леммы о смешении в КЧГ, по формулировке и доказательству являющейся обобщением генценовской леммы о смешении в логике предикатов.

### Лемма о смешении в теории КЧГ

Если  $\aleph$  и  $\aleph$  – выводы секвенций  $E = (\Delta \Rightarrow A)$  и  $G = (\Gamma \Rightarrow \Theta)$  в исчислении КЧГ,  $A^r$  – оснащенная **М**-формула,  $r$  – индикатор, то в КЧГ можно построить вывод  $\aleph$  секвенции  $H = (\Delta, \Gamma_A \Rightarrow \Lambda_A, \Theta)$ , где выражение вида  $\Phi_A$  обозначает результат вычеркивания из набора  $\Phi$  всех оснащенных **М**-формул  $B^r$  таких, что  $B$  есть **М**-формула и  $B \leftrightarrow A$  ( $B$  конвертируется в  $A$ , то есть в исчислении  $\lambda$ -конверсии [20, IV, п. 6] выводимы  $\lambda$ -секвенции  $A \rightarrow B$  и  $B \rightarrow A$ ), здесь  $\Rightarrow$  выступает как символ равенства по определению,  $\Delta, \Lambda, \Gamma$  и  $\Theta$  – наборы оснащенных **М**-формул.

### Доказательство леммы о смешении в КЧГ

Доказательство Леммы проводим индукцией по кортежу  $(r, \beta, \alpha)$ , где  $\beta = \varphi [\aleph]$  и  $\alpha = \varphi [\aleph]$  (выражение вида  $\varphi [\varphi]$  (длина вывода  $\varphi$ ) означает число вхождений в вывод  $\varphi$  дедуктивных секвенций). Разбор случаев (включая базис и шаг индукции) осуществляем обычным образом по Г. Генцену 1934 г., например, аналогично тому, как это делается в статье «Теорема о сечении для  $\aleph$ -теорий в комбинаторно полных системах» [24] и других публикациях (список некоторых работ автора см., например, в [16] на с. 414–416).

Доказательство (с точностью до языка) отличается от известного логико-предикатного генценовского тем, что из наборов вычеркиваются не только все вхождения оснащенной **М**-формулы  $A^r$  (как по Генцену в случае формул логики предикатов), но и все конвертирующиеся в  $A^r$  оснащенные **М**-формулы  $B^r$  [ $B \leftrightarrow A$ , в частности,  $A \leftrightarrow A$ ].

Кроме того, новые (по отношению к генценовским) правила  $*\lambda^*$ , связывающие между собой два яруса КЧГ, не нарушают индикаторов вводимых **М**-формул (см. эти правила  $*\lambda^*$ ) и относятся к структурным генценовским правилам.

Кортеж, сопоставленный выводам  $\aleph$ ,  $\aleph$  и индикатору  $r$ , будем иногда обозначать через  $\mu [ \aleph, \aleph, r ]$ , где  $r = \text{ind}(A)$ .

В доказательстве используем (в случае последних правил  $\Pi^*$  в  $\aleph$  и  $\ast\Pi$  в  $\aleph$ ) лемму о подстановке **М**-термов вместо переменных в выводы исчисления КЧГ, сохраняющую все рассматриваемые параметры, формулируемую и доказываемую индукцией по длине данного вывода аналогично тому, как это делается в [16] и др. публикациях с 1970 г.

Кортежи  $(e, f, \gamma)$ , где  $e, f, \gamma$  – индикаторы, считаем упорядоченными лексикографически, то есть  $(e_1, f_1, \gamma_1) < (e_2, f_2, \gamma_2)$  тогда и только тогда, когда выполняется одно из условий: (1)  $e_1 < e_2$ ; (2)  $e_1 = e_2$  и  $f_1 < f_2$ ; (3)  $e_1 = e_2, f_1 = f_2$  и  $\gamma_1 < \gamma_2$ .

Используя структурные правила  $\ast K^*$ ,  $\ast W^*$ ,  $\ast C^*$ ,  $\ast \lambda^*$ , вывод  $\exists$  можно построить: (1) как продолжение  $\aleph$  при  $A \overset{\uparrow}{\leftrightarrow} \Lambda$  (выражение  $c \overset{\uparrow}{\leftrightarrow} \Phi$  означает, что в наборе  $\Phi$  нет оба  $h$  такого, что  $c \leftrightarrow h$ ) или при  $A \leftrightarrow \Theta$ ; (2) как продолжение  $\aleph$  при  $A \overset{\uparrow}{\leftrightarrow} \Gamma$  или при  $A \leftrightarrow \Delta$ .

В дальнейшем предполагаем, что  $A \overset{\uparrow}{\leftrightarrow} \Delta \Theta$ ;  $A \leftrightarrow \Lambda$ ;  $A \leftrightarrow \Gamma$ . Пусть  $\gamma = (r, \beta, \alpha)$ .

Гипотеза индукции: допустим, что по любым выводам  $\aleph$  и  $\aleph$  и индикатору  $\tau$  при  $\mu [ \aleph, \aleph, \tau ] < \gamma$  можно указать вывод  $\exists$  секвенции  $H$ .

Рассмотрим выводы  $\aleph$  и  $\aleph$  и индикатор  $r \geq 0$  такие, что  $\mu [ \aleph, \aleph, r ] = \gamma$ . Здесь  $\varphi [ \aleph ] > 1$  и  $\varphi [ \aleph ] > 1$ . Доказательство существования вывода  $\exists$  проведем путем разбора случаев построения заключительных секвенций  $E$  и  $G$  выводов  $\aleph$  и  $\aleph$ .

Если  $E$  или  $G$  получена по одному из операциональных правил  $\ast P^*$ ,  $\ast P^*$ ,  $\ast |^*$ , то индикатор главного его члена  $Pfg$ ,  $|g$  или  $\Pi f$  как слова (оба), независимо от выводов, убывает при переходе к указанным в посылках оснащенным **М**-формулам  $f$  и  $g$ ,  $g$  или  $fc$ .

Заметим, что случаи  $\varphi [ \aleph ] = 1$  или  $\varphi [ \aleph ] = 1$  уже рассмотрены: вывод  $\exists$  получен как продолжение одного из данных выводов  $\aleph$  или  $\aleph$ .

Разбор различных случаев далее в доказательстве леммы о смешении осуществляем точно так, как это делается в литературе, начиная с публикаций Г. Генцена (1934 г.) для логики предикатов 1-го порядка и автора (с 1973 г.) в классах выводов таких, как **М** или КЧГ, бестиповых двухъярусных секвенциальных исчислений. Остановимся на «новых» правилах  $\ast \lambda^*$ .

Рассмотрим случаи применения правил  $\ast \lambda^*$ , являющихся последними в  $\aleph$  или  $\aleph$ . При вычеркивании  $B^r$ , когда, например, применяется правило  $\lambda^*$  со второй посылкой  $C \rightarrow D$ , где  $C \leftrightarrow A$ , по гипотезе индукции, обращаясь к выводу первой (дедуктивной) посылки правила  $\lambda^*$ , вычеркиваются и все оснащенные **М**-формулы  $D^r$ , поскольку  $D \leftrightarrow A$ , т.е. сразу строится искомым вывод  $\exists$  секвенции  $H$ . Так, можно сказать, «согласованы» правила  $\ast \lambda^*$  и лемма о смешении.

Лемма о смешении в КЧГ доказана.

Доказательство непротиворечивости КЧГ относительно отрицания

Предположим, что КЧГ противоречива. Тогда по определению в КЧГ выводимы, в частности, секвенции  $\Rightarrow B^S$  и  $\Rightarrow (\neg B)^S$  для некоторой оснащенной **М**-формулы  $B^S$ . Из секвенции  $\Rightarrow B^S$  по правилу  $* \neg$  следует секвенция  $(\neg B)^S \Rightarrow \otimes$ , где  $\otimes$  – пустой набор. Применив к выводам секвенций  $\Rightarrow (\neg B)^S$  и  $(\neg B)^S \Rightarrow \otimes$  Теорему Cut, получим в КЧГ вывод «пустой» секвенции  $\otimes \Rightarrow \otimes$ , что невозможно (см. [20, 21]) – поскольку постулаты исчисления КЧГ не кончаются «пустыми» секвенциями. Поэтому теория КЧГ непротиворечива относительно отрицания, что и требовалось доказать.

Теперь читателю рекомендуется прочесть еще раз [20, 21], заменяя исчисление **М** на КЧГ и понимая под непротиворечивостью и непротиворечивость относительно отрицания  $\neg$ .

Обратим внимание на работу «Solution of Hilbert Central Problem Following Kolmogorov» [25], являющуюся переводом [21] на английский язык. В [25] название [21] получило естественное уточнение.

Центральная проблема Гильберта впервые решена автором по Колмогорову; при этом в [20, 21, 25] непротиворечивость теорий **М** и, следовательно, КЧГ понимается только абсолютно, в данной работе доказывается, что КЧГ непротиворечива и относительно отрицания  $\neg$ .

В доказательстве непротиворечивости КЧГ относительно отрицания используется абсолютная непротиворечивость КЧГ, то есть если теория КЧГ непротиворечива абсолютно, то она непротиворечива и относительно отрицания. Из абсолютной непротиворечивости КЧГ следует её непротиворечивость и относительно отрицания. А абсолютная непротиворечивость вытекает непосредственно из построения КЧГ точно так же, как в [20, 21] это делается для теории **М**.

Поэтому для теории КЧГ, как у Г. Генцена для логики (предикатов 1-го порядка), непротиворечивость относительно отрицания является следствием построения соответствующего исчисления КЧГ или (как у Генцена) логики предикатов.

При этом КЧГ выступает как логика с генценовскими постулатами ( $i$ ),  $K* - *$ , расширенная алгоритмическим исчислением  $\lambda$ -конверсии А. Чёрча [20, IV, п. 6], представляющим теоретико-множественное свертывание Г. Кантора; связь между свертыванием и логикой осуществляется правилами  $*\lambda*$ . В таком смысле можно рассматривать теорию КЧГ как логический вариант (с аналогом теоремы К. Гёделя 1930 г. о полноте) формализации КТМ (канторовской теории множеств). Причем (с точностью до языка) «парадокс» КТМ, формульно выражаемый  $A$  и  $\neg A$ , представляется в КЧГ выводимыми дедуктивными секвенциями  $\Rightarrow A^k$  и  $\Rightarrow (\neg A)^l$  с различными индикаторами  $k$  и  $l$ .

Итак, два яруса исчисления КЧГ, как аналоги двух составляющих КМ, позволяют в КЧГ применением двух принципов КТМ, представ-

ленных в КЧГ логическими и алгоритмическими постулатами, получить в соответствии с программой А.Н. Колмогорова доказуемо полным и непротиворечивым образом все выводы КМ без ограничений естественно на основе колмогоровского пакета законов рассуждений, заданного указываемыми по постулатам исчисления КЧГ всеми свойствами выводимости при замене символа  $\Rightarrow$  на знак выводимости  $\vdash$ , или секвенциально в исчислении КЧГ.

Первый алгоритмический ярус КЧГ с его неразрешимостью используется не только при построении 2-го, логического, яруса КЧГ, но и при задании исходных элементов (**М**-термов и **М**-формул) исчисления КЧГ, а также в других случаях.

### **Заключение**

#### **Реализация программы А.Н. Колмогорова**

В статьях [20, 21, 25] и настоящей работе показано, что КМ, имея две компоненты и базируясь на двух принципах КТМ, доказуемо полностью и доказуемо непротиворечиво представляется двухъярусным секвенциальным исчислением (теорией) Кантора – Чёрча – Генцена (КЧГ).

Таким образом, программа А.Н. Колмогорова по основаниям математики реализована.

Формализацию областей математики, основания которых образует КТМ или ее части, можно осуществить в КЧГ аналогично тому, как, например, Уильям С. Хэтчер [26] проводит формализацию в известной аксиоматической системе Г. Фреге.

При этом соответствующие определения и конструкции в КЧГ можно ввести, следуя за определениями и построениями В.К. Захарова и А.В. Михалева (см., например, [27]) общей концепции произвольной математической системы в рамках теории множеств NBG Неймана – Бернаиса–Гёделя.

Подробности о неразрешимых бестиповых алгоритмических исчислениях чистой комбинаторной логики Шейнфинкеля–Карри и  $\lambda$ -конверсии Чёрча (не образующих теории 1-го порядка, но каждое из которых, обладающее неограниченной комбинаторной полнотой, можно положить в основу двухъярусной секвенциальной теории типа КЧГ) хорошо изложены, например, в монографиях Барендрегта 1985 г. и Вольфенгагена 2004 г. [28, 29] (см. также библиографию в [28, 29], работу Резуса 1982 г. [30], труд Хиндли и Кардоне 2005 г. по истории  $\lambda$ -исчисления и комбинаторной логики [31]). Указанная неограниченная комбинаторная полнота ( $\lambda$ -полнота) является принятой в КЧГ формализацией неограниченного канторовского теоретико-множественного свертывания.



С логическими секвенциальными конструкциями, используемыми в КЧГ, читатель может предварительно познакомиться, например, по известным трудам Генцена, Клини, Ершова и Палютина [32–34].

Когда данная работа была уже подготовлена к печати, автор ознакомился со статьей В.Я. Яцука [35], которую можно рассматривать как введение в предмет; процитирую частично резюме из [35]:

«АКЛ (аппликативная компьютерная логика) представляет собой систему прикладного исчисления секвенций с неограниченным принципом свертывания в форме комбинаторной полноты и состоит из двух частей, первая из которых называется теорией выводимости и является инвариантной к рассматриваемым областям знания. Вторая часть отражает конкретное знание о заданной предметной области. Теория выводимости реализуется в виде двух ступеней: алгоритмической и структурной. Логические связки (*с некоторыми ограничениями* – А.С.К.) определяются в рамках заданного комбинаторно-полного базиса и могут быть выведены автоматически» ([35], с. 29).

В статье [35] имеется интересное неформальное введение (с. 29–30).

Настоящая работа выполнена на кафедре теории вероятностей в кабинете истории и методологии математики и механики механико-математического факультета МГУ. Впервые основные её положения были доложены 22 апреля 2002 г. на научной конференции «Ломоносовские чтения МГУ».

Автор искренне благодарен всем, кто явно или неявно содействует реализации программы А.Н. Колмогорова по основаниям математики, особенно Юрию Васильевичу Прохорову, представившему в ДАН статьи [7, 20, 21], Анатолию Тимофеевичу Фоменко, опубликовавшему мою работу [6], Сергею Ивановичу Адяну, непонимание которым статьи [7] весьма помогло автору выделить более явно теоретико-множественный колмогоровский путь в основаниях современной науки, а также всем сотрудникам механико-математического и философского факультетов МГУ и других научных организаций России и зарубежья, обсуждавшим мои результаты.

## ЛИТЕРАТУРА

1. Кузичев А.С. Паранепротиворечивость интеллектуальных систем компьютерных логик. //Сб. «Технология проектирования интеллектуальных систем», М., МДНТП, 1989, с. 132–139.
2. Кузичев А.С. О роли теорем Гёделя о неполноте в основаниях наук. //»Философия и будущее цивилизации. Тезисы докладов и выступлений IV Российского философского конгресса», в 5 тт., Том 1, М., Современные тетради, 2005, с. 726–727.
3. Вейль Г. Математика и логика (1946). //Вейль Г. Избранные труды. Математика. Теоретическая физика. М., Наука, 1984, с. 328–340.
4. Рид К. Гильберт. С приложением обзора Германа Вейля математических трудов Гильберта (1970). М., Наука, 1977, 368 с.
5. Крайзель Г. Биография Курта Гёделя. Успехи математических наук (УМН), 1988, март – апрель, т.43, вып.2 (260), с.175–216; май – июнь, т.43, вып.3 (261), с.203–238.
6. Кузичев А.С. О негёделевской перестройке арифметики и других аксиоматических теорий первого порядка по Колмогорову. Доказательство их непротиворечивости. М., Изд-во механико-математического ф-та МГУ, 2004, 36 с.
7. Кузичев А.С. Колмогоровская редукция и непротиворечивость. ДАН, 1999, т. 367, № 2, с. 161–163.
8. Кузичев А.С. Арифметические теории, строящиеся на основе  $\lambda$ -конверсии. Доклады Академии наук (ДАН), 1981, т.261, № 4, с.792–796.
9. Кузичев А.С. Арифметически непротиворечивые  $\lambda$ -теории. ДАН, 1982, т. 262, № 4, с.795–799.
10. Кузичев А.С. Аксиоматические теории в комбинаторно полных системах. ДАН, 1982, т. 264, № 3, с.538–542.
11. Кузичев А.С. О представлении теорий первого порядка в бестиповых комбинаторно полных системах. ДАН, 1982, т.266, № 1, с.23–27.
12. Кузичев А.С. Арифметически непротиворечивые  $\lambda$ -теории бестиповой логики. ДАН, 1983, т.268, № 2, с.288–292.
13. Кузичев А.С. Теория множеств в бестиповых комбинаторно полных системах. «Вестник Московского университета», серия «матем., механ.», 1983, № 3, с.36–42.
14. Кузичев А.С. Непротиворечивость системы NF Куайна. ДАН, 1983, т. 270, № 3, с.537–541.
15. Кузичев А.С. Арифметическая полнота бестиповой логики. ДАН, 1983, т. 270, № 6, с.1323–1327.
16. Кузичев А.С. Об одной арифметически непротиворечивой  $\lambda$ -теории. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, 1983, Bd. 29, H. 5, s. 385–416.
17. Кузичев А.С. Теорема о непротиворечивости системы ZF Цермело–Френкеля. ДАН, 1983, т.273, № 5, с.1053–1057.
18. Колмогоров А.Н. О принципе Tertium non datur // Математический сборник, 1925, т. 32, № 4, с. 646–667.
19. Колмогоров А.Н. Математика и механика. М., Наука, 1985, 470 с.

20. *Кузичев А.С.* Вариант формализации канторовской теории множеств // ДАН, 1999, т. 369, № 6, с. 740–742.
21. *Кузичев А.С.* Решение проблемы Гильберта по Колмогорову // ДАН, 2000, т. 371, № 3, с. 303–306.
22. *Рыбников К.А.* История математики. М., Изд-во МГУ, 1994, 494 с.
23. *Leibniz G.W.* Dissertatio de arte combinatoria (1666). Die philosophische Schriften. Berlin, 1880, Bd. 4, s. 27–102.
24. *Кузичев А.С.* Теорема о сечении для  $\mathfrak{R}$ -теорий в комбинаторно полных системах // Вестник МГУ, Сер.1 Математика, механика, 1977, № 1, с. 62–67.
25. *Kuzichev A.S.* Solution of Hilbert Central Problem Following Kolmogorov // *Doklady Mathematics*. Vol. 61, № 2, 2000, pp. 212–215.
26. *Hatcher William S.* The Logical Foundations of Mathematics. Oxford., Pergamon Press. 1982, X+320 pp.
27. *Zakharov V.K., Mikhalev A.V.* Mathematical systems // Proceedings of the International Algebraic Conference. Berlin: Walter De Gruyter Publishers. 2000, pp. 370–383.
28. *Барендрегт Х.* Лямбда-исчисление. Его синтаксис и семантика. М., Мир, 1985, 606 с.
29. *Вольфенгаген В.Э.* Методы и средства вычислений с объектами. Аппликативные вычислительные системы. М., Центр ЮриИнфоР, 2004, XVI+789 с.
30. *Rezus A.* A Bibliography of Lambda-Calculi, Combinatory Logics and Related Topics. Amsterdam, 1982, 86 p.
31. *Hindley J.R., Cardone F.* History of  $\lambda$ -Calculus and Combinatory Logic. Preprint. Abertawe, 2005, 96 p.
32. *Генцен Г.* Исследования логических выводов. В сб. «Математическая теория логического вывода», М., Наука, 1967, с. 9–76.
33. *Клини С.К.* Введение в метаматематику. М., Изд-во иностранной литературы, 1957, 526 с.
34. *Ершов Ю.Л., Палютин Е.А.* Математическая логика. М., Наука, 1987, 336 с.
35. *Яцук В.Я.* Аппликативная компьютерная логика: неограниченные возможности и возможные ограничения // С-Пб, Вестник Петровской Академии, 2005, № 1, с. 29–35.

---

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЕГО ПРИЛОЖЕНИЯ В ЭКОНОМИКО- МАТЕМАТИЧЕСКИХ МОДЕЛЯХ\*

*А.Р. Бахтизин*

## **Введение**

Мотивы поведения человека в социально-экономической среде занимали умы величайших философов и ученых на протяжении всей истории человечества.

Согласно классической теории поведения потребителя, принимаемые им решения, исходят из соображений полной рациональности. Т.е. предполагается, что человек всегда выбирает наилучшее действие с целью максимизации полезности от приобретаемых им благ или его поведение в плане поиска работы оптимально с точки зрения будущего дохода. При этом также предполагается, что человек знает все возможности выбора и осведомлен о последствиях каждой альтернативы.

Нобелевский лауреат Герберт Саймон (Саймон (2000)) подверг резкой критике данный подход с его «абсурдно всеведущей рациональностью» и показал, что такая модель поведения человека далека от реальности.

Основной аргумент Г. Саймона заключался в том, что в основе поведения человека как работника лежит не только желание получить максимальный денежный доход, но еще и моральное удовлетворение, связанное с удовлетворением его социальных потребностей или с реализацией творческих возможностей.

Несмотря на то, что такое понятие как «моральное удовлетворение» не рассматривается в классической экономической теории, в психологии оно имеет первостепенное значение. В большинстве психологических теорий, основным мотивом человека, побуждающим его к действиям, является неудовлетворенность стремлений. Причем уровень стремлений непостоянен, а зависит от жизненного опыта, в связи с чем, достижение удовлетворенности является итерационным и непрекращающимся процессом.

---

\* Работа выполнена при поддержке Российского гуманитарного научного фонда (проект № 05-02-02039а), Фонда «Научный потенциал», Фонда содействия отечественной науке, Гранта Президента РФ «Разработка и апробация вычислимой модели общего экономического равновесия с применением технологий искусственного интеллекта» (проект № МК-1085.2005.6).

По мнению Г. Саймона в человеческом поведении много иррационального, а границы рациональности для каждого индивидуума не статичны и могут изменяться в зависимости от окружения. Например если ситуация хорошо знакома, а реакция внешней среды предсказуема, то принимаемое решение может быть оптимальным, однако в случае изменения какого-нибудь параметра внешней среды, с одной стороны могут поменяться границы рациональности, а с другой – не факт, что принимаемое решение будет оптимальным.

Традиционное принятие решений на основе рационального поведения было раскритиковано Г. Саймоном в разработанной им концепции ограниченной рациональности, согласно которой «способности человеческого мышления формулировать и решать комплексные проблемы в очень малой степени сравнимы с масштабом проблем, решение которых требуется для объективного рационального поведения в реальном мире или даже для приемлемой аппроксимации к такой объективной рациональности» (Пью, Хиксон (1999)).

Также необходимо отметить, что множество рассматриваемых человеком альтернатив гораздо меньше их реального числа, и поэтому невозможно точно предсказать последствия любой из них. Кроме того, цели, достигаемые человеком, являются неоднозначными и поэтому их нельзя измерить количественно.

Разработанная Г. Саймоном теория ограниченной рациональности (*bounded rationality*), имеет множество последователей. К наиболее известным работам в этой области экономической теории можно отнести (Gigerenzer, Selten (2002)), (Rubinstein (1997)) и (Sargent (1994)).

В нашей работе мы решили отказаться от традиционного моделирования поведения человека посредством максимизации функции полезности в пользу нестандартного моделирования ограниченной рациональности, посредством применения технологий искусственного интеллекта.

## **1. Моделирование поведения человека в социально-экономической среде**

Вопросами моделирования поведения человека занимаются многие ученые и практики, а само направление уже давно сформировалось в отдельную отрасль информатики – искусственный интеллект (ИИ). Это направление включает в себя разработку систем, обладающих возможностями, связываемыми с естественным интеллектом.

Необходимо отметить, что в проводимых исследованиях «экономическая составляющая» искусственного интеллекта не является первостепенной, а во главу угла ставится вопрос о создании автономных сущностей, способных к совершенствованию и, при необходимости, к воспроизведению себе подобных.

Рассмотрим современное состояние дел в области создания искусственного интеллекта и определим в этом ключе свою позицию при выборе инструмента для разработки нашей модели.

Термин искусственный интеллект (*Artificial Intelligence*) был предложен А. Тьюрингом, который в 1950 году сформулировал свой знаменитый тест Тьюринга, определяющий способность машины мыслить (Turing (1950)). Однако сейчас под ИИ понимают не робота, способного пройти тест Тьюринга, а как уже было отмечено, целое направление информатики.

Перечислим основные направления развития ИИ, получившие общественное признание и уже зарекомендовавшие себя:

**1. Нейронные сети** – наиболее популярный аппарат в области ИИ, который может быть применим в любой ситуации, где есть связь между входными и выходными переменными. Математически доказано (Круглов, Борисов (2002)), что любая непрерывная функция может быть равномерно приближена функциями, вычисляемыми нейронными сетями, если функция активации нейрона дважды непрерывно дифференцируема.

Идея нейронных сетей появилась в процессе исследований в области искусственного интеллекта и основывается на биологической модели нервных систем (Patterson (1996)).

На представленном ниже рисунке 1а представлена модель биологического нейрона. Он состоит из *ядра* (тела нейрона), двух видов нервных волокон – *аксонов*, передающих нервные импульсы и принимающих их *дендритов*, а также *синапсов*, влияющих на силу сигнала. Нейрон получает импульсы от аксонов других нейронов через дендриты и при этом сила импульса меняется в зависимости от типа синапса. Если суммарный импульс превышает некоторый порог, то нейрон возбуждается, формирует свой собственный импульс и передает его дальше по аксону. Мозг человека содержит около  $10^{11}$  нейронов, каждый из которых связан с  $10^3 - 10^4$  другими нейронами.

Когда человек получает новую информацию, то в отдельном участке мозга веса некоторых синапсов меняются и, соответственно, меняется поведение нейронов мозга. Таким образом, при решении какой-нибудь конкретной задачи, в мозгу задействуются соответствующие нейроны, возбуждаемые синапсами, подстроенными под поиск решения. Чем чаще человек сталкивается с задачами одного класса, тем более устойчивыми становятся веса синапсов и меньше времени требуется для поиска решения.

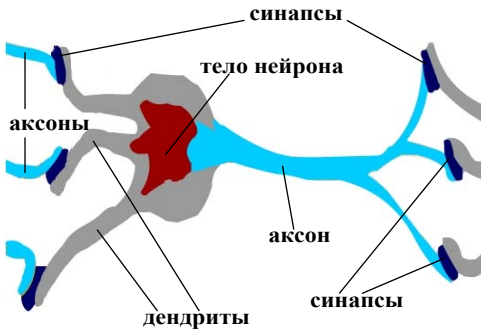


Рисунок 1а. Модель биологического нейрона

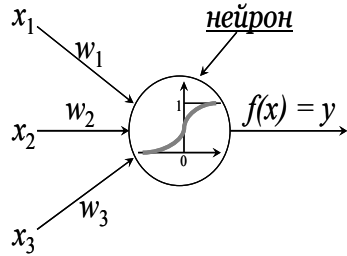


Рисунок 1б. Модель искусственного нейрона

Математическая формализация биологического нейрона была предложена в середине 50-х годов прошлого века. Модель искусственного нейрона задается в виде скалярной функции векторного аргумента.

На рисунке 1б представлена модель нейрона с тремя входами (дендритами), синапсы которых имеют веса  $w_1$ ,  $w_2$  и  $w_3$ . Пусть поступающие к синапсам импульсы имеют силу  $x_1$ ,  $x_2$  и  $x_3$ . Тогда к нейрону поступает суммарный импульс с силой  $x = x_1 * w_1 + x_2 * w_2 + x_3 * w_3$ .

Выше уже говорилось, что если суммарный импульс превышает некоторый порог, то тогда нейрон преобразует его с помощью некоторой передаточной (активационной) функции  $y = f(x)$ . Хотя существуют и другие функции активации нейронов, на практике, как правило, чаще всего используется нелинейная логистическая функция или сигмоид вида:

$$f(x) = \frac{1}{1 + e^{-x}}$$

При разработке нейронных сетей для решаемой нами задачи мы использовали именно сигмоид, поскольку эта функция обладает свойством усиливать слабые сигналы и тем самым предотвращает насыщение от больших сигналов, которые соответствуют тем областям аргумента, где сигмоид имеет пологий наклон.

Выходное значение сигмоида лежит в интервале (0;1), а область чувствительности для входов чуть шире интервала (-1;1). Поскольку при решении реальных задач приходится иметь дело с различными

значениями, то для подачи данных на вход сети используются алгоритмы шкалирования данных (к примеру, алгоритм минимакса). Такие алгоритмы, используемые на этапе предварительной обработки данных, называются *пре-процессированием*, а на этапе заключительной обработки – *пост-процессированием* (Bishop (1995)).

Существует множество различных алгоритмов обучения нейронных сетей, однако наиболее часто используется алгоритм обратного распространения ошибки (*error back propagation*), входящий в группу алгоритмов «обучения с учителем». Этот итеративный градиентный алгоритм обучения используется с целью минимизации среднеквадратичного отклонения текущих от требуемых выходов многослойных сетей с последовательными связями. При этом сигналы ошибки от выходов нейронной сети распространяются к ее входам, т.е. в направлении, обратном прямому распространению сигналов, с целью соответствующей подстройки весов синапсов. Таким образом, ошибки нейронной сети служат для оценки производных функции ошибок по отношению к регулируемым весам.

Подробное описание алгоритма обратного распространения ошибки приводится в книге «Искусственные нейронные сети» (Круглов, Борисов (2002)).

**2. Эволюционные вычисления**, т.е. автономное и адаптивное поведение компьютерных приложений. Прежде всего, эта отрасль затрагивает аспекты самовосстановления и самоконфигурирования сложных систем, состоящих из одновременно функционирующих модулей. Помимо этого, к эволюционным вычислениям относятся автономные агенты, несущие в себе функции электронного секретаря, ассистента, отбирающего нужные сведения в Интернете и т.д.

Примером такого агента может быть свободно распространяемая программа, разработанная фирмой GATOR ([www.gator.com](http://www.gator.com)), хотя мнения ее пользователей неоднозначно. Многие фирмы называют GATOR «сетевым паразитом», поскольку ее агент, следя за предпочтениями пользователя Интернет, навязывает ему посещение сайтов со смежной тематикой и тем самым наносит финансовый ущерб, поскольку уводят пользователя к конкуренту. Однако быстрый и хаотичный рост Web-пространства, несомненно, приведет к появлению интеллектуальных автономных агентов нового поколения, способных к самообучению, взаимодействию с себе подобными и проявляющим самостоятельность при принятии решений. Однако отношение общества к таким «работникам» пока трудно предсказать.

**3. Экспертные системы (ЭС)**, создание которых традиционно считается классическим занятием специалиста по ИИ. Под ЭС понимают систему (чаще всего воплощенную в виде компьютерной программы), основанную на знаниях экспертов, помогающую специалисту при принятии решений. В каждой области человеческой деятель-



ности существуют знания, которые трудно формализовать математическими формулами, что и обусловило появление ЭС, а сама задача получения знаний выделилась в новое направление инженерии знаний (четкая формулировка знаний эксперта-специалиста и внесение их в базу знаний компьютера).

Для ответа на поставленные пользователем вопросы требуется система извлечения ответов, называемая машиной логического вывода или интерпретатором. В качестве такой системы, компанией PDC был предложен особый инструмент в программировании приложений ИИ – декларативный язык программирования Visual Prolog (или язык ПРОграммирования ЛОГики), работа которого основана на исчислении предикатов первого порядка. Полное описание языка и обзор по экспертным системам можно найти в книге (Адаменко, Кучуков (2003)). Как уже было сказано, Visual Prolog язык декларативный, что отличает его от традиционных процедурных языков типа С и Basic. В процедурных языках программист должен четко прописать пошаговый алгоритм решения задачи, а в декларативном языке нужно предоставить только описание задачи и основные правила для ее решения. Урезанную по возможностям версию языка Visual Prolog (Visual Prolog 6.1 Personal Edition) можно бесплатно скачать с сайта [www.visual-prolog.com](http://www.visual-prolog.com).

Из наиболее известных примеров ЭС можно отметить экспертно-справочную систему Сус (en-Cyc-lopedia), разработанную компанией Cycorp ([www.cyc.com](http://www.cyc.com)). Эта система содержит более 1 млн. утверждений, охватывает все области знаний и способна делать логические выводы. Кстати, широко известный «скрепыш» – помощник офисных приложений компании Microsoft (Office Assistant), является ни чем иным как самоорганизующейся экспертной системой, в задачу которой входит конфигурирование справочной системы Microsoft Office в соответствии с часто затрагиваемыми темами. Для этих целей «скрепыш» постоянно отслеживает поведение пользователя в рамках семейства офисных программ.

**4. Нечеткая логика (fuzzy logic)** – направление, предложенное в 1965 году профессором Калифорнийского университета Лофти Заде, сочетающее в себе подходы математической логики и теории вероятностей. В нечеткой логике, в отличие от обычной, высказывания бывают не только истинными или ложными, что позволяет учитывать неопределенности при моделировании ИИ. Для решения практических задач в этой области, качественные переменные описываются некоторой функцией распределения, после чего они уже используются как точные (например, с помощью нечеткой логики можно более точно определить понятие величины дохода человека – «маленький», «средний», «большой», «очень большой»). Согласно знаменитой теореме FAT (Fuzzy Approximation Theorem), доказанной Б. Кос-

ко (Kosko (1992)), «любая математическая система может быть аппроксимирована системой, основанной на нечеткой логике.

Основные исследования в области нечеткой логики проводятся в США и Японии, а в числе результатов этих исследований – многочисленные микрочипы, используемые как в бытовой технике (стиральные машины, СВЧ-печи), так и в моторных отсеках автомобилей, в поездах метрополитена и т.д. Свое место нечеткие системы управления нашли в военной технике (например, при проектировании «умных» ракет).

Классическим примером использования нечеткой логики в системах управления стал эксперимент по управлению грузовиком при въезде в узкий гараж. Не делая никаких лишних движений, грузовик каждый раз без труда находил оптимальный путь для заезда. Система, управляющая грузовиком и заменяющая водителя, включала в себя всего лишь небольшой набор нечетких правил типа «если капот направлен влево, то возьми правее» и т.д. (всего 35 правил и 12 нечетких условий). Разработка такой системы посредством обычного математического аппарата потребовала бы на порядок больших усилий, да и вообще могла бы давать сбой в процессе эксплуатации.

Изложение основ нечеткой логики заняло бы слишком много места, поэтому для более детального изучения этого направления мы можем порекомендовать бестселлер Барта Коско (Kosko (1992)), в котором приводится фундаментальное изложение теории нечетких множеств.

Среди русскоязычных источников представляют интерес книги (Левнер, Птускин, Фридман (1998)), (Пивкин, Бакулин, Кореньков (1998)), а также научно-популярная статья А. Масаловича (Масалович (1995)). Кроме того, большая подборка материалов в этой области выложена по адресу:

[http://dir.yahoo.com/Science/Computer\\_Science/Artificial\\_Intelligence/Fuzzy\\_Logic](http://dir.yahoo.com/Science/Computer_Science/Artificial_Intelligence/Fuzzy_Logic)

**5. Генетические алгоритмы (ГА)** это последовательность управляющих действий и операций, моделирующих эволюционные процессы на основе аналогов механизмов генетического наследования и естественного отбора, почерпнутых из биологии (Круглов, Борисов (2002)). Из определения уже ясно, что это направление берет свое начало из теории эволюции, согласно которой каждый биологический вид непрерывно развивается, чтобы наилучшим образом приспособиться к окружающей среде. Путем естественного отбора природа решает задачу оптимизации, в результате чего выживают только более приспособленные особи. Используя биологическую терминологию можно представить задачу поиска экстремума функции многих переменных следующим образом:

*Хромосома* – последовательность нулей и единиц, каждая позиция которой называется *геном*. В задаче поиска экстремума это *аргумент* нашей функции. Для использования значений аргумента в ГА, сначала необходимо перевести их в двоичный формат.

*Особь* – набор хромосом, представляющих генетический код. В нашей задаче это *значение целевой функции*.

В биологии каждая особь состоит из клеток, содержащих в себе набор молекул ДНК. При размножении происходит взаимодействие ДНК особей, в результате чего образуется ДНК потомка. Существует три основных способа взаимодействия:

*Кроссовер* (или *скрещивание*) – операция, при которой две хромосомы обмениваются частями.

*Мутация* – случайное изменение хромосомы (изменение состояния одного из генов на противоположное).

*Инверсия* – изменение порядка генов в хромосоме (аргументе функции) путем циклической перестановки.

В ГА чаще используются первые два способа взаимодействия хромосом.

Последовательность действий ГА выглядит следующим образом:

Шаг 1. создание начальной популяции, т.е. присвоение аргументам целевой функции случайных значений;

Шаг 2. вычисление значения целевой функции, в соответствии со значениями аргументов, заданных на первом шаге;

Шаг 3. преобразование аргументов функции в двоичный формат;

Шаг 4. применение операций скрещивания и мутации для всего набора хромосом (популяции). На этом шаге количество особей в популяции возрастает;

Шаг 5. *селекция* популяции – формирование новой популяции из старой, путем отбора новых хромосом и удаления старых, после чего старая популяция вымирает;

Шаг 6. обратное преобразование аргументов функции в десятичный формат;

Шаг 7. расчет целевой функции и проверка ее значений со значениями, полученными в предыдущем цикле вычислений. Если оно стало больше (в случае поиска максимума функции), то такой набор хромосом запоминается;

Шаг 8. если значение целевой функции раз от разу не может превысить экстремального значения, полученного на одном из шагов итераций, то экстремальное значение найдено, а в противном случае – переход к шагу 3.

Сравнивая ГА с другими способами решения задач оптимизации – переборным и локально-градиентным, можно отметить, что ГА вобрала в себя лучшие стороны обоих подходов. С одной стороны этот

алгоритм гарантирует нахождение глобального максимума или минимума (в отличие от градиентного метода), а с другой – экстремум находится за разумное время (в отличие от метода перебора).

Несомненно, что описанные выше технологии привлекают внимание военных (к примеру, военно-научное агентство DAPRA является крупнейшим финансистом проектов в области ИИ и робототехники). При разработке современного оружия сейчас активно используются системы ИИ (в основном нейронные технологии и нечеткие экспертные системы). Например, реализация режима автономного полета на небольшой высоте и в плохих условиях без использования заранее подготовленной компьютерной базы рельефа требует применения высокоэффективных механизмов синхронизации движения с данными, получаемыми от систем навигации, видеокамер, радаров и других датчиков. Использование ИИ позволяет с помощью относительно малых ресурсов получать достаточно точные результаты, для вычисления которых классическими методами численной математики понадобилось бы использование суперкомпьютеров (Круглов, Борисов (2002)).

Более подробный обзор систем ИИ можно найти в работах С. Бобровского (Бобровский (2001)), а одна из самых больших подборок материалов по этой теме находится здесь:

[http://dir.yahoo.com/science/computer\\_science/artificial\\_intelligence](http://dir.yahoo.com/science/computer_science/artificial_intelligence).

И, наконец, отметим, что область ИИ привлекает большое количество исследователей, и на данный момент существуют различные общества, издающие периодические журналы и проводящие многочисленные конференции по этой теме. Наиболее известными из них являются:

1. The Journal of Artificial Societies and Social Simulation (JASSS, <http://jasss.soc.surrey.ac.uk>);
2. American Association for Artificial Intelligence (AAAI, [www.aaai.org](http://www.aaai.org));
3. International Society of Artificial Life (ISAL, [www.alife.org](http://www.alife.org));
4. IEEE Neural Networks Society (IEEE, [www.ieee-nns.org](http://www.ieee-nns.org)).

Нельзя не упомянуть также про крупнейшую лабораторию, занимающуюся исследованиями в области ИИ – MIT Artificial Intelligence Laboratory ([www.ai.mit.edu](http://www.ai.mit.edu)).

Из перечисленных в этом пункте систем ИИ, для моделирования поведения человека в социально-экономической среде подходят нейронные сети, экспертные системы и аппарат нечеткой логики. Эволюционные вычисления применяются для других целей (см. выше), а генетические алгоритмы в основном используются для задач оптимизации. Поскольку в своей работе мы исходим из теории ограниченной рациональности, то решение оптимизационных задач здесь неуместно.

Что касается оставшихся трех направлений ИИ, то для разрабатываемой нами модели они равнозначны, но есть отдельные тонкие моменты, исходя из которых, мы отдали предпочтение нейронным сетям.

Как будет рассказано дальше, в модели мы имитируем поведение множества людей, которые воплощены в виде одного «совокупного потребителя». Для его обучения с последующим встраиванием в модель, использовались данные реально проводимых опросов нескольких тысяч респондентов. Результаты работы нейронных сетей, обученных на большом количестве наблюдений, на наш взгляд будут больше соответствовать действительности, чем экспертные системы (исчисляющие предикаты из базы знаний, полученной путем опроса нескольких экспертов) и системы нечеткой логики (использующих правила, также закладываемые несколькими людьми).

## **2. Моделирование экономических агентов с использованием технологий искусственного интеллекта**

Перечисленные в предыдущем пункте системы ИИ нашли свое применение в самых различных областях – бизнесе, медицине, физике и, в том числе, экономике. Однако эти системы в большинстве случаев используются автономно, а не в составе имитационных моделей. К примеру, те же нейронные сети используют для прогноза макроэкономических показателей, так же как и одиночные регрессионные уравнения (Beltratti, Margarita, Terna (1996)), (Garson (1998)), (Engelbrecht (2002)).

Лишь за последние несколько лет стали появляться публикации, авторы которых делятся опытом использования систем ИИ в составе сложных моделей для более адекватного представления деятельности экономических агентов. В то же время моделей, использующих в качестве ИИ нейронные сети, совсем немного.

Хотя новатор в области ограниченной рациональности – Г. Саймон еще в 1970-ых годах отметил, что «за последние 20 лет, благодаря исследованиям в области искусственного интеллекта и когнитивной психологии, наше понимание процедурной рациональности заметно продвинулось. Использование этих достижений в экономической теории могло бы существенно углубить наши представления о динамической рациональности и воздействии на процессы выбора институциональных структур, в рамках которых осуществляется выбор» (Simon (1978)).

Такое отставание во времени связано в первую очередь с неспособностью вычислительных машин того времени численно разрешать модели большой размерности. Теперь современные компьютеры позволяют проводить такие вычисления.

Приведем примеры наиболее известных публикаций по этой теме.

В работе (Baldassarre (1997)) рассматривается модель Бертрана, в которой агенты, каждый из которых представлен совокупностью 30 нейронных сетей, принимают решения об установлении цены на свою продукцию. Веса синапсов этих сетей регулируются с помощью генетических алгоритмов. В процессе «игры» агент выбирает результаты, выдаваемые только одной сетью, просчитывая при этом последствия (прибыль, реакцию других агентов) от применения результатов других сетей. Проведенное исследование было теоретическим, и в нем не использовались реальные статистические данные. Автор хотел показать способность экономических агентов принимать решения посредством применения технологий ИИ.

В другой работе (Zizzo, Sgroi (2000)) рассматривается игра, в которой достигается равновесие Нэша. Один из игроков представлен нейронной сетью, обученной на множестве примеров игр с равновесием по Нэшу. В новых играх, которые «игрок – нейронная сеть» еще не встречал, равновесие достигалось в большинстве случаев.

Немецкий ученый Гротманн (Grothmann (2002)) построил нейронную сеть, имитирующую работу валютного рынка. В отличие от прямого использования нейронных сетей для получения прогнозных оценок, в разработанной им сети каждый нейрон, являясь участником валютного рынка, «принимает решение» о покупке валюты той или иной страны. Результатом одновременно принятых решений всей совокупности нейронов сети является ставка обменного курса национальной валюты.

В статье (Kooths (1999)) описывается своего рода гибрид нейронной сети и системы нечеткой логики, называемый «нейронно-нечетким генератором ожиданий» (Neuro-Fuzzy Expectation Generator – NFEG), который встроен в экономическую модель, рассматривающую товарные и валютные рынки, а также рынки ценных бумаг и рабочей силы.

Основываясь на теории ограниченной рациональности, автор работы критикует подход, согласно которому люди знают все экономические последствия принимаемых ими решений. В противовес классической теории, автором была предложена модель, работающая следующим образом:

1) изначально задается набор нечетких правил типа «ЕСЛИ безработица будет расти, ТО инфляция будет снижаться» и т.д.;

2) далее, эти правила преобразуются в нейронную сеть, веса синапсов которой определяются с помощью модифицированного метода обратного распространения. Измененный автором алгоритм обучения нейронной сети позволяет контролировать процесс «познания» агентами модели окружающей их экономической среды до некоторой степени, недостаточной для полного понимания системы

(это сделано специально, чтобы подчеркнуть ограниченную рациональность);

3) и, наконец, обученная нейронная сеть снова трансформируется в нечеткую систему, которая взаимодействует с другими частями экономической модели.

Помимо перечисленных работ по использованию систем ИИ в экономических моделях, необходимо упомянуть про новое направление в прикладной экономике – «Вычислимой экономике агентов» (Agent-based Computational Economics, ACE), основой которого является моделирование виртуального мира, «населенного» автономными агентами (экономическими, биологическими и т.д.). В проект по созданию подобных миров вовлечено много исследователей, разработки которых выложены на сайте <http://www.econ.iastate.edu/tesfatsi/ace.htm>. Управление созданным виртуальным миром, в соответствии с методологией ACE, осуществляется без вмешательства извне, т.е. только посредством взаимодействия агентов (Tesfatsion (2002)). При этом агенты должны обладать способностью к обучению.

Наиболее популярным прикладным пакетом для моделирования параллельно распределенного виртуального мира является универсальный пакет SWARM, разработанный в Институте Санта Фе (Santa Fe Institute). По своей сути SWARM является набором библиотек, написанных на языке Objective-C, служащих основой для разработок сложных мульти-агентных систем. Этот пакет в свободном доступе выложен в сети по адресу <http://wiki.swarm.org>.

Разработка виртуального мира в SWARM осуществляется за два основных этапа:

1) Создание среды виртуального мира, в которой будут сосуществовать агенты, определяемые на следующем этапе.

2) Создание агентов – объектов виртуального мира (к примеру, людей), с описанием их атрибутов и правил взаимодействий. В процессе своего существования, созданные агенты будут анализировать получаемые от окружающей среды данные, реагировать на них и пополнять свой опыт (обучаться). Этот этап разработки модели наиболее важен, поскольку агенты, корректно отражающие черты своего реального прообраза – залог адекватно построенного виртуального мира.

Более подробно про пакет SWARM можно узнать на официальном сайте (см. выше) и из книги (Economic simulation in SWARM (2000)).

Однако надо отметить, что практически все упомянутые выше модели, использующие ИИ, являются теоретическими. Т.е. они используют абстрактные данные, а цель их разработки заключалась в проверке возможности применения такого инструментария. В отличие от такого

рода разработок, наша модель, как будет показано далее, использует реальные данные и способна выдавать адекватные результаты.

Помимо этого, большинство экономических моделей, в которых используются технологии ИИ, являются итерационными, но в то же время количество итераций ограничивается самими разработчиками, которые «на глазок» определяют точку останова, а в нашей модели итеративный пересчет продолжается до совпадения совокупного спроса и предложения.

В этой работе, в качестве базовой экономической системы, в которую будет встроена совокупность нейронных сетей, мы решили использовать вычислимую модель общего экономического равновесия (Computable General Equilibrium Model, CGE модель). Модели этого класса сами по себе являются новым направлением в прикладной экономике, получившим широкое распространение во всем мире.

Соединяя CGE модель и совокупность нейронных сетей, мы получим симбиоз экономической системы и эмулятора мозга людей из рассматриваемых в модели «виртуальных обществ». О создании таких обществ будет рассказано далее, а пока мы дадим краткую справку о CGE моделях.

### 3. CGE модели

CGE моделям посвящено большое количество иностранной литературы, однако в нашей стране моделям этого класса не уделялось должного внимания. Тем не менее, в публикациях последних лет (Макаров (1999)), (Макаров, Бахтизин (2001)), (Бахтизин (2003)), (Бекларян (2002)), (Бахтизина (2003)), (Alekseev, Tourdyeva, Yudaeva (2003)), (Бесстремянная, Бахтизин (2004)) были описаны недавно созданные CGE модели России и ее регионов, а также был проведен большой обзор зарубежных аналогов, в связи с чем, в этой статье мы не будем проводить анализ работ в области CGE моделирования, а просто дадим характеристику этих моделей.

CGE модели можно определить в трех ключевых аспектах. Во-первых, они включают в себя экономических агентов, результаты деятельности которых, находят отражения во всей экономической системе. Именно поэтому CGE модели называются *общими*. Обычно в число агентов входят домашние хозяйства, фирмы и правительства. Во-вторых, CGE модели включают в себя систему нелинейных уравнений, посредством решения которой, достигается равновесие на рынке каждого товара, услуги и фактора производства. Благодаря этому модели становятся *равновесными*. В-третьих, модели выдают количественные результаты, что позволяет называть их *вычислимыми*.

### 4. Экономические агенты – действующие лица CGE модели

Модель представлена семью экономическими агентами. Первые три из них являются агентами-производителями.



*Экономический агент №1* – государственный сектор экономики. Сюда входят предприятия, доля государственной собственности в которых более 50 процентов.

*Экономический агент №2* – рыночный сектор, состоящий из легально существующих предприятий и организаций с частной и смешанной формами собственности.

*Экономический агент №3* – теневой сектор. Понимание теневого сектора в данной модели двоякое. С одной стороны к теневому сектору относятся нерегистрируемые в статистической отчетности экономические единицы, производящие товары и услуги, а с другой – легально существующие предприятия. В последнем случае в теневом секторе учитывается только их скрытая деятельность. К примеру, работники зарегистрированной фирмы могут производить услуги по ремонту квартир в соответствии с официально заключенным договором (регистраруемая деятельность), но в то же время и за «черный нал» (нерегистрируемая деятельность).

*Экономический агент №4* – совокупный потребитель, объединяющий в себя домашние хозяйства России. Этот экономический агент представлен в модели совокупностью «виртуальных обществ», которые принимают различные решения исходя из сложившейся экономической ситуации. К таким решениям относятся 1) смена работы; 2) изменение потребительских предпочтений. Как уже говорилось выше, процесс принятия решений осуществляется посредством применения нейронных сетей.

*Экономический агент №5* – правительство, представленное совокупностью федерального, региональных и местных правительств, а также внебюджетными фондами. Кроме того, в этот сектор входят некоммерческие организации, обслуживающие домашние хозяйства (политические партии, профсоюзы, общественные объединения и т.д.).

*Экономический агент №6* – банковский сектор, включающий в себя Центральный банк России и коммерческие банки.

*Экономический агент №7* – внешний мир. В данной версии модели все экономические показатели внешнего мира задаются экзогенно. Это значит, что отечественные производители не могут экспортировать больше, чем внешнему миру нужно.

## 5. Общая схема модели

Многие переменные, используемые в этой работе, имеют одинаковую структуру, характерную для CGE моделей. На приведенном ниже рисунке отображена схема, объясняющая индексацию переменных.

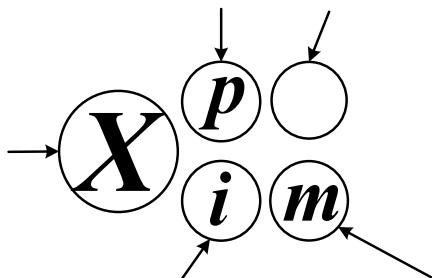


Рисунок 2. Обозначение переменных в модели

Символ  $X$  означает действие, осуществляемое агентом  $i=1\dots7$ . Примером такого действия может быть  $S$  – предложение товара,  $D$  – спрос на товар,  $O$  – определение агентом доли бюджета, идущей на то или иное направление и т.д. Как правило, из контекста понятно о каком действии идет речь.

Индекс  $m$  определяет рынок, на котором агент осуществляет свои действия. Этот индекс присутствует только у переменных, имеющих отношение к торговле. Более детальное описание механизмов торговли будет приведено ниже, а здесь мы перечислим рассматриваемые в модели рынки:

$c^1$ ,  $c^2$  и  $c^3$  – рынки конечных товаров для домашних хозяйств с государственными, рыночными и теневыми ценами соответственно;

$g^1$ ,  $g^2$  – рынки конечных товаров для экономического агента №5 (см. выше) с государственными и рыночными ценами;

$k^1$ ,  $k^2$  – рынки капитальных товаров с государственными и рыночными ценами;

$i^1$ ,  $i^2$  – рынки инвестиционных товаров с государственными и рыночными ценами;

$l^1$ ,  $l^2$ ,  $l^3$  – рынки рабочей силы с государственными, рыночными и теневыми ценами;

$ex$  – рынок экспортных товаров.

**Осуществляемое  
экономическим  
агентом действие**

Конечные товары для домашних хозяйств продаются по всем трем ценам (3 рынка). Конечные товары для экономического агента №5 продаются только по государственным и рыночным ценам (2 рынка). Инвестиционные и капитальные товары также продаются только по государственным и рыночным ценам (еще  $2 + 2 = 4$  рынка).

Таким образом, мы имеем  $3 + 2 + 4 = 9$  внутренних товарных рынков и 1 внешний рынок, на котором продаются экспортные товары.

Помимо товарных рынков в модель включены 3 рынка труда.

Общее число рынков в модели  $9 + 1 + 3 = 13$ .

И, наконец,  $P$  это индекс, означающий, что переменная имеет отношение к продаже или покупке товара по цене  $P_j$ , где  $j = 1, 2, 3$ ,  $ex$  – кодировка цены («1» – регулируемые государством цены, «2» – рыночные цены, «3» – цены теневого рынка, « $ex$ » – экспортные цены). Единицами измерения цен являются их индексы относительно базового периода.

Таким образом, в соответствии с рисунком 2, некоторая переменная  $S_{ic}^{p2}$  означает, что государственный сектор (1) предлагает ( $S$ ) конечный продукт для домашних хозяйств ( $c$ ) по рыночной цене ( $p2$ ).

Отдельные переменные не содержат всех перечисленных индексов, однако общий индекс, обозначающий номер агента ( $i$ ), присутствует во всех переменных.

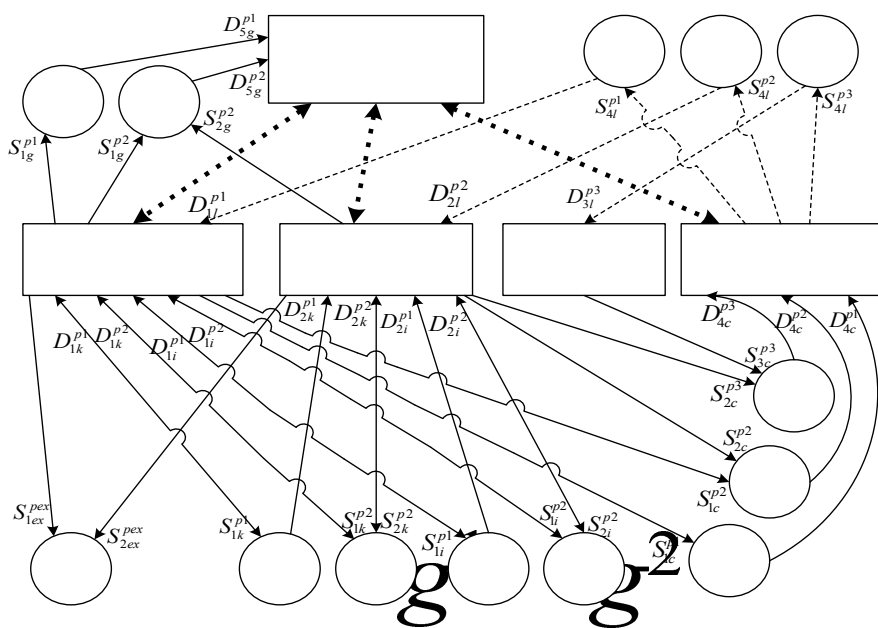
На рисунке 3 представлена концептуальная схема, отражающая работу модели в общем виде.

*Обозначения схемы:*

□ – экономический агент;

○ – рынок, на котором происходит торговля соответствующим товаром между рассматриваемыми в модели экономическими агентами.

«Входящая» на рынок стрелка  $\longrightarrow \bigcirc$  означает, что агент предлагает товар на рынке, а «исходящая» стрелка  $\bigcirc \longrightarrow$ , что агент покупает товар. Тонкая пунктирная стрелка  $\dashrightarrow$  отражает действия агентов, связанные со спросом и предложением рабочей силы, а жирная пунктирная стрелка  $\dashrightarrow$  – налоговые платежи и субсидии.



Го  
со  
пра

Рисунок 3. Концептуальная схема CGE модели

Согласно рисунку 3, государственный и рыночный сектора производят продукт, распределяемый по четырем направлениям:

конечный продукт для домашних хозяйств ( $S_{1c}^{p1}$ ,  $S_{1c}^{p2}$  и  $S_{2c}^{p2}$ ,  $S_{2c}^{p3}$ ), включающий в себя потребительские товары текущего потребления (продукты питания и т.д.), товары длительного потребления (бытовая техника, автомобили и т.д.), а также услуги.

конечный продукт для государственного сектора (Государственный сектор (Г)) ( $S_{1g}^{p1}$ ,  $S_{1g}^{p2}$  и  $S_{2g}^{p2}$ ), состоящий из:

конечного продукта для государственных учреждений (по методологии СНС – расходы государственных учреждений на приобретение конечной продукции), включающего в себя:

бесплатные услуги для населения, оказываемые предприятиями и организациями в области здравоохранения, образования и культуры;

услуги, удовлетворяющие потребности общества в целом, т.е. общее государственное управление, охрана правопорядка, национальная оборона, нерыночная наука, жилищное хозяйство и т.д.;

конечного продукта для некоммерческих организаций, обслуживающих домашние хозяйства, включающего в себя бесплатные услуги социального характера;

инвестиционные товары – затраты на улучшение произведенных и непроизведенных материальных активов (иными словами затраты на создание основного капитала) –  $S_{li}^{p1}$ ,  $S_{li}^{p2}$  и  $S_{2i}^{p2}$ . В этот вид товара не входят государственные (или правительственные) инвестиции, поскольку они учитываются в предыдущем виде товара. В соответствии с методологией СНС, этот вид товара определяется как сумма валового накопления основного капитала и изменения запасов материальных оборотных средств «минус» стоимость приобретенных новых и существующих основных фондов (за вычетом выбытия). Основные фонды в модели выделены в отдельный вид товара.

экспортные товары –  $S_{lex}^{pex}$ ,  $S_{2ex}^{pex}$ . Поскольку одной из составляющих рассмотренных ранее товаров являются импортные товары, то для избежания двойного счета, в экспортные товары входит только чистый экспорт (т.е. экспорт «минус» импорт);

Помимо произведенного продукта, государственный и рыночный сектора торгуют основными фондами (в модели это капитальные товары) –  $S_{lk}^{p1}$ ,  $S_{lk}^{p2}$  и  $S_{2k}^{p2}$ .

Теневой сектор продает только один вид товара – конечный продукт для домашних хозяйств –  $S_{3c}^{p3}$ . Этот экономический агент не платит налоги и не получает субсидии. Для своего производства теневой сектор нанимает и оплачивает работников –  $D_{3i}^{p3}$ .

Таким образом, всего в модели используется 5 видов товаров.

Для производства конечного продукта государственный и рыночный сектора покупают факторы производства:

рабочую силу –  $D_{li}^{p1}$ ,  $D_{2li}^{p2}$ ;

основные фонды –  $D_{lk}^{p1}$ ,  $D_{lk}^{p2}$  и  $D_{2k}^{p1}$ ,  $D_{2k}^{p2}$ ;

инвестиционные товары  $D_{li}^{p1}$ ,  $D_{li}^{p2}$  и  $D_{2li}^{p1}$ ,  $D_{2li}^{p2}$ .

Экономический агент №5 устанавливает налоговые ставки, определяет доли бюджета, идущие на субсидирование производителей и на социальные трансферты, а также расходует средства своего бюджета для покупки конечных товаров –  $D_{5g}^{p1}$ ,  $D_{5g}^{p2}$ , произведенных государственным и рыночным секторами.

Банковский сектор определяет проценты для привлеченных депозитов и выпускает в обращение деньги.

Совокупный потребитель покупает конечные товары, производимые государственным, рыночным и теневым секторами –  $D_{4c}^{p1}$ ,  $D_{4c}^{p2}$  и  $D_{4c}^{p3}$ . Кроме того, в рамках этого сектора определяется предложение рабочей силы для государственного, рыночного и теневого секторов –  $S_{4i}^{p1}$ ,  $S_{4i}^{p2}$  и  $S_{4i}^{p3}$ .

## Рынки

В процессе итеративного пересчета модели, на рынке каждого товара и услуги определяются суммарный спрос и предложение, уравниваемые посредством использования трех различных механизмов:

### 1. Механизм уравнивания на рынке с государственными ценами

Допустим, что суммарный спрос  $D_s^p$  на товар не равен суммарному предложению  $S_s^p$ . Т.е. мы имеем неравенство  $D_s^p > S_s^p$ , либо  $S_s^p > D_s^p$ . Для устранения дисбаланса вводится корректирующий коэффициент, называемый «индикатором дефицитности»  $I = S_s^p / D_s^p$ , который умножается на величину спроса, корректируя ее на каждом шаге итерации. В итерационном процессе индикатор дефицитности стремится к единице.

Поскольку в модели в ряде случаев суммарный спрос  $D_s^p$  на товар есть сумма спросов нескольких агентов, то в реальности, введенный нами коэффициент корректирует долю бюджета каждого агента, идущую на покупку соответствующего товара.

Предположим, что  $D_1^p$  – спрос агента 1, а  $D_2^p$  – спрос агента 2 на один и тот же товар. Спрос обоих агентов в модели определяется следующими соотношениями:  $D_1^p = (O_1^p \cdot B_1) / P_1$  и  $D_2^p = (O_2^p \cdot B_2) / P_1$ , где  $O_1^p$ ,  $O_2^p$  – доли бюджетов  $B_1$ ,  $B_2$  первого и второго агента соответственно. Для корректировки совокупного спроса, доли  $O_1^p$ ,  $O_2^p$  следует умножить на индикатор дефицитности  $I$ , к примеру так:  $O_{i[Q]}^p = O_{i[Q]}^p \cdot \eta + O_{i[Q]}^p \cdot I \cdot (1 - \eta)$ , где  $Q$  – шаг итерации, а  $0 < \eta < 1$  – модельная константа. При ее увеличении равновесие достигается медленнее, однако при этом система уравнений становится более устойчивой.

Ниже приводятся используемые в модели формулы, определяющие индикаторы дефицитности для рынков с государственными ценами:

(1)  $I_m = S_{sm}^{p1} / D_{sm}^{p1}$ , где  $m = l, g, i, k$  (т.е. в этих формулах рассчитываются индикаторы дефицитности для рынков рабочей силы, конечных продуктов для экономического агента №5, инвестиционных и капитальных товаров)

(2)  $I_c = D_{sc}^{p1} / S_{sc}^{p1}$  – индикатор дефицитности для рынка конечных продуктов для домашних хозяйств

Как видно индикатор дефицитности есть частное от деления предложения продукта на его спрос. Исключение составляет формула 2, где спрос делится на предложение. Дело в том, что доля бюджета домашних хозяйств, идущая на покупку конечных товаров, вычисляется с помощью нейронной сети, поэтому для устранения дисбаланса корректируется идущая на продажу доля готового продукта.

### 2. Рыночный механизм уравнивания спроса и предложения

Этот механизм стандартен и выглядит следующим образом:

(3)  $P[Q+1]=P[Q]+(D_{sm[Q]}^{p_2}-S_{sm[Q]}^p)/C$ , где  $Q$  – шаг итерации, а  $C$  – положительное число, называемое «*константой итераций*». При его уменьшении, экономическая система быстрее приходит в состояние равновесия, однако при этом увеличивается опасность ухода цены в отрицательную область.

Теперь запишем формулу модели, отражающие рыночный процесс изменения цен:

(4)  $P_{2m}[Q+1]=P_{2m}[Q]+(D_{sm[Q]}^{p_2}-S_{sm[Q]}^p)/C$ , где  $m=l, c, g, i, k$  (т.е. это уравнения для цен на рабочую силу, конечные товары для домашних хозяйств, конечные товары для экономического агента №5, а также на инвестиционные и капитальные товары)

### 3. Механизм уравнивания спроса и предложения на рынке с теневыми ценами.

Равновесная цена на теневых рынках образуется так же, как и в случае рыночной цены. Ниже приводится соответствующая формула.

(5)  $P_{3m}[Q+1]=P_{3m}[Q]+(D_{sm[Q]}^{p_3}-S_{sm[Q]}^p)/C$ , где  $m=l, c$  (т.е. это уравнения для цен на рабочую силу и конечные товары для домашних хозяйств).

Мы рассмотрели три механизма уравнивания совокупного спроса и предложения на рассматриваемые в модели товары и услуги. Рассмотрим теперь, каким образом образуется **спрос отдельного агента на конкретный товар или фактор производства**.

Итак, спрос на конкретный фактор  $D$  определяется двумя составляющими: суммой денег  $B$ , которую агент готов потратить на покупку фактора и ценой фактора  $P$ , т.е.  $D=B/P$ . Таким образом, экономический агент разделяет свой бюджет на доли по числу покупаемых факторов плюс часть, оставляемая про запас. Конечная формула спроса выглядит следующим образом:

(6)  $D=(O \cdot B)/P$ , где  $O$  – доля бюджета, идущая на конкретное направление.

#### **Образование предложение отдельного агента**

Предложение произведенного агентом товара задается следующим уравнением:

(7)  $S=E \cdot Y$ , где  $E$  – доля от общего выпуска  $Y$ , идущая на конкретное направление.

Ранее уже отмечалось, что суммарные спрос и предложения на рассматриваемые в модели товары формируются как сумма спросов и предложений нескольких агентов.

Для экономии места мы не будем приводить соответствующие формулы, т.к. исходя из рисунка 3 и его описания ясно, какие агенты и по каким ценам формируют конечный спрос и предложения каждого товара.

Переходим к формальному определению действий экономических агентов, в соответствии данным выше описанием:

## 7. Основные формулы модели

По умолчанию, переменные во всех формулах относятся к периоду времени  $t$ . Если переменные относятся к предыдущему  $(t-1)$  или к будущему  $(t+1)$  периоду, то это указывается явно.

### Уравнение производственной функции

Производственные возможности первых трех агентов задаются с помощью производственной функции Кобба-Дугласа

$$Y = f(K, L) = A \cdot K^\alpha \cdot L^\beta.$$

Входными факторами являются труд ( $L$ ) и капитал ( $K$ ). Значение производственной функции показывает добавленную стоимость (конечный продукт), произведенную соответствующим сектором. Оценка ее коэффициентов (таблица 1) производилась в соответствии с вкладом факторов:

коэффициент при труде  $\alpha$ : фонд заработной платы сектора, поделенный на валовую добавленную стоимость, произведенную сектором;

коэффициент при основных фондах:  $\beta = 1 - \alpha$

Таблица 1. Коэффициенты ПФ для трех секторов экономики

Сектор экономики	Труд	Основные фонды
Государственный	0,33	0,67
Рыночный	0,40	0,60
Теневой	0,49	0,51

Запишем производственные функции первых трех агентов (агентов – производителей).

$$(8) Y_1 = A_1^r \cdot \left( (K_{1(t)} + K_{1(t+1)}) / 2 \right)^{\alpha_1^r} \cdot (D_{1t}^{p1})^{\beta_1^r},$$

$$(9) Y_2 = A_2^r \cdot \left( (K_{2(t)} + K_{2(t+1)}) / 2 \right)^{\alpha_2^r} \cdot (D_{2t}^{p2})^{\beta_2^r},$$

$$(10) Y_3 = A_3^r \cdot \left( (K_{3(t)} + K_{3(t+1)}) / 2 \right)^{\alpha_3^r} \cdot (D_{3t}^{p3})^{\beta_3^r},$$

где  $A_1^r$ ,  $A_2^r$ ,  $A_3^r$  – коэффициенты ПФ, первый из которых является коэффициентом размерности, а два последних – коэффициенты при основных фондах и труде.

Аргументами производственных функций (ПФ) являются ресурсы: основные фонды  $K_i$  и труд  $D_{it}^{pj}$ . В модели используется производственная функция с эффективностью использования ресурсов, не зависящей от масштаба производства, иными словами она возрастает пропорционально росту количества ресурсов.

В качестве второго аргумента ПФ берется спрос экономического агента на рабочую силу. Полученное значение производственной функции объявляется предложением конечных товаров. Формально это выглядит так, что спрос на факторы преобразуется в предложение товаров, что с точки зрения логики выглядит бессмысленным, если



рассматривать это действие отдельно от итерационного процесса, смысл которого состоит в уравнивании спроса и предложения.

Ранее, в общем виде уже описывался процесс образования предложения продукта и спроса на факторы производства, в связи с чем, для краткости изложения конечные формулы по каждому продукту мы опускаем.

В следующих формулах подсчитывается **выручка секторов-производителей** от продаж произведенного продукта по различным направлениям и от продаж основных фондов (товарные потоки были представлены на схеме рисунке 3).

$$(11) Y_1^p = S_{1c}^{p1} \cdot P_{1c} + S_{1c}^{p2} \cdot P_{2c} + S_{1g}^{p1} \cdot P_{1g} + S_{1g}^{p2} \cdot P_{2g} + S_{1i}^{p1} \cdot P_{1i} + S_{1i}^{p2} \cdot P_{2i} + S_{1k}^{p1} \cdot P_{1k} + S_{1k}^{p2} \cdot P_{2k} + S_{1ex}^{pex} \cdot P_{ex}$$

$$(12) Y_2^p = S_{2c}^{p2} \cdot P_{2c} + S_{2g}^{p2} \cdot P_{2g} + S_{2i}^{p2} \cdot P_{2i} + S_{2k}^{p2} \cdot P_{2k} + S_{2ex}^{pex} \cdot P_{ex}$$

$$(13) Y_3^p = (S_{2c}^{p3} + S_{3c}^{p3}) \cdot P_{3c}$$

Далее, к выручке от продаж плюсятся субсидии из средств консолидированного бюджета  $G_1^s, G_2^s$  и внебюджетных фондов  $G_1^f, G_2^f$  (только для государственного и рыночного секторов). Помимо этого добавляются эмиссионные деньги  $M_1, M_2$  и средства, находящиеся на банковских счетах (с учетом процентов по вкладам)  $B_i^b \cdot (1 + P_{b\%(t-1)})$ .

Таким образом формируется **бюджет агента-производителя**:

$$(14) B_1 = B_1^b \cdot (1 + P_{b\%(t-1)}) + Y_1^p + G_1^s + G_1^f + M_1$$

$$(15) B_2 = B_2^b \cdot (1 + P_{b\%(t-1)}) + Y_2^p + G_2^s + G_2^f + M_2$$

$$(16) B_3 = B_3^b \cdot (1 + P_{b\%(t-1)}) + Y_3^p$$

**Динамика остатков средств** агентов-производителей определяется следующим образом:

$$(17) B_{1(t+1)}^b = O_1^s \cdot B_1$$

$$(18) B_{2(t+1)}^b = O_2^s \cdot B_2$$

$$(19) B_{3(t+1)}^b = O_3^s \cdot B_3$$

В этих формулах  $O_i^s$  – **неизрасходованная доля бюджета**, определяемая с учетом затрат на факторы производства и уплату налогов в консолидированный бюджет  $O_1^i, O_2^i$  и во внебюджетные фонды  $O_1^f, O_2^f$ :

$$(20) O_1^s = 1 - O_{1l}^{p1} - O_{1k}^{p1} - O_{1k}^{p2} - O_{1i}^{p1} - O_{1i}^{p2} - O_1^i - O_1^f$$

$$(21) O_2^s = 1 - O_{2l}^{p2} - O_{2k}^{p1} - O_{2k}^{p2} - O_{2i}^{p1} - O_{2i}^{p2} - O_2^i - O_2^f$$

$$(22) O_3^s = 1 - O_{3l}^{p3}$$

И, наконец, приведем формулы для определения **динамики основных фондов** государственного, рыночного и теневого секторов:

$$(23) K_{1(t+1)} = K_1 \cdot (1 - E_{1k}^{p1} - E_{1k}^{p2}) \cdot K_a + D_{1k}^{p1} + D_{1k}^{p2} + D_{1i}^{p1} + D_{1i}^{p2}$$

$$(24) K_{2(t+1)} = K_2 \cdot (1 - E_{2k}^{p2}) \cdot K_a + D_{2k}^{p1} + D_{2k}^{p2} + D_{2i}^{p1} + D_{2i}^{p2}$$

$$(25) K_3 = \gamma \cdot (K_1 + K_2)$$

где  $K_a$  – коэффициент выбытия основных фондов;

$\gamma$  – доля основных фондов государственного и рыночного секторов, используемых в теневой экономике.

Теперь, в общих чертах, определим другие экономические агенты.

### **Совокупный потребитель (домашние хозяйства)**

Выше уже говорилось, что домашние хозяйства представлены в модели совокупностью нескольких виртуальных обществ, принимающих различные решения исходя из сложившейся ситуации. Эти решения, как уж было отмечено, формулируются с помощью нейронных сетей – одного из направлений ИИ.

Приведем некоторые формальные зависимости, определяющие поведение экономического агента №4.

### **Предложение рабочей силы**

В модели существуют следующие виды перетоков рабочей силы:

государственный сектор → рыночный сектор (доля  $L_1^2$  от величины  $L_1$ );

рыночный сектор → государственный сектор (доля  $L_2^1$  от величины  $L_2$ );

государственный и рыночный сектора → теневой сектор (доля  $L_{12}^3$  от величины  $L_1 + L_2$ ).

Эти доли определяются с помощью нейронных сетей. Об этом будет рассказано чуть позже.

Ниже представлены **уравнения, определяющие баланс рабочей силы в разных секторах-производителях:**

*в государственном секторе:*

$$(26) L_1 = L_{1(t-1)} \cdot (1 - L_{1(t-1)}^2 + L_{1(t-1)}^a - L_{1(t-1)}^i) + L_{2(t-1)} \cdot L_{2(t-1)}^1$$

*в рыночном секторе:*

$$(27) L_2 = L_{2(t-1)} \cdot (1 - L_{2(t-1)}^1 + L_{2(t-1)}^a - L_{2(t-1)}^r) + L_{1(t-1)} \cdot L_{1(t-1)}^2$$

*в теневом секторе:*

$$(28) L_3 = (L_1 + L_2) \cdot L_{12}^3$$

где  $L_1^a$ ,  $L_1^r$ ,  $L_2^a$ ,  $L_2^r$  – доли прибивающих (к примеру, начавших свою трудовую деятельность в конкретном секторе) и выбывающих (к примеру, вышедших на пенсию) работников.

**Бюджет домашних хозяйств задается следующей формулой:**

$$(29) B_4 = B_4^b \cdot (1 + P_{b\%}^b) + B_{4(t-1)} \cdot O_{4(t-1)}^s + W_1 + W_2 + W_3 + G_4^r + G_4^f + M_4$$

Бюджет формируется из денег, отложенных на счетах в банках, нераспределенных наличных денег, остающихся с предыдущего периода  $B_{4(t-1)} \cdot O_{4(t-1)}^s$ , заработной платы, получаемой в государственном  $W_1$ , рыночном  $W_2$  и теневом секторах  $W_3$ , а также пенсий, пособий и субсидий, получаемых из средств консолидированного бюджета и внебюджетных фондов  $G_4^r$ ,  $G_4^f$ . Помимо перечисленного, в модели учитывается эмиссия наличных денег  $M_4$ .

**Заработная плата** работников различных секторов определяется следующим образом:

$$(30) W_1 = D_{1l}^{p1} \cdot P_{1l}$$

$$(31) W_2 = D_{2l}^{p2} \cdot P_{2l}$$

$$(32) W_3 = D_{3l}^{p3} \cdot P_{3l}$$

Что касается **остатка наличных средств**, переходящих на следующий период, то динамика этого показателя определяется так:

$$(33) O_4^s = 1 - O_{4c}^{p1} - O_{4c}^{p2} - O_{4c}^{p3} - O_4^{tax} - O_4^b - O_4^s$$

Здесь учитываются расходы на конечные товары по государственным  $O_{4c}^{p1}$ , рыночным  $O_{4c}^{p2}$  и теневым  $O_{4c}^{p3}$  ценам, а также средства идущие на покупку валюты  $O_4^s$ , уплату налогов  $O_4^{tax}$  и на сбережения в банках  $O_4^b$ .

Доли бюджета  $O_{4c}^{p1}$ ,  $O_{4c}^{p2}$ ,  $O_{4c}^{p3}$ ,  $O_4^s$  и  $O_4^b$  также определяются нейронными сетями.

Все перечисленные здесь доли бюджета участвуют в уравнениях, в которых формируется спрос на потребляемые продукты. Из-за экономии места мы не будем приводить здесь конечные уравнения, а просто сошлемся на приведенный выше способ (пункт 6) образования упомянутых величин.

Теперь настала пора рассказать про используемые в модели нейронные сети.

### **Обучение нейронных сетей**

Выше уже говорилось, что для обучения используемых в модели нейронных сетей использовались данные реально проводимых опросов.

Опишем вкратце процесс обработки социологических баз данных RLMS, используемых нами для обучения нейронных сетей.

RLMS – *Russian Longitudinal Monitoring Survey* или Российский мониторинг экономического положения и здоровья населения (РМЭЗ) представляет собой серию проводившихся в Российской Федерации в 1992 – 2001 гг. репрезентативных общенациональных опросов, реализованных в два этапа (две различные выборки). Второй этап мониторинга включает в себя шесть волн обследований: 1) 5-ая волна – осень 1994 года; 2) 6-ая волна – осень 1995 года; 3) 7-ая волна – осень 1996 года; 4) 8-ая волна – осень 1998 года; 5) 9-ая волна – осень 2000 года; 6) 10-ая волна – осень 2001 года.

Каждое обследование представляет собой ответы на более чем 3 тысячи вопросов, сгруппированных по трем вопросам, включающих в себя информацию по нескольким разделам:

вопросник для взрослых (разделы: *миграция, работа, медицинское обслуживание, оценка здоровья, бюджет времени*);

вопросник для детей (разделы: *уход за детьми, медицинское обслуживание, оценка здоровья, бюджет времени*);

семейный вопросник (разделы: *информация о семье, жилищные условия, земледелие и животноводство, расходы, доходы*).

В среднем, в каждой волне опрашивалось около 10000 взрослых, 2000 детей и 4000 домохозяйств.

Уникальность этого исследования заключается в том, что интервьюеры старались найти всех людей участвовавших в исследовании ранее, и когда находили переехавших, то опрашивали их по новым адресам. Таким образом, благодаря проведенным обследованиям можно проследивать поведение людей в динамике.

Для своих целей мы использовали данные второго этапа мониторинга, поскольку на первом этапе использовалась другая репрезентативная выборка.

В модель включены пять нейронных сетей, три из которых определяют поведение человека в плане смены работы, а остальные две определяют способ распределения бюджета домохозяйства.

Ниже описывается процедура построения **нейронной сети №1**, которая определяет долю  $L_1^2$  от величины  $L_1$ , т.е. количество работников государственного сектора, переходящих в рыночный сектор.

Для построения этой сети использовались анкеты для взрослых, из которых были отобраны вопросы, относительно работы респондентов. Полученный массив данных обрабатывался следующим образом:

Этап 1. Отбирались только те люди, которые участвовали во всех волнах обследований (для того, чтобы можно было проследить поведение человека во времени).

Этап 2. Отбирались только те, у кого есть работа. Среди них отбирались респонденты, указавшие свою зарплату, а уже среди них отбирались ответившие на вопрос о том, какая форма собственности у предприятия-работодателя.

Этап 3. В каждую волну обследований была включена новая переменная  $Z_1^j$ , представляющая собой частное от деления индекса номинальной заработной платы каждого респондента и индекса потребительских цен:

$$(34) Z_1^j = \left( \frac{W_{1(t)}^j}{W_{1(t+1)}^j} \right) / \left( \frac{P_{1c(t+1)} + P_{2c(t+1)} + P_{3c(t+1)}}{P_{1c(t)} + P_{2c(t)} + P_{3c(t)}} \right),$$

где  $W_{1(t)}^j$  – заработная плата отдельного респондента  $j$ , работающего в государственном секторе в момент времени  $t$ , взятая из анкет RLMS.

Как видно, в этой формуле используется суммарный индекс потребительских цен всех трех рынков конечного продукта для домашних хозяйств.

Поскольку в массиве данных представлено шесть волн, то таких переменных получилось пять: 1) значения 1995 года к значениям 1994 года; 2) 1996 к 1995; 3) 1998 к 1996; 4) 2000 к 1998; 5) 2001 к 2000. На данном этапе обработки, массив данных представляет собой таблицу, столбцами которой являются упомянутые выше переменные, повторяющиеся по годам, а строками – наблюдения.

Этап 4. Для обучения нейронной сети нам необходимы только две переменные: 1)  $Z_1^t$  и 2) индикатор, отражающий форму собственности предприятия-работодателя в момент времени  $t+1$ , при условии что в момент времени  $t$  человек работал в государственном секторе. Этот индикатор принимает два значения: «1» – работник остался в государственном секторе, «2» – работник перешел в рыночный сектор. Таким образом, остальные переменные массива уже обработаны и больше не нужны, поэтому мы их удалили. Оставшиеся две переменные были «склеены» в массив из двух переменных, содержащий 1097 наблюдений для непосредственного обучения нейронной сети.

Этап 5. Введем в рассмотрение еще одну переменную  $V_1^j$ , показывающую разницу между переменной (34) для отдельно взятого работника и переменной  $Z_1$  «совокупного работника»:

$$(35) V_1^j = Z_1^j / Z_1, \text{ где}$$

$Z_1 = ((W_{1(t+1)} / W_{1(t)}) / ((P_{1c(t+1)} + P_{2c(t+1)} + P_{3c(t+1)}) / (P_{1c(t)} + P_{2c(t)} + P_{3c(t)})))$  – частное от деления индекса номинальной заработной платы «совокупного работника» государственного сектора и индекса потребительских цен.

Для обучения нейронной сети переменная (35) не потребуется, однако она нужна для встраивания сети в CGE модель. Но об этом чуть позже.

Этап 6. Вкратце, мы уже упоминали о процессе обучения (пункт 1), однако нет необходимости его автоматизировать самостоятельно, поскольку для этих нужд в настоящий момент существует достаточное количество прикладных программ, которые помимо всего прочего позволяют конструировать наиболее подходящую для конкретного случая архитектуру сети. Самыми известными среди них являются *NeuroSolutions* от компании *NeuroDimension* и *STATISTICA Neural Networks* от компании *StatSoft*. Эти пакеты предоставляют возможность получить значения весов синапсов нейронов, которые в дальнейшем были использованы нами для симулирования работы нейронных сетей (в пакете *Microsoft Excel*) в составе CGE модели.

Как уже говорилось выше, результатом последнего этапа обработки данных стал массив из двух переменных содержащий 1097 наблюдений. В некоторых случаях значения переменной  $Z_1^j$  были слишком большими, что сделало необходимым очистить массив данных от подобных выбросов, в результате чего окончательное число наблюдений стало равным 839.

Таким образом, с помощью пакета *STATISTICA Neural Networks*, первая нейронная сеть была обучена на 839 наблюдениях.

### Включение нейронной сети в CGE модель

Все наблюдения, ранее используемые для обучения нейронной сети, в CGE модели интерпретируются как 839 человек «виртуального общества», которые *принимают решение о переходе в рыночный сектор или о продолжении работы в государственном секторе*. Принимаемое с помощью нейронной сети решение базируется на размышлении о целесообразности дальнейшей работы в секторе, исходя из изменения своей покупательной способности. Иными словами *в процессе работы CGE модели каждому члену «виртуального общества» подается следующая информация:*

(36)  $U_1^j = Z_1 \cdot V_1^j$  – входная переменная нейронной сети, где  $V_1^j$  – константа «различия людей», рассчитанная ранее по формуле (35), а  $Z_1$  – общая для всех работников переменная, изменяющаяся в процессе итеративного пересчета.

Теперь что касается выходной переменной нейронной сети

В соответствии с выбранной нами топологией нейронной сети и функции активации ее нейронов, для вычисления выходной переменной используется следующая формула:

(37)  $N_1^j = (\sum (1/(1 + e^{-(U_1^j \cdot r_{(1)scale}^{in} + r_{(1)shift}^{in}) \cdot w_1^{2k} - \tau_1^3)})) \cdot w_1^{3k} - \tau_1^3 - r_{(1)shift}^{out}) / r_{(1)scale}^{out}$ , где  $r_{(1)scale}^{in}$ ,  $r_{(1)shift}^{in}$ ,  $\tau_1^{2k}$ ,  $\tau_1^3$ ,  $w_1^{2k}$ ,  $w_1^{3k}$ ,  $r_{(1)scale}^{out}$ ,  $r_{(1)shift}^{out}$  – технические переменные сети, определяемые программным пакетом.

Естественно, что нейронная сеть не выдает значения в точности равные «1» или «2», в связи с чем, в модели они округляются до ближайшего целого. Таким образом, выходная переменная сети  $N_1^j$ , получаемая по преобразуется следующим образом:

(38)  $N_1^j = 1$  если  $1,5 \leq N_1^j < 2,5$ ;  $0$  если  $0,5 < N_1^j < 1,5$

Такое преобразование необходимо для вычисления доли работников, от их общего количества  $L_1$ , перешедших в рыночный сектор:

(39)  $L_1^2 = (\sum_{j=1}^{839} N_1^j) / 839$

Полученная таким образом доля используется в уравнениях (26)–(28).

По тому же принципу строились:

- **нейронная сеть № 2**, определяющая долю  $L_2^1$  от величины  $L_2$ , т.е. количество работников рыночного сектора, переходящих в государственный сектор;
- **нейронная сеть № 3**, определяющая долю  $L_{12}^3$  от величины  $L_1 + L_2$ , т.е. количество работников государственного и рыночного секторов, работающих в теневом секторе.

Полученные с их помощью доли также использовались в уравнениях (26) – (28).

**Нейронные сети №№ 4, 5**, определяющие доли бюджета  $O_{4c}^{p1}$ ,  $O_{4c}^{p2}$  и  $O_{4c}^{p3}$  совокупного потребителя, идущие на покупку конечных товаров по государственным, рыночным и теневым ценам, а также доли бюджета  $O_4^s$ ,  $O_4^b$ , идущие на покупку валюты и на вклады в банках строились по похожему принципу, но только для их обучения использовались семейные вопросники, из которых были отобраны вопросы, конкретизирующие расходы домашних хозяйств на конечные товары, покупку валюты и сбережения.

### Государство (экономический агент №5)

Как уже говорилось, этот экономический агент представлен совокупностью федерального, региональных и местных правительств, а также внебюджетными фондами. Этот экономический агент устанавливает типы налогов и налоговые ставки (в модели учитываются основные налоги, среди которых НДС  $T^{vad}$ , ЕСН  $T^{esn}$ , а также налоги на прибыль организаций  $T^{pr}$ , имущество  $T^{prop}$  и доходы физических лиц  $T^{pod}$ ).

Собранные деньги формируют консолидированный бюджет и внебюджетные фонды:

$$(40) B_5 = O_1^i \cdot B_1 + O_2^i \cdot B_2 + O_4^i \cdot B_4 + B_5^{other} + B_5^b \cdot (1 + P_{b\%(t-1)})$$

$$(41) F_5 = O_1^f \cdot B_1 + O_2^f \cdot B_2 + F_5^b \cdot (1 + P_{b\%(t-1)})$$

где,  $B_5^{other}$  – экзогенно вводимое в модель значение, представляющее собой сумму других налогов (не вошедших в перечень рассматриваемых), неналоговых доходов и прочих доходов консолидированного бюджета. Кроме различного вида поступлений, к полученным суммам приплюсовываются средства, находящиеся на банковских счетах  $B_5^b \cdot (1 + P_{b\%(t-1)})$ ,  $F_5^b \cdot (1 + P_{b\%(t-1)})$ .

В свою очередь, сумма средств на банковских счетах определяется следующим образом:

$$(42) B_{5(t+1)}^b = O_{5b}^s \cdot B_5$$

$$(43) F_{5(t+1)}^b = O_{5f}^s \cdot F_5$$

где,  $O_{5b}^s$ ,  $O_{5f}^s$  – доли нераспределенного консолидированного бюджета нераспределенных средств внебюджетных фондов, которые образуются по аналогии с уравнениями (20) – (22) и (33) с учетом затрат на покупку конечных товаров  $O_{5g}^{p1}$ ,  $O_{5g}^{p2}$ , субсидий государственного и рыночного секторов  $O_{5g}^{p1}$ ,  $O_{5g}^{p2}$ ,  $O_5^{f1}$ ,  $O_5^{f2}$  и социальных трансфертов населению  $O_5^{pr}$ ,  $O_5^{f4}$ .

Отметим, что, как и в рассмотренных ранее случаях эти доли участвуют в уравнениях, в которых формируется спрос на потребляемые продукты. Из-за экономии места мы не будем приводить здесь конечные уравнения, а просто сошлемся на приведенный выше способ обозначения упомянутых величин.

### Банковский сектор (экономический агент №6)

Банковский сектор в модели включает в себя Центральный банк и коммерческие банки и осуществляет функции по эмиссии денег  $M_1$ ,  $M_2$ ,  $M_4$  и устанавливает ставки по депозитам для предприятий  $P_{b\%}$  и физических лиц  $P_{b\%}^h$ . Естественно, что функции рассматриваемых структур в реальной жизни гораздо шире, но в модели этот сектор нужен только для баланса денежных потоков.

### Интегральные показатели модели

Приведем некоторые формулы, рассчитывающие интегральные показатели экономики России.

ВВП России (в ценах базового периода):

$$(44) Y = Y_1 + Y_2$$

ВВП России (в текущих ценах):

$$(45) Y_p = Y_1^p + Y_2^p$$

Инфляция потребительских цен:

$$(46) P = P_{2c(t)} / P_{2c(t-1)}$$

Количество людей, занятых в экономике:

$$(47) L = L_1 + L_2 + L_3$$

Основные фонды:

$$(48) K = K_1 + K_2$$

## 8. Механизм работы модели

Переменные в модели подразделяются на экзогенные и эндогенные. Экзогенные переменные разбиваются на *заданные переменные*: доли бюджета и произведенного продукта и *переменные экономической политики*: доли консолидированного бюджета, идущие на субсидии и трансферты, процентные ставки налогов и т.д.

Вначале рассчитываются значения экзогенных переменных, затем рассчитываются параметры уравнений модели и в результате одновременного пересчета уравнений модели находятся значения эндогенных переменных: валового внутреннего продукта, цен на товары, доходы населения и т.д.

Пересчет уравнений модели происходит до совпадения совокупного спроса и предложения на рынке каждого товара и услуги, рассматриваемых в модели, посредством итеративного процесса с помощью соответствующих прикладных пакетов. Сходимость модели достигается примерно на 10000 шаге итераций.

Как уже говорилось выше, на каждой итерации происходит корректировка долей бюджета экономических агентов (в случае государственной цены) или изменение цены (в случае рыночного или теневого механизма уравнивания спроса и предложения).

Что касается действий потребителя, то на каждой итерации три нейронные сети определяют параметры трудовой мобильности, а еще



две определяют доли бюджета домашних хозяйств, идущие на покупку конечных товаров, на сбережения и на покупку валюты.

При проведении вычислительных экспериментов сначала мы меняем какой-нибудь экзогенный параметр, а затем запускаем процедуру итеративного пересчета, в процессе которого модель имитирует реальные действия экономических агентов.

### **9. Вычислительный эксперимент**

После разработки модели мы провели вычислительный эксперимент, для того чтобы показать применимость модели для решения актуальных задач российской экономики.

Так, в модели был снижен единый социальный налог в соответствии с предложенным правительством вариантом его снижения.

Объяснение необходимости снижения ставки ЕСН заняло бы слишком много места, поэтому мы сразу перейдем к краткому изложению результатов.

Итак, по результатам расчетов, снижение ЕСН привело к небольшому приросту ВВП (к 2010 году прирост составил 0,759% по сравнению с базовым вариантом развития, т.е. без снижения ставки налога).

Годовая зарплата одного работника государственного сектора практически не изменилась, а в рыночном секторе увеличилась (на 2,8% в 2010 году). Поскольку значения инфляции потребительских остались на прежнем уровне, то дополнительные доходы работников «не съедаются» из-за роста цен.

Бюджет домашних хозяйств незначительно уменьшился. По всей видимости, это связано с тем, что в модели были использованы не компенсирующие механизмы и внебюджетные фонды получили недостаточно денег для запланированных выплат пенсий и пособий.

Количество работников теневого сектора осталось прежним.

Несмотря на значения отдельных показателей можно сказать, что в модели, реформа по снижению ЕСН оказала положительное влияние на экономику страны

## ЛИТЕРАТУРА

1. *Адаменко А. Н., Кучуков А.М.* (2003): Логическое программирование и Visual Prolog. СПб.: БХВ-Петербург.
2. *Бахтизин А.Р.* (2003): Вычислимая модель «Россия: Центр – Федеральные округа». Препринт # WP/2003/151. М.: ЦЭМИ РАН.
3. *Бахтизина Н.В.* (2003): CGE модель конкурирующих партий России // Материалы четвертого всероссийского симпозиума «Стратегическое планирование и развитие предприятий». М.: ЦЭМИ РАН.
4. *Бекларян Г.Л.* (2002): Анализ эффективности экономической политики России с помощью вычислимой модели общего равновесия, описывающей взаимодействие совокупного потребителя, совокупного производителя и государства. Препринт # WP/2002/143. М.: ЦЭМИ РАН.
5. *Бесстремляная Г.Е., Бахтизин А.Р.* (2004): Вычислимая модель «Социальная Россия». Препринт # WP/2004/173. М.: ЦЭМИ РАН.
6. *Бобровский С.* (2001): Досье искусственного интеллекта. <http://www.computer-museum.ru/frgnhist/ai1.htm>.
7. *Гилл Ф., Мюррей У., Райт М.* (1985): Практическая оптимизация. М.: Мир.
8. *Круглов В.В., Борисов В.В.* (2002): Искусственные нейронные сети. Теория и практика. М.: Горячая линия-Телоком.
9. *Левнер Е.В., Птускин А.С., Фридман А.А.* (1998): Размытые множества и их применение. – М.: ЦЭМИ РАН.
10. *Макаров В.Л.* (1999): Вычислимая модель российской экономики (RUSEC). Препринт # WP/99/069. М.: ЦЭМИ РАН.
11. *Макаров В.Л., Бахтизин А.Р.* (2001): Эффективный способ оценки государственной политики // Экономика и управление. № 4.
12. *Масалович А. И.* (1995): Этот нечеткий, нечеткий, нечеткий мир. <http://www.tora-centre.ru/library/fuzzy/fuzzy.htm>.
13. *Пивкин В.Я., Бакулин Е.П., Кореньков Д.И.* (1998): Нечеткие множества в системах управления. [http://idisys.iae.nsk.su/fuzzy\\_book/content.htm](http://idisys.iae.nsk.su/fuzzy_book/content.htm).
14. *Пью Д.С., Хиксон Д. Дж.* (1999): Исследователи об организациях. Хрестоматия. М.: ЛИНК.
15. Российский мониторинг экономического положения и здоровья населения (RLMS). Ежегодные обследования. Carolina Population Center at the University of North Carolina at Chapel Hill.
16. *Саймон Г.* (2000): Теория принятия решений в экономической теории и науке о поведении. М.: Теория фирмы.
17. *Alekseev A., Tourdyeva N., Yudaeva K.* (2003): Estimation of the Russia Trade Policy with the Help of the Computable General Equilibrium Model. CEFIR Academic papers.
18. *Baldassarre G.* (1997): Neural networks and genetic algorithms for the simulation models of bounded rationality theory: An application to oligopolistic markets // Rivista di Politica Economica. V. 12, pp. 107-146.
19. *Beltratti A., Margarita S. and Terna P.* (1996): Neural Networks for Economic and Financial Modelling. London: International Thomson Computer Press.
20. *Bishop C.* (1995): Neural Networks for Pattern Recognition. Oxford: University Press.

21. Economic simulation in SWARM: agent-based modeling and object oriented programming (2000) / edited by Francesco Luna and Benedikt Stefansson. Kluwer Academic Publishers.
22. *Engelbrecht A.P.* (2002): Computational Intelligence: An Introduction. Chichester: John Wiley.
23. *Garson G.D.* (1998): Neural Networks: An Introductory Guide for Social Scientists. London: Sage Publications.
24. *Gigerenzer G., Selten R.* (2002): Bounded Rationality: The Adaptive Toolbox. MIT Press.
25. *Grothmann R.* (2002): Multi-Agent Market Modeling based on Neural Networks. Thesis presented for the Degree of Doctor of Economics. Bremen University.
26. *Koohs S.* (1999): Modelling Rule- and Experience-Based Expectations Using Neuro-Fuzzy Systems. University of Muenster. Germany. <http://www-wiwi.uni-muenster.de/~09/makromat/cef99/cef99-koohs.pdf>.
27. *Kosko B.* (1992): Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Englewood Cliffs, NJ: Prentice Hall.
28. *Patterson D.* (1996): Artificial Neural Networks. Singapore: Prentice Hall.
29. *Rubinstein A.* (1997): Modeling Bounded Rationality. MIT Press.
30. *Sargent T.J.* (1994): Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures. Clarendon Press.
31. *Simon H.A.* (1978): Rationality as Process and as Product of Thought. Richard T. Ely Lecture // American Economic Review. V. 68, no.2, p.1–16.
32. *Tesfatsion L.* (2002): Agent-Based Computational Economics: Modelling Economies as Complex Adaptive Systems. <http://www.econ.iastate.edu/tesfatsi>.
33. *Turing A.M.* (1950): Computer machinery and intelligence // MIND: A quarterly review of psychology and philosophy. V. LIX. no 236.
34. *Zizzo D.J., Sgroi D.* (2000): Bounded-Rational Behavior by Neural Networks in Normal Form Games. Nuffield College Oxford Economics Discussion Paper. No. 2000-W30

# VI

## СОЦИОКУЛЬТУРНЫЕ АСПЕКТЫ

---

---

### СТАНОВЛЕНИЕ ИНФОРМАЦИОННОГО ОБЩЕСТВА В РОССИИ И НАЦИОНАЛЬНАЯ БЕЗОПАСНОСТЬ

*К.К. Колин*

#### **1. Состояние «электронной готовности» России в процессе формирования глобального информационного общества**

За последние годы в России созданы серьезные предпосылки для формирования основ информационного общества. Об этом свидетельствуют следующие характеристики процессов развития российского общества:

- Сформировался и быстро прогрессирует отечественный рынок информационных и коммуникационных технологий, продуктов и услуг.
- В значительной степени компьютеризированы многие отрасли хозяйства, в том числе, банковская сфера и сфера государственного управления.
- Создана база для законодательного и нормативного обеспечения развития информационно-коммуникационных технологий (ИКТ).
- В общественном мнении складывается понимание актуальности задачи использования ИКТ в бизнесе, политике и управлении, в здравоохранении и культуре, в науке и образовании.
- Высокий уровень образования населения России создает благоприятные предпосылки для дальнейшего информационного развития страны. Доля лиц с высшим и незаконченным высшим образованием в экономически активном населении России составляет сегодня 20%, а по численности студентов на 1000 человек населения Россия сегодня находится на уровне передовых экономически развитых стран мира.
- Растет число студентов, получающих образование в сфере ИКТ, и по их числу, приходящемуся на 1000 человек, Россия не уступает таким странам, как Франция, Швеция и Германия.

Таким образом, информационное развитие России идет достаточно быстрыми темпами, о чем свидетельствуют также и следующие показатели:

Объем отечественного рынка ИКТ ежегодно растет более чем на **20%**.

Ежегодный прирост компьютерного парка составляет не менее **25%**.

Положительная динамика характерна для всего сектора информационной экономики. Уверенно растет реальный объем информационно-коммуникационных услуг населению.

Идет активное внедрение ИКТ в такие социально важные сферы деятельности, как образование, государственное и корпоративное управление, экономика и другие области общественной жизни.

И, тем не менее, в рейтинге *«готовности к электронному развитию»*, представленном участникам Всемирного экономического форума, Россия занимает только **63 место** (из 102 стран мира).

Однако, существующие в настоящее время темпы роста в области информационных технологий свидетельствуют о том, что для России эта область является одним из наиболее перспективных секторов рынка [8]. Так, например, на долю России в 2004 году пришлось около 47% всех затрат на информационные технологии в Центрально-восточном европейском регионе в то время, как в 2003 году эти затраты составляли 44%.

Приведем также и некоторые другие количественные характеристики информационного развития России:

- Российские инвестиции в основной капитал отрасли составили в 2004 году \$4,1 миллиардов, увеличившись за год на 24,6%.
- Иностранные инвестиции, по сравнению с 2003 годом, возросли вдвое и составили порядка \$1,4 млрд. долларов.
- Рост ИТ-рынка в 2004 году составил 20 процентов, а в четвертом квартале 2004 года рынок персональных компьютеров (ПК) вырос по сравнению с аналогичным периодом 2003 года более чем на 35%.
- В течение 2004 года произошло удвоение российского рынка мобильных ПК, что более чем в 3 раза превышает аналогичный средний показатель для стран Западной Европы.
- За 2005 год общее количество мобильных телефонов в России возросло почти на 70% и достигло значения 126 млн. (88% от общей численности населения).

По имеющимся прогнозам, по уровню инвестиций в информационные технологии в ближайшие годы будут лидировать производственный, финансовый и государственный секторы экономики России.

## **2. Развитие национальной информационно-коммуникационной инфраструктуры и производства современных ИКТ**

Серьезной проблемой институционального характера в России продолжает оставаться недостаточное развитие национальной информационно-коммуникационной инфраструктуры, а также ее высокая стоимость. Это существенно тормозит развитие в стране новых сетевых структур организации жизни общества и государства, которые являются необходимым условием социального прогресса и успешного движения к информационному обществу, увеличения спроса на информацию и знания, перехода к высоким и, прежде всего, компьютерным и коммуникационным технологиям, формирования в России конкурентного инновационного рынка.

Тем не менее, в последние годы в России продолжается *интенсивное внедрение современных информационных технологий в экономику, государственное управление и общественные процессы*, что является важнейшей составляющей структурных преобразований в экономике и реформы государственного управления.

Устойчиво растет спрос на внутреннем ИТ – рынке. В период с 2000 рынок информационных технологий в России вырос более чем в два раза и к концу 2004 г. достиг уровня 7,5 млрд. долл. США.

Однако, несмотря на высокие темпы роста, абсолютные объемы сектора ИТ в России остаются недостаточными (рынок ИТ составляет всего 1,4% от ВВП России, в то время, как в США объем сектора ИТ превышает 500 млрд. долл. и более 5% от ВВП).

При этом сохраняется сильная зависимость России от импортного аппаратно-технологического обеспечения (рынок отечественных ИТ-услуг составляет лишь 30% от общего объема, а рынок программных продуктов – 14%).

Доля России на международном рынке информационных технологий все еще не превышает 1%.

## **3. Информационные технологии и макроэкономическая среда российского предпринимательства**

В настоящее время в России уже сформировано достаточно представительное сообщество компаний и фирм, которые ведут профессиональную деятельность на рынке ИКТ и обслуживают все сегменты этого рынка.

Новым этапом в развитии российской индустрии стал 2005 год, когда по личной инициативе Президента России В.В. Путина в стране началось создание первых отечественных технопарков в сфере ИТ-технологий. Предполагается, что к 2010-2012 годам такие технопарки будут созданы в Дубне, Новосибирске, Санкт-Петербурге и Сарове и получат статус особых экономических зон.

Активное участие в создании технопарков принимает сегодня и отечественный бизнес. Так, например, его инвестиции в социальную инфраструктуру технопарка, создаваемого в Санкт-Петербурге, могут составить около 80%.

Однако, создание ИТ-технопарков выявило и ряд проблем. Одна из них – кадровая. Дело в том, что отставание системы образования от современных требований информатизации не обеспечивает подготовки достаточного количества квалифицированных пользователей ИТ и специалистов этой отрасли. Уже начала сказываться нехватка специалистов в области ИТ, особенно с учетом ее прогнозируемого многократного расширения. Так, например, оказалось, что в Новосибирске сегодня имеется острый дефицит квалифицированных программистов. По имеющимся оценкам, сегодня в области развития информационных технологий в России в особенности не хватает специалистов среднего звена, а также руководителей проектов.

Кроме того, в России сегодня также не развиты и механизмы *венчурного финансирования* инновационных проектов и это требует соответствующих изменений в законодательстве и налоговой политике. Неразвитость венчурного финансирования мешает появлению новых ИТ-компаний, развитию экспорта, тормозит внедрение инноваций.

Дорогая и недостаточно развитая информационная инфраструктура нашей страны сегодня тормозит развитие целых направлений бизнеса, например, аутсорсинга бизнес-процессов, и мешает привлечению в отрасль иностранных инвестиций. Так, например, стоимость выделенного канала связи в России сегодня в десятки раз превышает стоимость аналогичного канала в странах-конкурентах.

В настоящее время Мининформсвязи подготовлена и одобрена Правительством РФ *Концепция развития отрасли информационных технологий*, в результате реализации которой планируется к 2010 году достичь объемов производства порядка 30 млрд. долларов США. При этом занятость в отрасли составит 3,5 млн. человек, или около 5% работающего населения (для сравнения, эта занятость в 2001г. составляла лишь 1,4%).

#### **4. Готовность потребителей и российского бизнеса к вхождению в информационное общество**

По оценкам специалистов, сегодня основными потребителем ИКТ в России является государство (около **30%**), а также несколько крупнейших российских компаний (Газпром, Российские железные дороги, РАО ЕЭС, Аэрофлот, Связьинвест) – еще **30%**. Оставшийся объем спроса приходится на предприятия финансовой и нефтегазовой сферы, а также, в меньшей степени, на предприятия торговли и связи. Металлургия, машиностроение, транспорт и другие отрасли нашего хозяйства все еще значительно отстают в развитии и использовании ИКТ.

Необходимо отметить, что спрос компаний на ИКТ в России так же, как и в других странах мира, возрастает вместе с общим ростом управленческой культуры и улучшением бизнес-климата в стране. Поэтому, помимо мер, направленных на улучшение бизнес-климата, государство может увеличить спрос на ИКТ предприятий всех отраслей экономики, если будет способствовать:

- увеличению числа специалистов в области ИКТ;
- общему повышению компьютерной грамотности населения;
- повышению доступности современного оборудования за счет снижения импортных пошлин и уменьшения сроков амортизации компьютерной и другой информационной техники.

В качестве первоочередных мер, которые необходимо принять для решения данной проблемы, специалисты Министерства информационных технологий и связи РФ в разработанной ими Концепции развития рынка ИКТ рекомендуют следующие:

- введение ускоренных (до 2-х лет) норм амортизации компьютерной техники и программного обеспечения;
- применение ускоренных норм амортизации в случае передачи ПК в личное пользование сотрудникам, либо в качестве благотворительных пожертвований учебным заведениям;
- снижение налогооблагаемой базы налога на прибыль на сумму расходов предприятий по финансированию учебных заведений ИКТ – направленности, а также на обучение и переподготовку собственного персонала в области информационных технологий.

Со стороны населения России спрос на информационную технику и технологии также быстро растет, но все еще остается недостаточным, особенно, в регионах Сибири, Севера и Дальнего Востока. Проиллюстрировать это можно на примере повышения спроса на услуги мобильной связи, который в феврале 2005г. вырос вдвое по сравнению с 2004 годом и составил 56%. Как уже отмечалось выше, количество абонентов сотовой связи в России к концу 2005г. превысило 126 млн. (88%), в то время, как число пользователей ПК и сети Интернет составляет пока всего лишь 16-19 млн. (менее 14% от общей численности населения).

Однако, нужно отметить, что лавинообразный рост мобильной связи в России начался лишь после того, как ее клиентами стало более 20% населения. На этом рубеже сейчас находится и российская отрасль ИТ-технологий. По прогнозам специалистов Мининформсвязи, можно ожидать, что после того, как 20%-й барьер в России будет преодолен, возникнет лавина приобретений ПК и подключения к системе Интернет неквалифицированных пользователей. Поэтому сегодня необходимо всячески содействовать скорейшему преодолению этого барьера.



## **5. Развитие нормативной правовой базы в информационной сфере России**

По оценкам специалистов, существующая в России нормативная правовая база требует существенной корректировки и развития. В особенности это касается области развития информационных технологий и информационной экономики. Отсутствие законодательной базы для отрасли ИКТ в целом, а также для их отдельных приложений (например, электронный документооборот, электронно-цифровая подпись) существенным образом тормозит развитие как спроса, так и предложения информационных технологий.

Не разработанный механизм защиты прав на интеллектуальную собственность и обеспечения патентного права приводит к потере доходов российскими экспортерами и мешает привлечению в Россию крупных международных компаний. Отсутствие точных формулировок, касающихся ИКТ, в налоговом законодательстве приводит сегодня к неопределенности в вопросах начисления налогов, увеличению расходов и рисков компаний ИКТ-отрасли.

В разработанной Мининформсвязи РФ Концепции правового регулирования в сфере информационных технологий предусматривается внесение изменений в действующее законодательство (законы «Об информации, информатизации и защите информации», «О средствах массовой информации», «О рекламе»), а также принятие целого ряда новых законов («Об электронном документообороте», «Об информации персонального характера», «Об электронной торговле», «О праве на информацию», «Об участии в международном информационном обмене» и др.).

Кроме того, Институтом государства и права РАН уже подготовлена Концепция развития информационного законодательства в Российской Федерации [3], которая предусматривает системное развитие информационного законодательства в нашей стране, в том числе, и на основе принятия ряда новых Федеральных законов («Об экономике и организации сектора ИКТ России», «Об информационном взаимодействии федеральных и региональных органов власти», «О регулировании информационных ресурсов субъектов РФ» и др.).

## **6. Социокультурное развитие России и проблема информационной безопасности**

В Доктрине информационной безопасности Российской Федерации, утвержденной Президентом России в 2000 году [1], отмечается, что наибольшую опасность в сфере духовной жизни представляют следующие угрозы:

- Деформация системы массового информирования за счет монополизации СМИ, а также за счет неконтролируемого расширения сектора зарубежных СМИ в отечественном информационном пространстве;

- Неспособность современного гражданского общества России обеспечить формирование у порастающего поколения и поддержание в обществе общественно необходимых нравственных ценностей, патриотизма и ответственности за судьбу страны.

Именно эти угрозы остаются наиболее опасными и сегодня, так как за прошедший период действенных мер по их нейтрализации не принято, а соответствующие программы действий по защите национальных интересов страны и в области СМИ, и в сфере культуры в настоящее время отсутствуют [11]. Поэтому представляется необходимым специально рассмотреть эти проблемы на заседании Совета Безопасности РФ и выработать соответствующие рекомендации Правительству России.

### **7. Уровень распространенности информационных технологий среди населения России и проблема информационного неравенства**

Уровень распространенности информационных технологий среди населения России быстро растет, хотя по своим показателям он ещё очень далек от показателей развитых стран. Так, например, по числу семей, обладающих компьютерами (11 компьютеров на 100 семей), Россия далеко отстает не только от США и Западной Европы, но и от стран Восточной Европы, Балтии, а также от ряда развивающихся стран.

При этом наблюдается значительная разница в уровнях информатизации между Москвой и регионами. Так, например, в Москве до 40 семей из 100 имеют дома компьютеры, в то время, как в небольших городах и в сельской местности – менее 5 семей из 100.

Сегодня, даже в экономически благополучных регионах, многие мало обеспеченные семьи не могут позволить себе покупку персонального компьютера. Поэтому и дети, и взрослые в таких семьях не могут приобрести необходимых им базовых навыков компьютерной грамотности, что мешает и получению хорошего образования, и устройству на высокооплачиваемую работу.

Таким образом, наличие «цифрового неравенства» в России сегодня усиливает социальное расслоение общества и поэтому является одной из угроз для национальной безопасности страны. Действенной мерой для уменьшения этой угрозы могла бы стать государственная программа поддержки покупки населением персональных компьютеров. Программы такой поддержки существуют как в развитых странах (Великобритания, Италия, Швеция и др.), так и в развивающихся (Румыния, Египет). Поэтому и в России также необходимо разработать *программу поддержки малообеспеченных семей при покупке компьютеров*, используя для этого уже существующий зарубежный опыт.

Снижение импортных пошлин также может стать важным стимулом для повышения возможности населения покупать компьютеры. В настоящее время, пошлины на компьютеры и комплектующие изделия к ним составляют от 5 до 20%, что идет вразрез с международной практикой. Поскольку Россия практически не производит компьютерной техники, то эти пошлины не несут никаких защитных функций, а лишь повышают стоимость компьютера для конечного потребителя. Ликвидация пошлин на компьютеры и комплектующие согласуется с позицией ВТО и должна быть осуществлена в России в самое ближайшее время.

Важной мерой представляется также и *создание общедоступных компьютерных центров в регионах России*. Эта мера необходима для повышения компьютерной грамотности и спроса населения на информационную технику и услуги, а также как база для организации компьютерных курсов.

Кроме того, представляется остро необходимым дополнительное выделение средств из федерального и региональных бюджетов для ускорения реализации программы подключения школ к Интернету, в том числе с использованием современных средств доступа. По заявлениям Министра образования и науки РФ А. Фурсенко, процесс подключения всех российских школ к сети Интернет планируется завершить в 2007 году.

## **8. Информатизация общества и местное самоуправление**

В последнее время в социокультурной сфере России возникла еще одна угроза, которая обусловлена передачей финансирования городских и сельских библиотек на уровень местных, в том числе сельских, бюджетов. По информации с мест, это приведет к массовому закрытию сельских и городских библиотек, у которых уже начали отбирать здания и помещения. Существует серьезное опасение, что дальнейшее развитие этого процесса приведет к разрушению сложившейся в нашей стране библиотечной системы, обострению информационного неравенства между городским и сельским населением, а также к снижению уровня общей культуры в России.

Поэтому для того, чтобы избежать столь серьезных и пагубных для дальнейшего развития нашей страны последствий, необходимо безотлагательно внести соответствующие изменения в Федеральный Закон «Об общих принципах организации местного самоуправления в Российской Федерации», а также принять дополнительные меры по обеспечению деятельности сельских и городских библиотек и, в особенности, специализированных библиотек для детей и юношества. Представляется также необходимым существенным образом улучшить социальное положение работников этих библиотек, а также работников школьных библиотек, которые сегодня не имеют статуса преподавателей. Ведь в информационном обществе к этим работни-

кам предъявляются очень высокие профессиональные требования, соответствие которым потребует существенного повышения уровня их информационной компетентности.

## **9. Российская наука и образование на этапе формирования информационного общества**

По оценкам отечественных и зарубежных специалистов, современное состояние фундаментальной и прикладной российской науки в области информатики, развития информационной сферы общества и массовых коммуникаций еще не соответствует современным требованиям и новым вызовам 21-го века, а также имеющемуся в стране интеллектуальному потенциалу, который в настоящее время используется недостаточно эффективно.

В этой части особую тревогу вызывают негативные тенденции, обусловленные *старением научных кадров*, что делает все более реальной угрозу потери преемственности поколений в российской науке и образовании и утраты многих научных и научно-образовательных школ, на восстановление которых потребуются десятилетия [9].

Дальнейшее развитие этих тенденций представляет собой серьезную угрозу для национальной безопасности страны и должно быть решительно приостановлено. Одной из радикальных мер по решению данной проблемы могло бы стать придание статуса государственных служащих ученым и преподавателям государственных научных и образовательных учреждений, что существенным образом повысило бы их социальный статус в российском обществе.

*Недостаточно эффективно используется и интеллектуальный потенциал высшей школы России*, где сегодня работают более 170 тысяч кандидатов и докторов наук. Причина этого заключается не только в низком уровне оплаты труда научных работников, но также и в том, что они находятся в ВУЗах на положении второстепенных специалистов по отношению к основному профессорско-преподавательскому составу. Для устранения этого положения нужно всемерно содействовать расширению научной и инновационной деятельности ВУЗов и, в первую очередь, в информационной сфере.

Следующая мера заключается в разработке и реализации государственной политики России, направленной на привлечение талантливой молодежи в сферу науки и образования, закрепления их в этой сфере при помощи различных моральных и материальных стимулов, а также на повышение эффективности подготовки научных кадров и высококвалифицированных специалистов для развития информационной сферы общества и информационной экономики, основанной на знаниях.

Необходимо также существенным образом увеличить объемы финансирования фундаментальных и прикладных исследований в области информатики и проблем становления информационного обще-

ства. Причем, это финансирование должно осуществляться как на федеральном, так и на региональном уровнях, а также из внебюджетных фондов и средств бизнес-структур.

Основным научным фондам России (РФФИ и РГНФ) необходимо также существенно увеличить объемы финансирования по грантам, связанным с исследованиями информационных проблем развития российского общества, а также с подготовкой и изданием научной и научно-образовательной литературы по этим проблемам.

### **10. Информатизация образования и проблема становления информационного общества**

В области информатизации российского образования самым негативным фактором сегодня является фактический отказ Минобрнауки от признания данного направления в качестве приоритетного в процессе модернизации российского образования, как это определено Концепцией модернизации образования РФ на период до 2010 года.

В новой Программе развития образования, одобренной Правительством РФ в декабре 2004г., проблемы информатизации образования практически не рассматриваются, хотя дела в этой области обстоят далеко не лучшим образом. Так, например, результаты проведенного в 2004г. регионального социологического мониторинга в Ростовской области показали, что решение этой проблемы в России организовано неудовлетворительно и находится еще в самой начальной стадии.

Первым шагом здесь могло бы стать создание постоянно действующей национальной системы социологического мониторинга процессов модернизации и информатизации образования, методологические предпосылки для которой уже имеются. Это позволит повысить эффективность управления процессом информатизации образования и более рационально использовать финансовые, технические и другие ресурсы. Снижение темпов информатизации образования ведет к увеличению отставания России от передовых стран в этой области и снижению качества образования.

Определенную угрозу национальной безопасности страны представляет сегодня и низкий уровень общей и профессиональной подготовки педагогических кадров общеобразовательной школы в области информатики, который еще не соответствует современным требованиям. Для снижения уровня этой угрозы дальнейшему развитию нашей страны необходимо осуществить в ближайшие годы массовую переподготовку педагогов и решить наиболее острые проблемы оплаты их труда и социального обеспечения.

## **11. Проблемы России в области противодействия информационной экспансии и задачи государственной информационной политики**

По оценкам многих специалистов, сегодня Россия является ареной хорошо организованной и тщательно скоординированной перманентной информационной войны, которая ведется в соответствии с действующей доктриной информационной экспансии стран Запада и, в особенности, США. Управление и координация этой войной осуществляются специально созданными зарубежными инфраструктурами.

Области проведения активных информационных операций – это политика, экономика, промышленное производство, научные исследования и разработка технологий, информационно-технологическая среда систем управления, территориальные интересы и социальная обстановка, вооруженные силы, культура, история.

Информационная экспансия, как форма государственной информационной политики, реализуется сегодня как государственными институтами, так и негосударственными структурами западных стран. Эта экспансия сильно связана с экономической гегемонией промышленно развитых стран, а ее эффективность сопоставима с использованием промышленного и финансового капитала.

Главная цель информационной войны — экономическая и состоит в том, чтобы получить экономические преимущества за счет экономических потерь России. Для нейтрализации этой угрозы необходимо сделать информационную оборону нашей страны более активной и упреждающей. Она должна стать важнейшей частью государственной информационной политики России.

## **12. Международное сотрудничество в информационной сфере в целях поддержания мировой стабильности и международной безопасности**

Вхождение нашей страны в глобальное информационное общество невозможно без развития глобального международного сотрудничества в сфере развития ИКТ, информатизации государственного управления, науки, образования и культуры.

К сожалению, в настоящее время Россия еще слабо участвует в осуществлении международного сотрудничества в области нейтрализации угроз национальной и международной безопасности в информационной сфере. Необходимо также усиление сотрудничества с другими государствами, прежде всего, в направлении унификации правоотношений в информационной сфере, правовой фиксации общественных отношений, носящих прогрессивный и развивающий характер, и правового запрета на отношения, приносящие ущерб личности, обществу и государству.

В глобальном информационном обществе уже сложились устойчивые общественные отношения, особенностью которых является то, что они действуют на граждан разных государств одновременно. Такое положение требует согласованной правотворческой политики государств – участниц процесса формирования глобального информационного общества. Представители России должны более активно включиться в этот процесс – от этапа формирования концепции построения информационного общества, до этапа реализации норм права. При этом деятельность, которую в данной сфере сегодня осуществляют частные лица и общественные организации, должна не только находить поддержку и помощь со стороны государства, но и само государство должно активно включиться в эти процессы.

Государственные органы, на которые возложена обязанность представлять интересы России в международных организациях, должны занять более активную позицию в международных правотворческих процессах в сфере ИКТ, что будет содействовать сбалансированному развитию международных отношений и становлению глобального информационного общества. В первую очередь здесь представляется необходимым приведение отечественного законодательства в информационной сфере в соответствие с требованиями международных норм, а также уже существующих нормативных актов других стран – участниц глобального информационного общества.

Требование унификации законодательства в информационной сфере особенно важно в связи со вступлением России в ВТО. Однако, этот процесс не должен быть односторонним. Россия имеет полное право предлагать другим странам свой взгляд на нормы международного права и отстаивать свои национальные интересы. Она должна выступать инициатором равного партнерства на мировой арене всех государств, независимо от уровня развития их экономики.

Развитие международного права не должно осуществляться в ущерб национальной безопасности, суверенитету, культуре и самобытности государств-участниц глобального информационного общества. Только при этих условиях можно обеспечить необходимое соблюдение баланса международных и национальных интересов.

## ЛИТЕРАТУРА

1. Доктрина информационной безопасности Российской Федерации, утвержденная Президентом РФ 09.09.2000г., № Пр-1895.
2. Концепция развития отрасли информационных технологий РФ, разработанная Мининформсвязи 01.10.2004г.
3. Концепция развития информационного законодательства в Российской Федерации. // Государство и право, № 7, 2005. – С. 47-61.
5. Концепция содействия международному сотрудничеству и внешнеэкономической деятельности России в сфере ИКТ, в том числе, с учетом процесса присоединения к ВТО.
6. Федеральный закон «Об общих принципах организации местного самоуправления в Российской Федерации».
7. *Колин К.К.* Индустрия информационных технологий и проблемы информатизации российского образования. // *Alma mater* (Вестник высшей школы), № 4, 2005. – С. 12-16.
8. *Колин К.К.* Информационные технологии как импульс к развитию инновационной экономики. // *Человек и труд*, № 4, 2006.
9. *Колин К.К.* Инновационный прорыв России и проблема интеллектуальной безопасности. // *Открытое образование*, № 6, 2004. – С. 17-23.
10. *Колин К.К.* Преемственность поколений в области высоких технологий как гуманитарная проблема национальной безопасности России. // *Alma mater* (Вестник высшей школы), № 11, 2005.
11. *Колин К.К.* Нравственные приоритеты молодого поколения и будущее России. // *Молодые в библиотечном деле*, № 2-3, 2006.



---

## ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО: ДИАЛОГ КУЛЬТУР

*В.В. Миронов*

Становление информационного общества – неизбежный и необходимый процесс, оказывающий влияние на структурообразующие компоненты всей системы культуры. Аналогичных процессов в истории человеческого общества не существовало. Стремительность развития информационного общества и его масштабный характер одновременно выявляют и определенное отставание общественного сознания от тех изменений, которые уже произошли или происходят на наших глазах. В результате информационное общество рассматривается либо только как позитивное явление, либо, напротив, как фактор, негативно воздействующий на человеческую культуру. Попытаемся, не занимая крайних позиций, отметить лишь один аспект происходящего становления, связанный с втягиванием в информационное пространство человека и культуры в целом. По отношению к традиционной культуре – это новый процесс, от осмысления которого философия не должна уходить. Задача философии состоит в том, чтобы рационально оценить и спрогнозировать возможные последствия данного процесса для культуры в целом.

Возникновение глобального информационного пространства резко изменяет ситуацию обмена информацией между культурами. Точнее, оно выступает само по себе некой общей основой, в которую одновременно погружаются все локальные традиционные культуры, причем условием общения становятся уже законы данного информационного пространства. Информационное пространство в буквальном смысле вынуждает вести диалог между культурами по своим законам и правилам, которые не носят локального (однокультурного) характера. Культуры адаптируются теперь к единому информационному пространству как некому метаязыку, а не друг к другу в рамках межкультурного общения, как это было ранее. Культура растворяется в информационном пространстве.

Активность информационных процессов столь высока, что подчиняет себе традиционные элементы культуры и, прежде всего, изменяет традиционную систему культурной коммуникации. Традиционная система культур представляла собой совокупность замкнутых локальных образований или, в используемой нами терминологии, устойчивых и замкнутых информационных систем. Соответственно, диалог между такими культурами, иначе говоря, обмен информацией между ними осуществлялся на базе незначительных сходных устой-

чивых информационных компонентов, а основной массив информации должен был быть адаптирован (переведен, интерпретирован) на язык конкретной локальной культуры. Никакого промежуточного метаязыка для этой цели не существовало.

Локальная культура как информационная система обладала такими свойствами как стационарность и устойчивость. Динамика развития описываемой модели культуры была относительно медленной, она представляла собой, по отношению к другим явлениям, некоторую стационарную, застывшую систему, которая оставалась одинаковой для многих поколений. Признаком стабильности был самовыражением культуры как таковой. Она функционировала по своим определенным законам, эволюционным путем впитывая в себя и адаптируя новую информацию.

Относительная стабильность (слабая динамика по отношению к индивидуальному сознанию), аристократический принцип отбора, элитарность и замкнутость привели к выработке такого адаптивно-адаптирующего механизма, который позволял каждой локальной культуре относительно безболезненно и незаметно для индивида приспособливаться к себе новые компоненты, включая их в культурную систему. Длительность такого рода изменений в локальных культурах, как правило, выходила за рамки индивидуальной жизни, поэтому для отдельного человека они были незаметны. Действительно, находясь внутри локальной культуры или во временном отношении недалеко от нее, эти изменения обнаружить почти невозможно. Лишь сравнивая локальные культуры, отдаленные друг от друга большим временным расстоянием, мы легко обнаруживаем различия между ними, свидетельствующие о тех или иных изменениях.

Данный механизм адаптации можно рассматривать как своеобразный культурный иммунитет к иному, к традициям и ценностям другой культуры, препятствующий проникновению чужого влияния. Важной частью и проявлением указанного механизма адаптации было выделение в культуре «верхней» и «нижней» частей, которые находились в противоречивом взаимодействии, взаимодополняя друг друга. Абстрактно-духовная, рафинированная часть культуры постепенно оформляется в истории человеческой цивилизации как Культура с большой буквы. Она принципиально удалена от повседневности, даже от конкретной личности, и требует определенной подготовки при ее восприятии. Такое понимание культуры настолько типично, что может сложиться впечатление, что кроме нее никакой иной культуры не существует.

Вместе с тем в человеческой культуре всегда сохранялся пласт неэлитарной, повседневной, низовой культуры, в потребление продуктов которой всегда была включена наибольшая масса людей, народа.

Она в силу своей открытости была менее устойчива, а, следовательно, в большей степени подвержена изменениям.

Локальная культура, таким образом, предстает перед нами как некая завершенная целостная символическая система культурных значений, отражающая завершенность бытия человека и человечества. Точно также завершенным считается и противопоставление в ней «верхней» и «нижней» частей, в котором господствующей культурой как таковой является именно «верхняя».

Поэтому в целом тезис о единой общечеловеческой культуре как целостной системе был в этот период неправомерен и представлял собой скорее лишь метафору. Не существовало целостной культуры, а имелась **система локальных культур**, отдаленных друг от друга даже пространственно. Это не означает, что такие локальные культуры не взаимодействовали между собой, не знали друг о друге. Общение, конечно, было, но в рамках временного семиотического диалога-пространства. Одновременно каждая из культур вырабатывала в себе мощнейший каркас, некий «иммунитет» к другой культуре, не пропускающий чуждых элементов и влияний.

Две культуры соотносились как два языковых пространства, и общение между ними выступало как особое коммуникационное пространство, которое часто обозначают как диалог культур. Это был относительно локализованный диалог, так как пространство пересечения культур было невелико, а область непересекающегося огромна.

Диалог культур – важнейший принцип окультуривания личности, так как познание своей культуры здесь совершается через познание иной, другой культуры в рамках семиотического пространства сравнения. Именно познание области несовпадения (изначального непонимания) культур обогащает их новыми смыслами и новыми ценностями, хотя и затрудняет сам факт общения и, в конечном счете, де-лает культуру непознаваемой для другой в абсолютном смысле.

Главным средством диалога выступает язык, знание которого является важнейшей предпосылкой понимания другой культуры. Зная иную язык, я необходимо адаптирую (перевожу) смыслы другой культуры. Сопоставляя же иную и свою культуры, я необходимым образом понимаю ценность и своеобразие собственной культуры. «Чтение иностранной литературы неизбежно сопровождается и знакомством с чужой, **иной страны** культурой, и конфликтом с ней. В процессе этого конфликта человек начинает глубже осознавать свою собственную культуру, свое мировоззрение, свой подход к жизни и к людям»<sup>1</sup>.

Данная мысль представляет, по сути, одну из формулировок принципа развития, в данном конкретном случае – развития лично-

---

<sup>1</sup> Тер-Минасова С.Г. Языки межкультурной коммуникации. – М., 200. –С.24.

сти. Действительно, конфликт – это форма противоречия, а противоречие, точнее, его разрешение является главным источником развития. Поскольку речь в данном случае идет о языке, носителем которого является индивид как представитель отдельной культуры, то саморазвитие индивида необходимо связано с разрешением данной формы противоречия – противоречия культур. Именно это, в свою очередь, не только обогащает наше понимание, но и взаимообогащает понимание культурами друг друга. Поэтому *язык* является не просто средством коммуникации или передачи информации, а представляет собой *важнейший механизм культурного общения*.

Диалог локальных культур осуществляется как адаптация одной культуры к другой посредством интерпретации смыслов, понимания стереотипов и традиций иной культуры. Для человека, познающего иную культуру, она предстает, прежде всего, и главным образом, как иной язык. При этом для диалога культур более важной является область их несовпадения, требующая адаптации, приспособления, перевода. Полное совпадение культур исключило бы возможность диалога. Диалог культур, таким образом, является пониманием, осознанием своей культуры через другую, а иной культуры через свою путем культурной интерпретации и адаптации культур друг к другу в условиях смыслового несовпадения и, нередко, конфликта.

Таким образом, если семиотически интерпретировать диалог культур, то он реализуется, по удачному обозначению Лотмана, внутри *«семиосферы как особого семиотического пространства, включающего в себя не только сумму отдельных языков, но и социокультурное поле их функционирования»*. Любой отдельный язык оказывается погруженным в некоторое семиотическое пространство, и только в силу взаимодействия с этим пространством он способен функционировать... Семиосфера – и результат, и условие развития культуры». Поэтому *диалог культур – это смысловая адаптация их друг к другу*.

Современное состояние культуры фиксирует стадию ее перехода от локального уровня к интеграционному. Глубина происходящих в культуре изменений побуждает рассматривать данный процесс как *кризис культуры, означающий сегодня ситуацию резкого изменения коммуникационного пространства, все более разрывающего границы между культурами и, вместе с тем, создающего предпосылки (новые культурообразующие компоненты), которые могут обеспечить связь культуры в единую систему*.

В условиях развития информационного общества возрастает объективная возможность интеграции культуры. Вместе с тем, новое формирующееся информационное пространство навязывает нам диалог, основанный не на различии культур, а на схожести, что существенно меняет характер диалога. Если раньше в его основе лежал смысл культуры, а расшифровка этого смысла выступала для другой

культуры средством адаптации, то сегодня система коммуникации диктует собственные правила диалога между культурами. Мир начинает говорить на языке господствующих в нем стран. В результате расширяется поле «псевдокультурного» общения, диалог в котором осуществляется в пределах наиболее доступных, совпадающих или почти совпадающих смысловых структур, являющихся наименее содержательной частью культуры. В данном коммуникационном поле господствуют общие стереотипы, оценки, нормы поведения, наиболее общедоступные и простые компоненты культуры. Это удобно, в определенном смысле, но вместе с тем лишает диалог важнейшего смысла и содержания. Он перестает играть роль фактора, способствующего взаимообогащению культур и обеспечению процессов их самоидентификации (своя культура познается через иную). Мы можем понять любого человека в любой точке Земли, но лишь на уровне совпадения или даже тождественности смыслов. Это общение без насыщения смыслами, по существу, общение со своим зеркальным отображением по заданным стереотипам коммуникации.

В новых условиях происходит разрушение старых систем ценностей и традиций, господствовавших несколько столетий, а возникновение новых происходит такими быстрыми темпами, что человек не успевает, как это было ранее, адаптироваться к изменениям, соотнести новые ценности и предыдущие. В результате ценности начинают просто потребляться. На семантическом уровне данный процесс проявляется, например, как быстрое «наполнение» языка новыми словами и оборотами, понятными молодым людям, но уже не доступными людям даже среднего поколения.

Второй смысл, которое я вкладываю в понятие «кризис» культуры, фиксирует резкое увеличение скорости разрушения старых ценностей, сжатие временных рамок этого процесса, что не позволяет новым символам и знакам адаптироваться к традиционной знаковой системе ценностей.

Действительно, с одной стороны, происходит резкое увеличение образований, претендующих на статус культурных, а с другой – их адаптация к старым ценностным системам происходит как бы в более сжатых временных рамках. Порой это уже можно наблюдать в период жизни одного человека или даже еще быстрее. В результате происходит разрушение старых систем ценностей и традиций, разрушение целостной знаковой системы культуры, которая господствовала на протяжении столетий. Символы и образы старой культуры исчезают или меняют свой смысл и значение. Новые ценности настолько расходятся с традиционными, что их культуuroобразующий смысл остается не всегда ясным и открытым для человека. Число людей, воспринимающих культурные новообразования, возрастает, но

это восприятие лишается той утонченности и глубины, той степени подготовки, которой оно требовало ранее.

Изменение средств коммуникации нарушило естественный баланс между высокой и низкой культурой, существовавший в традиционной культуре. В результате низовая культура стала выступать как превращенная форма официальной культуры. *«Низовая», массовая культура начинает доминировать, в каком-то смысле, по крайней мере, временно, подавляя «высокую».* Это третье значение, которое я вкладываю в понятие «кризис культуры».

Современная массовая культура, представляющая собой, по сути, типичную «низовую» культуру, все более приобретает черты шоу, для которого характерно не индивидуальное творческое участие, а участие само по себе без передачи какого-либо смысла. Возникает вопрос, насколько данная ситуация уникальна и типична для нашего времени? Существовали ли в рамках локальной культуры похожие явления? Безусловно, да. Это карнавалы. Но различие состоит в том, что в современном обществе такие карнавалы становятся образом жизни многих, карнавал длится не неделю, не месяц, а практически постоянно. Образцы «низовой», массовой культуры начинают доминировать, господствовать в культуре. Возникает своеобразная поп-культура.

Философия как часть культуры, ее душа, самосознание неизбежно оказывается втянутой в этот процесс. Так, разрушение классической философии и появление постмодернистских философских концепций является реализацией современной культурной ситуации внутри философии. В ней тоже начинает доминировать «низовая» культура. Постмодернизм чутко уловил процессы фрагментации современной культуры и противоречивые процессы, сопровождающие разрушение классических образцов, и стал своеобразным карнавальным зеркалом классической философии. Традиционные сюжеты постмодернизма о смерти автора, о бесконечности текста реализовались в семиотическом пространстве интернета. Но одновременно человек стал более свободным в творчестве и мыслеизъявлении, ведь нередко ему проще дать свое собственное объяснение тем или иным феноменам, не углубляясь в изучение их сущности, чем опереться на доказанные и проверенные истины.

Таким образом, революционное развитие информационных и коммуникационных технологий создало невиданные ранее возможности для взаимодействия культур, разрушая одновременно старую систему локальных культур. Возникает вопрос, возможно ли создание глобального коммуникационного пространства на основе гуманистических принципов? Какие потенциальные возможности для роста культурного разнообразия содержит формирующееся информационное пространство? Ответы на эти и многие другие непростые

вопросы, связанные с развитием информационного общества, призвана дать современная философия.

#### ЛИТЕРАТУРА

1. *Бахтин М.М.* Франсуа Рабле и народная культура средневековья и Ренессанса. – М.: 1990.
2. *Визель М.* Гипертексты по ту и эту стороны экрана // Иностранная литература: Электронная версия. – 1999. -№10. [http: //novosti.online.ru/magazin/inostran/n10-99/ vizel.htm](http://novosti.online.ru/magazin/inostran/n10-99/vizel.htm).
3. *Ионин Л.Г.* Социология культуры. – М.:1998.
4. *Кнабе Г.С.* Двудеинство культуры // Материалы к лекциям по общей теории культуры и культуре античного Рима. – М.: 1993.
5. *Лотман Ю.М.* Культура и взрыв. – М.:1992.

## АНТРОПОЛОГИЧЕСКИЙ ПОВОРОТ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ\*

*Е.А. Никитина*

Современная ситуация в когнитивных науках в целом, и в методологии искусственного интеллекта, в частности, во многом напоминает гуманитарный, антропологический поворот, который осуществился в неклассической и постнеклассической эпистемологии. Классическая когнитивистская парадигма уступает место неклассическим тенденциям, и аргументы, возражения сторонников перемен во многом напоминают критику положений классической теории познания.

В эпистемологии подвергаются критике такие установки классической новоевропейской теории познания как субъектоцентризм, полагающий непреложным факт существования субъекта познания, картезианская дихотомия «субъект – объект», наукоцентризм (установка, в соответствии с которой научное знание, точнее – математическое естествознание является высшим типом знания) и фундаментализм (требование выведения знания из единого основания). Наиболее развернутый и обстоятельный анализ тенденций неклассической и постнеклассической эпистемологии содержится в трудах В.А.Лекторского<sup>1</sup>.

Один из объектов критики – классическая трактовка субъекта познания как абстрактного «абсолютного наблюдателя», нацеленного на получение объективной истины (максимально объективного знания об объекте) и постоянно избегающегося от своей субъективности, препятствующей достижению цели. Такое понимание субъекта сложилось в классической европейской философии XVII – XVIII вв. под влиянием формирующегося естественнонаучного идеала знания и соответствующего типа научной рациональности<sup>2</sup>.

Не рассматривая подробно традицию критики классической субъектной парадигмы в истории философии и современной философии<sup>3</sup>, отметим главные возражения критиков (философия жизни и др.). Прежде всего, несогласие вызывал абстрактный, внеисторический характер рационалистического субъекта и классической гносеологии в це-

---

\* Работа выполнена при финансовой поддержке РГНФ. Проект 05-03-03134а

<sup>1</sup> См. об этом *Лекторский В.А.* Эпистемология классическая и неклассическая. – М.: 2001.

<sup>2</sup> *Степин В.С.* Теоретическое знание. – М., 2000, с.633.

<sup>3</sup> Существует широкий круг литературы по данной теме: труды Лекторского В.А., Степина В.С., Гайденоко П.П. и др.



лом. Гносеология должна исходить из целостной, спонтанной жизненной активности человека, включенного в реальный мир и систему отношений с другими субъектами. Так, например, В. Дильтей считал, что науки о духе должны преодолеть метафизическое представление о гносеологическом субъекте и включить в науку о познании ценностные установки, условия и способы деятельности. Дильтей, мысливший историческую действительность как континуальность знания и действия, считал, что такой подход позволит преодолеть теорию репрезентации в гносеологии, возникшую вследствие картезианского отделения внутреннего мира человека от внешнего мира.

Постепенно в русле критики классической теории познания в эпистемологии происходит своеобразный антропологический поворот к проблематике сложной структуры субъекта, роли телесности в познании, влияния ценностей и целей человека на познание.

Так, в современной отечественной эпистемологии сформировался экзистенциально-антропологический подход к субъекту познания, главной установкой которого является принцип доверия субъекту как целостному познающему человеку, «ответственно поступающему в получении истинного знания и в преодолении заблуждений»<sup>4</sup>. Основанием для принятия принципа доверия субъекту служат данные эволюционной эпистемологии (Лоренц, Фоллмер, Кезин, Меркулов), согласно которым субъективные структуры познания объективно соотношены с реальным миром, так как сформировались в процессе эволюции. Принцип доверия субъекту познания включает в себя требование исходить в философском осмыслении познания из живого, непосредственного, реального познания, учитывать всю полноту жизненного опыта и данные непосредственного наблюдения, личностные особенности познающего человека. В противном случае за порогом исследования останется сфера иррационального, хаотичного, непосредственного «живого» психологического опыта человека и человечества, недоступного рационализированному сознанию, мыслящему в пределах опозиции «субъект-объект».

Еще одна важная методологическая установка данного подхода состоит в признании равноценности рациональных и иррациональных познавательных способностей человека, что позволяет выявить сложную структуру субъекта и отойти от его абстрактного понимания. Хотелось бы в этой связи вспомнить, что немало схожих мыслей было высказано в онтологической гносеологии (Флоренский, Лосский, Франк и др), развивавшейся в рамках русской религиозной философии конца XIX – нач. XX вв., которая, к сожалению, недостаточно включена в контекст современных эпистемологических рассу-

---

<sup>4</sup> Микешина Л.А. Философия познания. Полемиические главы. – М.: Прогресс-Традиция, 2002, с.159.

ждений. Интересна мысль Флоренского, прозвучавшая в плане критики рационализма западной философии, о том, что для восстановления полноты взаимодействия человека с реальностью нужно восстановить равноправие различных средств познания. Множественности форм бытия не может не соответствовать множественность форм познания.

Обратимся к ситуации в методологии искусственного интеллекта. Ряд исследователей отмечают, что в последнее время в когнитивной науке «происходит концептуальный поворот, смена основного подхода к исследованию: от *вычислительного подхода* (computational approach) к *динамическому* (dynamical approach)»<sup>5</sup>. Он связан с именами таких ученых как Варела, Бир, Брукс, Чалмерс и др). Известно, что вычислительный подход – это основа методологии искусственного интеллекта. При этом аргументы критиков вычислительного подхода во многом похожи на аргументы критиков классической субъектной парадигмы.

В чем состоят возражения сторонникам вычислительного подхода?

1. Прежде всего, в том, что в рамках данного подхода все многообразие когнитивных функций человека сводится к интеллектуальным. При этом интеллект рассматривается абстрактно, вне процесса его естественного функционирования, вне связи с телесностью человека и вне ситуации взаимодействия человека со средой. Соответственно, подчеркивают критики, модель лишается связи с реальностью и объяснительной силы.

Это уже традиционное и привычное возражение. Но, строго говоря, это лишь констатация ограниченных возможностей точных методов в познании и воспроизведении живого. Формализации поддаются далеко не все познавательные функции человека. А вычислительный подход ограничен в той степени, в какой ограничено моделирование как один из методов научного познания, связанный с отвлечением, абстрагированием от каких-либо конкретных свойств объекта в соответствии с целями исследования. Причем в большинстве случаев ученые, обращаясь к математическому и компьютерному моделированию, решают проблемы в пределах своей предметной области, а не в области моделирования действительных познавательных функций человека. Например, идея квантового компьютера появляется в связи с задачами моделирования тех физических систем, где классическая система вычисления не «справлялась». И лишь затем проводятся аналогии между функционированием вычислительной машины и мозга, а сознание мыслится, исходя из компьютерной

---

<sup>5</sup> Князева Е.Н. Концепция инактивированного познания: исторические предпосылки и перспективы развития // Эволюция. Мышление. Сознание. (Когнитивный подход и эпистемология) – М.: Канон+, 2004, с.308.

метафоры. В этом случае естественное функционирование мозга лишается ряда когнитивных функций, они исключаются из рассмотрения как не поддающиеся формализации.

Более точным объектом критики в этом случае являются метатеоретические установки и ценности исследователей, редуцирующих сознание и познание человека к вычислительным процессам. Так, например Х.Патнэм обращает внимание на то, что идея теоретической редукции «всей психологии, имплицитно присутствующей в нашей повседневной практике атрибутирования ментальных состояний, к вычислительной психологии – бессодержательна»<sup>6</sup>. Он называет химерой, а не научным проектом замысел отождествления ментальных состояний с физическими или вычислительными.

Патнэм пишет о двух разновидностях данного проекта. Первая, более простая, состоит в отождествлении интенциональности наших мыслей (своеобразного внутреннего письма) с их каузальной ковариацией. По сути, речь здесь идет о сильной версии искусственного интеллекта, которая допускает, что вычислительные устройства, запрограммированные соответствующим образом, могут обладать разумом и способны, как и человек, понимать. Данный замысел, как правило, вызывает серьезные возражения, и, по его мнению, неосуществим: он разрушается множеством контрпримеров. Широко известен мысленный эксперимент Д.Серла «китайская комната», предложенный для опровержения данной позиции.

Более сложной, рассуждает Патнэм, была «моя собственная идея «функционализма», подразумевавшая отыскание таких вычислительных состояний, которые возможно было бы отождествить с существующими нам пропозициональными установками. При этом Патнэм оговаривается, что на тот момент полагал, что необходимое понятие вычислительного состояния точно было определено имевшимся тогда формализмом вычислительной теории (формализмом Тьюринга или теорией автоматов).

Патнэм пишет, что «формализм вычислительной теории *имплицитно определяет* каждое вычислительное состояние (причем всякое) тотальностью его вычислительных отношений (к примеру, отношений последовательности или вероятностной последовательности) ко всем другим состояниям данной системы. Иными словами, *имплицитно происходит одновременное определение* всего набора вычислительных состояний данной системы; и это имплицитное определение *индивидуализирует* каждое из состояний, в смысле отличия

---

<sup>6</sup> Патнэм Х. Химера когнитивной науки // Философия и будущее цивилизации: Тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т. 1. – М.: Современные тетради, 2005, с.277.

чения его от всех других вычислительных состояний»<sup>7</sup>. Но ни одна, подчеркивает он, реальная психологическая теория никогда не претендовала на то, чтобы представить законы, в соответствии с которыми можно было бы отличить состояние ревности по поводу мнимой любви Дездемоны к Кассио от всякой другой актуальной или возможной пропозициональной установки<sup>8</sup>. Таким образом, Патнэм считает идею формализации всех возможных психологических состояний и процессов утопической.

Тема принципиальных ограничений при моделировании естественных познавательных функций человека – одна из традиционных тем в философии и методологии искусственного интеллекта. В частности, с рядом трудностей сталкивается разработка интеллектуальных технологий и систем для образовательного процесса. Например, создание экспертных систем обучения и контроля знаний связано с изучением структуры экспертного знания и механизмов его функционирования, при этом возникает проблема с вербализацией неявного, личностного знания экспертов. Другая проблема состоит в том, как учесть при проектировании компьютерных обучающих комплексов индивидуальные особенности их будущих пользователей и др.

Таким образом, для ряда направлений искусственного интеллекта актуальной является задача исследования и понимания естественного интеллекта, который, помимо когнитивных функций, «включает широкий спектр других психических модальностей, таких как эмоции, воображение, желание, воля, саморефлексия, экзистенциальные состояния и др.»<sup>9</sup>

Вернемся в философскую сферу. Как представляется, можно провести параллели между современной критикой вычислительного подхода и критикой рационалистического субъекта в классической новоевропейской философии. Так, познавательная установка опытного математического естествознания (Галилей, Гюйгенс) тоже была основана на своеобразном «диктате» объекта и методологии его познания по отношению к субъекту. Реальность, данная в чувственном опыте, постигается на основе искусственного конструирования в эксперименте того, что исследуется, а это предполагает использование математики.

Соответственно, сформировалось требование к субъекту постоянно сверять полученное знание с тем, которое считалось нормой,

---

<sup>7</sup> Там же, с.276.

<sup>8</sup> Там же, с.277.

<sup>9</sup> Дубровский Д.И. Искусственный интеллект и проблема сознания // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛЛ, 2005, с.42.

соответствовало определенным критериям – эмпирической проверяемости, непротиворечивости, воспроизводимости, объективности и т.д. В результате, субъект познания мыслился как сугубо рационалистический субъект.

2. Обратимся еще к одному возражению сторонникам вычислительного подхода, состоящему в том, что в модели познания, интеллектуальные функции представлены как завершенные, готовые. Между тем, необходимо учитывать эволюционное происхождение данных функций (филогенез) и особенности их формирования в индивидуальном развитии человека (онтогенез).

В этом отношении обширный материал для методологии искусственного интеллекта содержится в эволюционной эпистемологии, исследующей биологические предпосылки познания на основе современной синтетической теории эволюции, конкретно-научных данных из области психологии, антропологии, этологии, психологии, нейрофизиологии, лингвистики и др. Эволюционная эпистемология или эволюционная теория познания является своеобразным «полем» междисциплинарного синтеза, новым междисциплинарным направлением.

К.Лоренц, австрийский этолог, лауреат Нобелевской премии, развивая во второй половине XX в. эволюционную теорию познания, исходил из необходимости научной переориентации философской теории познания. Он полагал, что умозрительное философское изучение познания исчерпало свой объяснительный потенциал. (Заметим, что в среде ученых-естествоиспытателей, изучающих познание, довольно широко распространен скептицизм в отношении возможностей философского изучения познания и выражено желание поставить эпистемологию «на твердую почву»). Успехи естественных наук позволяют, по мнению Лоренца, привлечь для изучения познания естественнонаучную и общенаучную методологию. Опираясь на нее, ученый исследовал формирование аппарата познания (нервную систему и мозг) вместе с поведением человека и животных в филогенезе. В ходе исследований было показано, что субъективные познавательные структуры человека сформировались в ходе эволюции. Причем закрепились те из них, которые в наибольшей степени соответствовали условиям жизни и способствовали выживанию. Это означает, что познавательные способности человека объективно соотносены с реальностью, а истоки абстрактного и символического мышления, самосознания содержатся в филогенезе.

Кстати, заметим, что данные эволюционной эпистемологии о том, что рациональное начало в субъекте познания эволюционно обусловлено, в определенной мере разрушают объект критики гуманитарно-ориентированной эпистемологии, а именно – образ рационально сконструированного, абстрактного субъекта классической новоевропейской философии. Ведь искусственно созданный (под влиянием

естественнонаучного идеала знания) рационалистический субъект на проверку оказывается «естественным» субъектом, так как истоки абстрактного мышления коренятся в самой природе человека. В таком случае, напряженность методологического противостояния гуманитарного и естественнонаучного способов изучения субъекта познания снимается. Становится очевидной их взаимодополнительность.

В трудах другого известного современного представителя эволюционной эпистемологии – немецкого ученого и философа Г.Фоллмера – также развивается мысль о том, что формирование когнитивных структур происходит в процессе эволюции. Из данной посылки, отмечает Фоллмер, вытекает ряд следствий<sup>10</sup>. Первое следствие – существование онтогенетического (а не только филогенетического) развития познавательных способностей человека – является для психологов и педагогов, в общем-то, очевидным фактом. В психологии процесс интеллектуального созревания ребенка в соответствии с некоторыми генетически обусловленными факторами исследовался в работах Ж. Пиаже (онтогенетическая эволюция ментальных структур).

Второе следствие, выводимое Фоллмером, – связь эволюции познавательных способностей с историческим развитием человеческого познания: с историей понятий, идей, науки.

И, в-третьих, из исходных посылок эволюционной эпистемологии следует существование связи между эволюцией познавательных способностей человека, когнитивным развитием индивида и эволюцией науки. Фоллмер подчеркивает в своих трудах, что, если исходить из методологической позиции «гипотетического реализма», сформулированной в эволюционной эпистемологии, то такие связи и общие структуры можно найти<sup>11</sup>.

В соответствии с гипотетическим реализмом познание носит гипотетический характер. Одновременно постулируется существование независимого от сознания, закономерно структурированного мира, который можно частично познать посредством восприятия, мышления и интересубъективной науки. Соответственно возникает задача сопоставления исторического формирования науки как интеллектуального инструментария с филогенезом и онтогенезом познавательных способностей человека. В частности, интерес представляет изучение соотношения гипотетико-дедуктивного метода построения научной теории и метода проб и ошибок в индивидуальном познании

---

<sup>10</sup> Фоллмер Г. Эволюционная теория познания: врожденные структуры познания в контексте биологии, психологии, лингвистики, философии и теории науки (Перевод с немецкого) – М.: Русский Двор, 1998, с.206-207

<sup>11</sup> Там же, с.217-222.

человека, выявление степени влияния биологических и психологических факторов на возникновение и рост научного знания и т.п.

Данные связи и проблемы, обращает внимание Фоллмер, изучены слабо, и не в последнюю очередь потому, что исследовательские задачи носят междисциплинарный характер. Ученый должен обладать знаниями в области эволюционной теории и психологии, истории и теории науки, философии и методологии науки, что встречается нечасто.

Современные российские философы (Меркулов, Кезин и др.), развивающие идеи эволюционной эпистемологии, также исходят из того, что «когнитивная эволюция – это один из аспектов биологической эволюции, тесно связанный с другим ее аспектом – с эволюцией поведения»<sup>12</sup>. Более того, И.П.Меркулов отмечает, что в настоящее время необходимо исходить из факта продолжающейся когнитивной эволюции. А это существенным образом влияет на наши представления об эволюции познания, человека и о факторах, влияющих на социальное и культурное развитие<sup>13</sup>.

Проблема субъекта познания в рамках эволюционной эпистемологии также рассматривается в новых аспектах: прослеживается эволюционное формирование познавательных способностей субъекта, онтогенетическая динамика субъекта познания, дается естественнонаучная интерпретация активности субъекта познания, изучается процессуальная сторона познания и т.д.

Идеи и данные эволюционной эпистемологии учитываются в классических методах и моделях искусственного интеллекта. В них вводятся, как отмечают исследователи, элементы эволюции, самоорганизации, динамики, обучения. Имеется в виду «построение открытых динамических баз знаний в русле прикладной семиотики, эволюционных вычислений, вычислений со словами и перцептивными образами Л.Заде, модифицируемых рассуждений, вопросов синтеза познавательных процедур с использованием комбинированных и многозначных логик, моделей эволюционной семиотики, вопросы происхождения логики на базе изучения биологической эволюции»<sup>14</sup>.

Так, например, в рамках бионического подхода к разработке систем искусственного интеллекта конструируются и исследуются искусственные «организмы» (в виде компьютерной программы или робота), способные приспосабливаться к внешней среде («анимать»).

---

<sup>12</sup> Меркулов И.П. Введение: формирование когнитивных представлений в эпистемологии // Эволюция. Мышление. Сознание. (Когнитивный подход и эпистемология) – М.: Канон+, 2004, с.7.

<sup>13</sup> Там же, с.11-12.

<sup>14</sup> Тарасов В.Б. Синергетический искусственный интеллект // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛЛ, 2005, с.232.

Известный специалист в данной области В.Г.Редько пишет: «Почему ... результаты, получаемые математиком, применимы к реальной природе? Или в общей формулировке: почему наше логическое мышление применимо к познанию природы?» И далее: «Можем ли мы думать об эволюционных корнях логических правил, используемых в математике?»<sup>15</sup> Наиболее четкий путь исследования этой проблемы, отмечает автор, – «построение математических и компьютерных моделей «интеллектуальных изобретений» биологической эволюции, таких, как безусловный рефлекс, привыкание (угасание реакции на биологически нейтральный стимул), условный рефлекс, цепи рефлексов..., логика. То есть целесообразно с помощью моделей представить общую картину когнитивной эволюции: эволюции когнитивных способностей животных и эволюционного происхождения интеллекта человека»<sup>16</sup>.

3. Еще одна группа критических замечаний в адрес классического вычислительного подхода в когнитивных науках касается того, что в нем не учитывался процессуальный, динамический и интерактивный характер познания, влияние социокультурного контекста познания, активность субъекта познания. Важным свойством когнитивных систем является то, что они появляются спонтанно в процессе самоорганизации.

Рассматривая изменения, происходящие в последнее время в методологии искусственного интеллекта, продолжим параллели с неклассическими и постнеклассическими тенденциями в философской теории познания. В последние десятилетия внимание философов привлекает процессуальная сторона познания, его интерпретационные и символические аспекты (коммуникативное понимание субъекта и др.). Коммуникация полагается универсальным условием человеческого бытия и познания, что побуждает рассматривать субъект-объектные отношения через призму общения, диалога субъектов, в контексте условий коммуникации. В значительной мере это связано с развитием информационного общества, возрастанием роли информационных и коммуникационных технологий в жизни человека. Возрастает количество исследований, посвященных изучению коллективного характера получения знания, выявлению социокультурной обусловленности познания и исторически изменчивых норм познавательной деятельности и т.д. Интересные исследования, в которых конкретизируется процесс познания, осуществляются в психологии, социальной психологии, психологии социального познания.

---

<sup>15</sup> Редько В.Г. Задача моделирования когнитивной эволюции // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛЛ, 2005., с.224.

<sup>16</sup> Там же, с.224.



Возрастает интерес к таким концепциям, как символический интеракционизм (Мид, Блумер) этнометодология (Гарфинкель), социальная драматургия (Гофман) и др. Символический интеракционизм изучает социальные взаимодействия преимущественно в их символическом содержании. При этом социальный процесс понимается как выработка и изменение социальных значений, постоянное определение и переопределение ситуаций взаимодействия их участниками. Различные группы создают, конструируют различные миры: эти миры меняются, когда объекты, представляющие их, меняют свои значения. В этнометодологии, развивающейся как методология социальных наук, социальная реальность также формируется как результат интерпретационной активности взаимодействующих людей. При этом в процессе интерпретации выявляются неосознаваемые, скрытые, неявные смыслы и значения, механизмы коммуникации.

Соответственно, похожие процессы происходят в разработке проблем искусственного интеллекта. Так, отмечается, что в ближайшие годы в искусственном интеллекте и информатике будут интенсивно развиваться такие направления как самоорганизующийся искусственный интеллект, синергетическая информатика<sup>17</sup>. При этом методология синергетического искусственного интеллекта претерпит существенные изменения. В философско-методологическом плане это переход от рационалистической методологии к синергетической. Так, например, природа интеллекта в синергетическом искусственном интеллекте понимается как коллективная, распределенная, тогда как в классическом она понималась как индивидуальная, сосредоточенная. В области архитектуры компьютера происходит переход к интеграции различных информационных технологий, что позволяет обеспечить компенсацию недостатков одних технологий и объединить преимущества дополняющих друг друга технологий.

«Область *синергетического ИИ* включает, в первую очередь, исследования процессов *формирования, деятельности, коммуникации, эволюции и кооперации* искусственных систем. Прежде всего, изучаются неустойчивые состояния, динамика, взаимные переходы, способы создания и разрушения сложных интеллектуальных систем. Примерами синергетических искусственных систем служат: гибридные интеллектуальные системы, системы «мягких вычислений» (Soft Computing), распределенные системы управления, интеллектуальные агенты, многоагентные системы, виртуальные организации, эволю-

---

<sup>17</sup> Тарасов В.Б. Синергетический искусственный интеллект // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛ, 2005, с.230-232.

ционирующие искусственные сообщества и пр.»<sup>18</sup>, – отмечает В.Б.Тарасов.

Изменения, происходящие в методологии искусственного интеллекта, по сути, представляют собой, как это ни парадоксально, антропологический поворот, последствия которого пока еще сложно определить. Интересно, что Лоренц видел в абстрактном мышлении человека опасность для дальнейшего выживания человечества в значительно измененной среде<sup>19</sup>. Опасность проистекает из связи процессов абстрагирования и символизации с агрессивным началом в человеке. Проблема состоит в том, что усиливается противоречивость между быстрым ростом массива надындивидуальных знаний, которые человек производит с помощью абстрактного и символического мышления, их внегенетической трансляцией и возможностями человеческой природы. По мнению Лоренца, понятийное мышление делает неконтролируемыми внутривидовой отбор и агрессивность, которые в новой быстрой меняющейся среде не могут изменяться так же быстро.

Антропологический поворот расширяет сферу применения технологий искусственного интеллекта, среда приобретает новое качество. Уже сейчас реальность все более предстает человеку своей технологически и лингвистически сконструированной стороной. Технологический подход распространяется на все сферы жизни: социальные, политические и др. В философской литературе все чаще обсуждается проблематика активности среды по отношению к человеку (Гибсон и др.). В результате философия сталкивается с новым феноменом, к осмыслению которого, быть может, не совсем готова. Речь идет уже не только об искусственном интеллекте, но и об искусственном человеке, искусственной личности, искусственных сообществах и искусственной жизни. И концептуальных возможностей отдельных дисциплин недостаточно для осмысления меняющейся реальности. Необходим междисциплинарный диалог, и, возможно, формирование нового направления в философии.

Актуальной становится задача осмысления антропологических и социальных последствий применения технологий искусственного интеллекта, анализ новых возникающих экзистенциальных противоречий, новых рисков человеческого существования.

<sup>18</sup> Там же, с.231.

<sup>19</sup> *Ребещенкова И.Г.* Потенциал эволюционной теории познания К.Лоренца для осмысления феномена «искусственный интеллект» // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред Д.И.Дубровского и В.А.Лекторского – М., 2005., с.223-224.

ЛИТЕРАТУРА

1. *Дубровский Д.И.* Информация, сознание, мозг, – М., 1980.
2. *Дубровский Д.И.* Гносеология субъективной реальности // Эпистемология и философия науки. 2004, №2.
3. *Зинченко В.П., Моргунов Е.Б.* Человек развивающийся. Очерки российской психологии. – М.: 1994.
4. *Кезин А., Фоллмер Г.* Современная эпистемология: натуралистический поворот. – Севастополь, НПЦ «ЭКОСИ-Гидрофизика», 2004.
5. *Лекторский В.А.* Эпистемология классическая и неклассическая. – М.: 2001.
6. *Микешина Л.А.* Философия познания. Полемиические главы. – М.: Прогресс-Традиция, 2002.
7. *Никитина Е.А.* О принципиальной неполноте «Я» // Философия и будущее цивилизации: тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.) В 5 т. Т. 1. – М.: Современные тетради, 2005.
8. Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛ, 2005.
9. *Серл Д.* Открывая сознание заново. – М.: «Идея-Пресс», 2002.
10. *Фоллмер Г.* Эволюционная теория познания: врожденные структуры познания в контексте биологии, психологии, лингвистики, философии и теории науки (Перевод с немецкого) – М.: Русский Двор, 1998.
11. Эволюция. Мышление. Сознание. (Когнитивный подход и эпистемология) – М.: Канон+, 2004.
12. *Юлина Н.С.* Головоломки проблемы сознания: концепция Дэниела Деннета. – М.: Канон+, 2004.

---

# МИРОВОЗЗРЕНЧЕСКИЕ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЭТИЧЕСКИХ ИССЛЕДОВАНИЙ ВИРТУАЛЬНОСТИ

*А.А. Скворцов*

Одной из самых интересных тем в философии искусственного интеллекта остается попытка наделения ЭВМ антропологическими характеристиками. Обычно вопрос о мере разумности машины ставится приблизительно так: может ли компьютер мыслить самостоятельно, или он есть лишь послушный инструмент человеческого мышления? Несомненно, что ответ на этот вопрос лежит в понимании сущности мышления. Но не только. В значительно большей степени решение указанной задачи будет касаться понимания сущности самого человека. Даже если мы обозначим всю совокупность электронных процессов, происходящих в компьютере, словом «мышление», то это повлечёт за собой множество других проблемных ситуаций. Например, в каких именно параметрах такое мышление повторяет человеческое? А если это всё-таки разум, похожий на человеческий, то есть ли основания полагать, что ЭВМ обладает ещё и субъектом мышления? И что же тогда представляет собой данный субъект: просто набор мета-логических операций, или некий личностный центр, наделённый свободной волей, задатками ценностного освоения мира, способностью обучаться и т.д.? Если сравнивать разум ЭВМ и человека, то следует сначала точно знать сущность качеств, которые считаются уникальными особенностями личности.

В наше время просвещенческое убеждение в том, что человек возвышается над всем остальным миром благодаря разуму, безнадежно устарело. Благодаря трудам по этологии и зоопсихологии (К. Лоренц, Н. Тимберген), мы узнали, к каким сложным мыслительным процессам способны животные. А формы группового поведения и взаимодействия у них в некоторых аспектах могут быть ещё сложнее человеческих. Философия XX века настаивала на другом важном различии, которое, впрочем, было очевидно для любой эпохи. Человека от животного отличает не наличие мышления, а нравственное сознание. Это убеждение стало настолько очевидным, что даже закрепилось в языке. «Человечным» мы называем того, кто добродетелен, а «бесчеловечным», наоборот, того, кто злобный и жестокий. Для крайних выражений зла часто используется другое слово – «зверство», хотя звери часто не способны на то, на что оказался способен «человек разумный». Правда, и здесь не всё так очевидно. Та же самая этология показала, как животные могут проявлять чувства сострадания, дру-

желюбия, любви и ревности. Да и никто никогда не сомневался, что наши меньшие братья способны к чувству жалости. Так, В.С. Соловьёв, называя чувство жалости важнейшим первичным фактом нравственности, отмечает, что её отсутствие у человека означает нахождение по своему нравственному развитию «ниже скота». Но, как совершенно очевидно доказала кантианская традиция в этике, нравственные чувства ещё не означают наличие моральности. Чувства – материя крайне изменчивая: к кому-то я их испытываю, а к кому-то нет. Но нельзя же допускать, что мы имеем право поступать непорядочно по отношению к человеку, не вызывающему в нас симпатии. К тому же, какая же заслуга личности в чувствах, данных нам от природы? Даже Аристотель полагал, что доброта, в отличие от добродетели, является природным качеством. И вся античная этика понимала нравственность как руководство чувствами посредством разума. Кант же отстаивал понимание моральности как исключительно рациональное усмотрение необходимости жить согласно моральному закону и поступать только из уважения к этому закону. Именно данное обстоятельство, по его мнению, становится ключевым различием между человеком как природным существом и человеком, как «вещи самой по себе», способным осуществлять свободный выбор. Но ещё определённое о человеке как единственном носителе нравственных ценностей и существе, способном реализовывать их в мире, высказалась в XX веке философия феноменологии. Только личность – считает Н. Гартман – «...может подчиниться идеальной власти ценностей. Остальное сущее глухо к зову идеального»<sup>1</sup>.

Итак, если использовать ещё один знаменитый образ Аристотеля, сущность человечности, т.е. моральность, занимает пространство между животным и Богом. Какое же место в ней или рядом с ней занимает искусственный интеллект? Этот вопрос кажется надуманным и некорректным, но только на первый взгляд. Действительно, разве кто-то когда-нибудь признавал ЭВМ мерилом ценностей или считал бы её нравственным существом? Так или иначе, но искусственный интеллект понимался как инструмент человеческого мышления, но не заменитель его. Однако рассматривать искусственный интеллект как изолированное явление, вне культурного контекста, – неверно. Он является порождением более широкой, технической реальности. А весь опыт философского постижения смысла техники показал, что её никак нельзя считать просто инструментом человеческой деятельности. Так, возможно, было в традиционном обществе, но в индустриальном техника стала не только приспособлением, но особой формой действительности, средой обитания человечества. Отныне она не инструмент жизни, а сама жизнь. Ещё более интересную мысль в начале

---

<sup>1</sup> Гартман Н. *Этика*. СПб.: Владимир Даль, 2002. С. 222.

XX века, полемизируя с инструментальным пониманием техники, высказал О. Шпенглер. Он назвал её «тактикой всей жизни в целом» и пояснил: «Она представляет собой внутреннюю форму способа борьбы, который равнозначен самой жизни... Речь идёт не о создании инструментов-вещей, а о способе обращения с ними; не об оружии, а о борьбе»<sup>2</sup>. Будучи рождённой в недрах рационалистического мировоззрения, и, став на время торжеством научной картины мира, техника со временем превратилась в орудие иррационального стремления жизни к могуществу, мировому господству. Конечно, за изобретением техники стоят люди и управляют ей тоже люди. Но дело в том, что эти же люди больше не могут жить без техники. Ни жить в смысле поддержания уровня достижений современной цивилизации, ни жить физически. «Творение<sup>3</sup> поднимается на творца... Властелин мира сделался рабом машины», – подводит итог Шпенглер. А там, где речь идёт об изменении человеческой жизни, неизбежно изменяется и ценностный мир. Неслучайно на рубеже XIX-XX века заговорили о нравственных проблемах техники. Это было лучшее философское подтверждение того, что техника серьёзно повлияла на человечность. Правда, единой «этики техники» нравственная философия создать не смогла. Попытки, несомненно, предпринимались, например, Фридрихом Юнгером, но такой стройной системы, как прикладные биомедицинская этика или этика бизнеса не получилось. Так, проблема подчинения личности правилам машинного производства находилась в поле внимания хозяйственной этики, появление оружия массового поражения – в рамках этики войны, деградация окружающей среды под воздействием техники относилась к проблематике экологической этики и т.д. Но иначе не могло быть: техника – это новая форма обитания человека, его повседневная жизнь, а о прикладной этике жизни как таковой говорить абсурдно.

Похожую эволюцию в отношении ценностного мира, в качестве важнейшей составляющей техники, испытал и искусственный интеллект. Во время появления первых ЭВМ никто бы не сказал, что они могут влиять на моральность. Как сугубо научный проект, «умные машины» преследовали цель существенно расширить возможности человеческого мышления, т.е. воспринимались опять же как инструмент познания. Но уже тогда в идее создания искусственного разума явно слышались поэтические мечты об интеллекте, являющем собой чистую объективность, лишённую эмоций; об уме, не устающем, не стареющем и, наконец, бессмертном. Частично этот рационалистический замысел был воплощён в компьютере. Но жизнь – если пользоваться понятием О.Шпенглера – не сводимая к чистому

---

<sup>2</sup> Шпенглер О. Человек и техника // Культурология. XX век: Антология. М.: Юрист, 1995. С. 457.

<sup>3</sup> Там же. С. 486.

мышлению, смогла приспособить искусственную ментальную реальность для завоевания новых пространств, для создания иных форм бытия. Совершенный разум вышел за свои пределы в результате создания виртуальной реальности. И этот переход стал не просто движением от одного технического открытия к другому; благодаря ему, была заложена основа перехода от индустриального типа цивилизации к информационному, т.е. от технологий накопления информации к безграничным возможностям её распространения. Исследования логики такого перехода – отдельная большая тема. Нам же интересен, в первую очередь, иной аспект произошедшего: создание на базе ЭВМ виртуальной реальности означало вторжение искусственного интеллекта из сугубо когнитивной сферы в ценностный мир. После полувековой истории развития виртуальности сегодня она видится нам как совершенно нелогичное, часто парадоксальное пространство, такое же, как сама жизнь. Раньше она выглядела как порождение техники, теперь она стала самостоятельной формой жизни наряду с техникой. Более того, дальнейший научно-технический прогресс невозможен без виртуальности. И крайне затруднительной без неё стала сама жизнь.

Ныне ситуация такова, что уместно говорить о значительном отдалении искусственного интеллекта от виртуальной реальности. Ещё в 70-х годах XX века, когда информационные сети были доступны всего нескольким тысячам людей, преимущественно имевшим отношение к научной сфере, существовали виды на рациональное развитие киберпространства. Казалось, что одно вытекает из другого: если искусственный интеллект может мыслить, то почему бы глобальной информационной сети не стать общечеловеческим «мировым разумом»? Но сейчас подобная логика очень далека от реального положения. Мы, конечно, понимаем, что функционирование сетевого пространства невозможно без работы компьютерной техники. Однако как явления культуры они в массовом сознании перестали ассоциироваться друг с другом. Учёный, рассуждающий над путями усовершенствования искусственного интеллекта, может вообще не пользоваться Интернетом, и, напротив, человек, не имеющий никакого познания в информатике и кибернетике, владеющий элементарными навыками программирования, способен создавать выдающиеся сетевые проекты. Если важнейшей характеристикой наступления технической эры считалось появление огромной группы людей-техников, занятых исключительно конструированием и обслуживанием машин, то ныне число людей, профессионально работающих в виртуальном мире, стремительно приближается к количеству техников. Все эти факты наглядно демонстрируют, что обе реальности – техническая и виртуальная – сегодня выглядят как равносильные ареалы человеческой жизни. Важнейшим и, в то же время, роковым различием между

ними стал сам вид культур, к которым относится то и другое явления. Техническая мысль до сих пор является преимущественно выражением интеллектуальной, научной деятельности, в то время как сетевое пространство – важнейшая часть массовой культуры. И не просто часть, а главный носитель и пропагандист. Ныне виртуальность выглядит радикально нерациональной, она не желает формализовываться и организовываться, она желает жить и расширять себя без всяких ограничений. Это не рациональная сфера, а поле кипения великих страстей, чьё многообразие позволяет ей быть второй действительностью, иной, но абсолютно достоверной для обитания человека реальностью. И очевидно, что тяга к созданию виртуальных миров неискоренима; отныне они станут вечным спутником человека, таким же, как и техника. А раз так, то никто уже не сможет отрицать, что виртуальность оказывает существенное влияние на ценностный мир личности.

Но как именно сетевая реальность повлияла на моральный мир человека? По этому вопросу высказываются диаметрально противоположные точки зрения, как будто бы речь идёт о совершенно различных явлениях. Оценки колеблются от признания мировой сети главным рассадником порока до веры в возможность организации глобального диалога с целью дипломатического решения всех социальных проблем. И этот разброс мнений уже свидетельствует, что однозначно оценить вторжение виртуальности в человеческую жизнь невозможно. Из истории цивилизаций мы прекрасно знаем, что самые яркие, гуманистические проекты, с которыми связывались надежды на переход общества к принципиально иной стадии существования, заканчивались отчуждением. Техника, обещавшая, казалось бы, освобождением человека от тяжёлой трудовой повинности, в итоге поставила цивилизацию на грань уничтожения. Не произойдёт ли то же самое с виртуальной реальностью? Как добиться того, чтобы она способствовала торжеству человечности, а не к ее деградации? Здесь мы вступаем в сферу этического поиска. Если оценивать появление виртуальности, экзистенциально, т.е. с точки зрения удовлетворённости человечества своим положением в мире, то надо признать: её стремительное распространение лишней раз показало трагическое одиночество человека в мире. Если бы нам было достаточно обычного общения и перспектив развития в его рамках своих способностей, то вряд ли бы потребовалось конструировать новую реальность. Но случилось так, что личность оказалась готова искать себе собеседника в искусственной действительности, часто соглашаясь идти на разрыв с реальной жизнью. Та, в свою очередь, становится для неё спасением от обыденности, местом проявления таланта, не нашедшего возможности реализовать себя в социуме. Несомненно, что быстрое завоевание Интернетом человеческих душ стало выражением кризиса межличностных отношений в современном обществе. С дру-



гой стороны, появление виртуального мира лишний раз доказала нам банальную мысль о том, что человек – не самодостаточное существо. Он вечно обречён развивать свой творческий потенциал, стремиться к целостности, осваивать новые формы бытия. Таковы антропологические предпосылки появления искусственного электронного пространства. Но кризис традиционных человеческих форм коммуникации необходимо влечёт за собой кризис моральности, ибо она по своей сути есть представление об идеальных отношениях между людьми.

Если проводить сугубо этический анализ виртуальной реальности, то, на первый взгляд, мы увидим немало отрицательных черт. Самая главная претензия, которую предъявляют сетевому сообществу – ничем неограниченная пропаганда и распространение порока. Колоссальные образовательные возможности, как и неограниченные ресурсы межкультурного общения, явно вытеснены из сетевого пространства индустрией развлечения. Но, подумаем: могло ли быть иначе? Интернет создавался в определённой духовной ситуации и набирал силу в творческих поисках людей, осознавших свою ничтожную роль в процессах глобализации. Люди изначально увидели в виртуальном пространстве возможность дополнительной свободы, в т.ч., от моральных требований; именно в этом пункте Интернет с самого начала стал зоной проявления нравственного кризиса. Он по своей сути содержит в себе соблазн совершить то, на что в реальной жизни человек бы никогда не решился. Отсутствие социального контроля вполне естественно повлекло за собой предосудительные формы поведения. Но разве в этом есть что-либо удивительное? Всегда любые новые, необычные социальные явления влекли за собой и новые соблазны. Например, телевидение, обещавшее едва ли не фантастический рассвет культуры, сегодня лишь процентов на пять осуществляет просветительскую миссию; всё остальное – это частные коммерческие интересы, которые на девяносто процентов реализуются так же через пропаганду порока. Поэтому за возмутительные вещи, обитающие в сетевом пространстве, несёт ответственность не сам Интернет как глобальное коммуникативное сообщество, а те люди, чья внешняя благопристойность держится не на внутренних моральных устоях, а на боязни социальных санкций. В этом смысле можно заметить, что сетевая реальность стала определенным индикатором морального состояния личности: кто считал порок допустимым, тот воспользовался Интернетом для его культивирования, тот же, кто считал недопустимым, отказался использовать виртуальность в корыстных целях. Другой часто встречающийся довод моральных обличителей Интернета касается утверждения, что виртуальность подменяет реальную жизнь иллюзорной. Пользователи сети становятся совершенно недееспособными в обыденных обстоятельствах.

На этой почве якобы развиваются психические болезни, происходит интеллектуальная деградация. Кто спорит: такое, к сожалению, случается. Иногда подростки впадают в зависимость от компьютерных игр, полностью уходят в виртуальное общение, а вполне зрелые люди заканчивают депрессией, если им не удаётся ежедневно просматривать порносайты. Однако все эти явления суть такие же виды социальных пороков, как алкоголизм и наркомания. И нельзя явные отклонения выдавать за норму. Очевидно, что далеко не все подвержены сетевым болезням. И, напротив, многим людям удалось победить пороки, благодаря самореализации в Интернет-проектах. Что же касается замены реальной жизни иллюзорной, то данный вопрос чрезвычайно сложен. Уже в самой его постановке видно совершенно ложное убеждение в полной оторванности одной сферы жизни от другой. Но ведь это совершенно не так. Человек может огромное количество времени проводить, работая в виртуальной реальности, но при этом быть очень даже решительным в повседневных делах. Если уж настаивать на неизбежных проблемах с адаптацией у жителей сетевого пространства, то следует обвинять в том же самом художников, учёных, писателей и вообще всех людей, живущих богатой внутренней, духовной жизнью. Кое-кто себе это позволяет, но мы прекрасно знаем, что все подобные представления – слишком превратны. Но самый главный недостаток сетования на излишнее увлечение иллюзорной действительностью заключается в совершенно некритичном превознесении реальности и очернении виртуальности. Во-первых, это неправильно, поскольку виртуальность – есть часть реальности, и отражает пороки, существующие в обыденной жизни. Во-вторых, разве мало стрессов случается в окружающей нас жизни? И разве люди не теряют социальные навыки адаптации из-за агрессивного воздействия на них внешних событий? И если уж считать Интернет исключительно иллюзорной реальностью, то он – отнюдь не самый её бесчеловечный вид. Существуют значительно более злобные проявления, как, например, тоталитарные секты и идеологии, бороться с которыми помогают, в т.ч., и просветительские сайты. То же самое можно сказать и о снижении интеллектуального уровня. Несомненно, виртуальная реальность косвенно этому способствует, ибо привыкшие визуально усваивать информацию подростки больше не могут читать книги, требующие от них хотя бы малейших умственных усилий. Но в таком положении дел, в первую очередь, виновато телевидение, а уж потом Интернет. С другой стороны, существует огромное количество образовательных порталов, которые могут более доходчиво преподнести информацию, чем неудачно написанные учебники. Посредством сетевого пространства можно изучать языки, знакомиться с мировой культурой, участвовать в интерактивных конференциях. Иными словами, перед подростком есть выбор: использовать компьютер и Интернет как развлечение, либо с их помо-

щью заниматься собственным образованием. И этот выбор будет зависеть не от факта наличия Интернета, а от воспитания и социального окружения подростка, а так же от нравственных устоев людей, влияющих на его жизнь.

Какие бы в дальнейшем отрицательные черты виртуальности мы не разбирали, станет ясно, что все они – не сущностная сторона сетевого пространства, а обычное проявление людских пороков, появившихся задолго до самых ранних идей об искусственном интеллекте. Если рассматривать этот вопрос более углублённо, то можно заметить, что в отличие от общения человека с техникой, которое носит обезличенный характер и заставляет быть такими же тех, кто ставит свою жизнь в зависимость от подобного общения, виртуальная реальность является местом проявления ярко выраженных личностных качеств. При помощи техники можно улучшить свою материальную жизнь, но не духовную. Поэтому несомненная ценность Интернета состоит в возможности проявить творческий потенциал. Более того, успеха на этом поприще достигают люди с ярко выраженными креативными способностями. Некоторые из них, не сумевшие проявить себя в обыденной жизни, достигли огромных успехов в сетевом пространстве и получили широкую известность. Иногда, знакомясь с некоторыми порталами, посвящёнными разным аспектам мировой культуры, испытываешь подлинное восхищение подвижническим трудом их авторов. Виртуальность – это сравнительно новое поле человеческой деятельности, но его новизна не делает его более трудным для освоения моральным сознанием, чем иные сферы. Как и в любом поле человеческого бытия, в ней есть место для проявления и добра, и для зла. И так же, как и везде, в сетевом сообществе добродетель похвальна, а порок занимает такое же точно место, как и в обычной жизни: ему многие подвержены, но большинство считает его моральным злом, уклонением от нормы человечности.

Однако самая большая опасность для человечности, связанная с появлением виртуальной реальности, лежит не в электронном тиражировании порока. Страшен не сам по себе порок, а тот, кто его оправдывает и считает приемлемым. В этом смысле весьма тревожной выглядит попытка некоторых интернет-технологов доказать, что в киберпространстве якобы действует особая «виртуальная мораль». То есть подспудно подразумевается, что всё, существующее в противоречивом сетевом сообществе, в принципе, допустимо. Недаром же используется слово «мораль», которое, несмотря на множество смыслов этого слова, в обыденном языке, как правило, воспринимается в качестве положительной ценности. Но из истории человечества мы знаем: всюду, где доселе утверждалось существование двух моралей – всечеловеческой и особой, существующей для избранных, – дело заканчивалось нравственной катастрофой. Ещё раз скажем: вирту-

альность – это неотъемлемая часть нашей действительности, и никакой особой морали там быть не может. Моральность, как в смысле высших нравственных ценностей свободы, жизни, добродетели, сострадания, прав человека и т.д., так и в смысле важнейших заповедей не убивать, не красть, не лгать, не причинять вреда своему ближнему и др. – абсолютна. И трудно представить, как факт появления сетевого пространства мог бы обесценить моральные абсолюты. В конечном итоге, миллионы людей, работающих в сфере виртуальных технологий, вряд ли бы согласились, чтобы осуждаемые всеми действия стали бы нормой их жизни. Тот, кто связал с Интернетом свою профессиональную деятельность, всё-таки желает работать в человеческом, а не в бесчеловечном пространстве. Но что же защитники «сетевой морали» понимают под этим выражением? Как правило, – изменение, произошедшее с самим субъектом электронной коммуникации. Например, у такого субъекта отсутствует телесность, следовательно, заповедь «не убий» теряет свой смысл. Но, во-первых, это, пожалуй, единственная заповедь, на которую может пасть подозрение в неактуальности для сетевой действительности; все остальные заповеди там весьма актуальны и действенны. А, во-вторых, сам этот пример ложен. Тот факт, что нельзя физически лишить жизни субъекта виртуального пространства, не отменяет оценку убийства как тяжчайшего преступления против личности. Поэтому пропаганда убийства и насилия в виртуальном пространстве также считается преступной и влечёт за собой как уголовное преследование, так и нравственное осуждение. Далее, утверждать полное отсутствие телесности обитателей сетей – совершенно необоснованно. В лучшем случае можно говорить, что тело отсутствует косвенно, т.е. информация, получаемая из искусственной реальности, очень слабо влияет на тело. Но, с другой стороны, за виртуальными проектами находятся живые люди, и для того, чтобы просто двигаться в сети и воспринимать данные, следует прикладывать физическую силу и обладать органами чувств. Но главное: моральность никогда не имела отношения только к телу. Она всегда взывала к разуму и чувствам человека; тело же является материальным инструментом для реализации намерения. Виртуальность также воздействует на чувства и разум личности, побуждая её к определённой деятельности. Что принципиальным образом здесь изменяется в сравнении с обыденной действительностью, чтобы возможно было говорить о разных моральных, – непонятно. Вторым ключевым аргументом сторонников двойственности добра и зла, применяемым в спорах, является указание на анонимность поступков, совершаемых в сети. В реальной жизни человек боится за свою репутацию и опасается совершить какой-нибудь экстраординарный поступок, а в Интернете не существует социальных ролей; там все абсолютно свободны и равны в способах самовыражения. Хорошо видно, как отсюда напрашивается радикальный вывод: в сетевом

пространстве всё позволено. А там, где всё позволено, действительно, трудно говорить о морали. Но, занимаясь подобными размышлениями, кто-то явно выдаёт желаемое за действительное. Допустим, некий деятель написал прокламацию откровенно фашистского содержания и анонимно опубликовал её в Интернете. Но уже факт анонимного обнародования свидетельствует о его боязни открывать своё имя, а значит, он понимает, что в данном обществе позволено далеко не всё. И публикация в Интернете может повлечь для него такие же правовые последствия, как и публикация в любом журнале. Если же анализировать работу сетевого пространства в целом, то мы видим, что ведущие, наиболее популярные проекты и издания никогда не позволят себе откровенно преступной пропаганды. Они очень дорожат своей репутацией и следят за собственным моральным обликом. Иное дело, есть масса сайтов, в самом деле работающих по принципу «всё позволено». Но к ним обычно вырабатывается соответствующее отношение как к форме девиантного поведения, и отнюдь не они играют главенствующие роли в глобальной сети. Наконец, сама характеристика действий как анонимных ещё не означает их обезличенности и, тем более, безнравственности. Как уже приходилось отмечать, миллионы людей приобщаются к Интернету ради нового общения. Отсюда столь большой популярностью пользуются форумы и, так называемые, сетевые «болталки». Конечно, это общение анонимно: никто ни с кого не требует анкетных данных; при желании можно представиться человеком противоположного пола, иного рода занятий и иного благосостояния. Можно приписывать себе совершенно несвойственные взгляды и поступки, но, в любом случае, суть подобного общения неизменна: одна личность желает общаться с другой личностью, желает найти родственную душу. Получается, что в такого рода общении мы наблюдаем не тенденцию к обезличиванию, а, наоборот – торжество личностного начала. Но если это так, то в межличностном общении никак не может быть всё позволенным. Человек дорожит общением с таким же, как он и не желает оскорбить его, дабы не потерять навсегда. При желании можно ещё привести несколько убедительных аргументов против существования автономной сетевой морали, но, по-моему, всё и так очевидно. Справедливо говорить лишь о соблазне её исповедовать, но не о реальном положении дел.

Однако, несомненно, что поведение людей в виртуальной реальности всё-таки обладает моральной спецификой. И если грамотно говорить на этическом языке, стараясь указать на не совсем обычный характер виртуальной реальности, то справедливо было бы говорить об особенном этосе обитателей сетевого пространства. Именно не об ином, а об особом, и только в смысле указания на определённые черты поведения, нехарактерные для обыденной жизни. А так же на своеобразные обстоятельства, в которых осуществляется виртуальная

коммуникация. Сюда, конечно, можно включить наблюдения о сведении к минимуму проявлений телесности, о тяге к анонимности, о молниеносном распространении информации, о возможности ролевых игр, о появлении новых широких перспектив для проведения пир-кампаний, и т.д. Но это будет лишь означать, что каждая указанная черта может способствовать как укреплению добродетели, так и развитию порока. Но раз мы заговорили об этосе, то его анализ применительно к сетевому пространству должен вращаться вокруг ответа на вопрос: какие меры обычно применяют жители виртуальных миров против распространения порока? И, напротив, что ими делается для очеловечивания данного пространства? Какие нравственные качества для этого требуется усилить, на какие опасности обратить особое внимание? Мы понимаем, что перечисление подобных качеств не выявит серьёзного отличия между нравственным миром человека как деятеля киберпространства и существа из обыденной реальности. Так же, впрочем, как и понимание добра и зла. По иному и быть не может, ибо обитатель и той, и другой реальности, как правило, – один и тот же человек. Но всё же каждая отмеченная черта виртуальности требует своего, особого нравственного ответа. Например, из молниеносного тиражирования информации следует необходимость большей ответственности личности, размещающей какие-либо материалы в сети.

Помимо указаний на особенности сетевого этоса, можно так же говорить и о прикладных этических исследованиях виртуальной реальности. Они будут касаться уже не специфики поведения человека в сетевом пространстве, а обсуждения проблем, возникающих при взаимодействии виртуальности и окружающей её действительности. Например, вопрос об авторских правах, о размещении рекламы на сайтах, о защите чести, достоинства, деловой репутации и т.д. Такие дискуссионные темы, кстати, лишний раз доказывают, что никакого принципиального морального различия между двумя сферами нет, и мы при решении практических вопросов пользуемся единым мериллом допустимого и недопустимого. В противовес голосам, возмущающим особую «сетевую мораль», всё настойчивее звучат голоса, требующие ценностного упорядочивания киберпространства. Предлагается, например, посредством использования Интернета в качестве поля для глобального диалога, выработать единые нормативные правила, общий нравственный кодекс жителей Сети. Установленные им правила принимались бы добровольно теми, кто желает работать в подлинно гуманитарном, а не нравственно-противоречивом пространстве. Ценностной основой подобного кодекса должно стать убеждение, что за каждым виртуальным проектом стоят люди со своими интересами, мировоззрением, моральными убеждениями. Участники виртуального диалога – личности, а не машины. И оскорбляя, нанося урон субъекту сетевой коммуникации, мы вредим человеку, а не безличной информации. Виртуальность – это естест-

венное продолжение и развитие человеческой творческой мысли. А там, где действует и творит человек – там всегда открыта возможность реализации нравственных ценностей. И не виртуальных, а настоящих, действенных. Как-то в процессе безоглядной критики Интернета теряется из вида, что его ресурсы можно использовать так же и для нравственного просвещения людей.

Этические исследования виртуальности – очень интересная, перспективная тема. Но перед ней стоит большая опасность превратиться в сугубо описательную дисциплину, просто фиксирующую происходящие процессы. Однако, точно так же, как природа не терпит пустоты, не терпит её и сетевое пространство. Если философская этика не возьмёт на себя миссию трансляции ценностей из обычной действительности в сетевую, то её роль исполнят всякого рода интернет-технологии и пиарщики; к каким причудливым выводам они обычно приходят – уже приходилось говорить в данной статье. Помимо сбора фактического материала, этика, ведущая исследования в указанной области, обязана искать пути реализации и укрепления нравственных ценностей в виртуальной реальности. И не просто искать, а, используя безграничные возможности сетевой коммуникации, показывать и пропагандировать примеры подлинных человеческих отношений, разрабатывая собственные сетевые проекты. В этой связи надо понимать, что появление виртуальной реальности дало нравственной философии уникальный шанс незамедлительно превращать слова в дело. Не использовать такую возможность было бы непостижимым упущением.

# «ЧЕЛОВЕЧЕСКОЕ, СЛИШКОМ ЧЕЛОВЕЧЕСКОЕ», ИЛИ ПРОЦЕССОР В ЭКСТАЗЕ

*В.И. Самохвалова*

В последнее время отношения человека с созданной им же техникой стали приобретать ощутимые черты проблемности. Машина не только выполняет технически взаимодополнительные функции, как этого изначально и хотел от нее человек, но и начинает обретать как бы независимое от него, самостоятельное бытие. Так, например, специалисты говорят о формировании техносферы — по аналогии с биосферой и антропосферой, — которая способна на самостоятельную внутреннюю эволюцию рядом с человеком и как бы отдельно от него; наряду с чувством удовлетворения от того, что человек благодаря машине получил могучего многофункционального помощника, он начинает ощущать растерянность и даже опасение. Всё чаще звучат предостерегающие голоса, сравнивающие компьютер, например, с троянским конем, которого когда-то давно с радостными возгласами втащили в ворота своей крепости сами же осажденные в ней; воспоминания о результате этой операции навсегда закрепились в истории. Утверждают, что человек ныне таким же образом недооценивает опасности распространения компьютеризации, не учитывает невероятных скоростей технологического усовершенствования машин. Так, японцы уже создали компьютер, который совершает 300 млрд операций в сек.; на подходе компьютер со скоростью 1 трлн операций в сек. В то время как Г.Каспаров проиграл компьютеру «Deep Blue», который просчитывал варианты *всего лишь* со скоростью 200 млн операций в сек. И т.д. Всё чаще говорят о естественном превосходстве искусственного машинного интеллекта над непоследовательным и ненадежным естественным, о моделировании человеческой личности, о том, что ЭВМ может выступать творцом, способным и к созданию произведений искусства. Разумная техника вообще грозит в будущем вытеснить человека как менее эффективную и менее рентабельную «модель».

Сложившаяся ситуация требует от человека, во-первых, не только новой определенности в его позиционировании в мире, где всё большее место и значение начинает принадлежать технике, но и методологической ясности в отношении содержания таких понятий, как «интеллект», «творчество» и т.п., и правомерности их применения, когда их субъектом выступает искусственный объект. А, во-вторых, сегодня человеку, как некогда Пигмалиону, создавшему Галатею, приходится — не слишком надеясь на известные «три закона робототехники», некогда сформулированные Айзеком Азимовым, — думать о том, как сохранить расположение своего творения и не утратить статус его творца.



Иначе складывается ситуация, о которой, пугая читателя, А.С.Спирин говорит как об уже свершившейся: «Человек проиграл компьютеру. Вероятно, в будущем он станет игрушкой для компьютеров, которые лучше и быстрее обрабатывают, что и как надо делать»<sup>1</sup>. Если оставить в стороне полемическую заостренность высказывания, то можно отчасти согласиться с тем, что быстрее — да, однако, что лучше — проблематично, тем более что это их «соображают» означает лишь оперирование с имеющимися данными и способность удерживать «в памяти» и мгновенно перебирать множество вариантов (именно поэтому ЭВМ вполне может выиграть у *среднего* шахматиста). Если понимать, например, науку только в таком операциональном смысле, тогда в определенном отношении компьютер действительно легко может заменить человека в подавляющем большинстве рутинных операций; однако в собственно творческом отношении, которое имеет столь большое значение и в науке, и в искусстве, компьютер вряд ли пока может составить человеку конкуренцию. Так, в середине декабря 2004 г. к выведенному на орбиту телескопу «Хаббл» отправили космонавтов, но не роботов, мотивировав это именно тем, что машина не сможет сориентироваться на месте и справиться с неполадками, которые предварительно еще надо адекватно квалифицировать. То есть *пока* еще автомат не способен действовать в экстремальных ситуациях, как это может делать человек.

Компьютер действительно выигрывает, когда речь идет о *количественных* аспектах дела, но там, где вопрос касается *качественной* стороны мышления, — приоритет по-прежнему остаётся за живым мозгом: гибкость и вариативность его функционирования, многообразие и подвижность его внутренних связей обеспечивают многоканальность обработки информации, возможность продуктивных «случайностей», неожиданных пересечений, что проявляется в непредсказуемости творчества, в совершаемых открытиях. Живой мозг, в отличие от машины, может действовать не только алгоритмически, но и эвристически, и эта эвристичность его поведения обнаруживает основополагающее отличие творческого поведения, имеющего, в том числе, и свои *биологические* предпосылки<sup>2</sup>, — от аналогового.

Именно внесение определенности в содержание тех понятий, которыми он пользуется для описания объектов и процессов моделирования, позволит ему увидеть, насколько, например, оправдан перенос понятий и представлений когнитивной психологии на область бихевиористики, концепциями которой он по сути руководствуется при описании искусственного интеллекта, разумного поведения и т.п.

---

<sup>1</sup> См.: Известия – Наука. 2002. № 16 (33). С. 1.

<sup>2</sup> Анализ биологических предпосылок творчества посвящено большинство статей книги «Красота и мозг. Биологические аспекты эстетики». М., 1995.

Конечно, при сравнении организации работы человеческого мозга и компьютера возникают определенные аналогии, однако говорить о *сходстве* здесь было бы некорректно. Подобное «сходство» возникает здесь не путем естественного, независимого генезиса, т.е. не является результатом, так сказать, гомотетической онтологии, но выступает как порождающее отражение, будучи результатом «искусственной» (направленной) человеческой деятельности. Ведь это человек создает искусственный интеллект по образу и подобию своего собственного, и мистифицирующие переносы смысла здесь неоправданны. ЭВМ «играет» роль, написанную для нее человеком с себя, ибо других образцов у него фактически нет. При этом к действующему лицу, как известно, подходят с требованиями экзистенциального плана (смысла, нравственности, последовательности и т.п.), от актера же, исполнителя его роли, требуют лишь соответствующей актерской техники, достоверности исполнения и т.п. ЭВМ «играет» человеческую личность, как более или менее искусный актер образ своего героя. Рассуждать о машинном разуме в терминах когнитивной психологии представляется некорректным, ибо у компьютера всё-таки нет ни психики как таковой, ни самостоятельной познавательной деятельности.

При этом человек к тому же спешит с созданием модели того, в чем он сам до конца не разобрался. Это не значит, что человек должен запретить себе поиск — даже сомнительный, и эксперимент — даже рискованный, но ему не следует допускать ментальную и методологическую неряшливость, торопясь выдать желаемое за действительное. Человек до сих пор до конца не знает и «собственный материал»: так, известно, что человек использует свой чрезвычайно сложно устроенный мозг едва лишь на 15%, не зная, что делают и для чего нужны остальные 85%, как они влияют на деятельность этих 15% — или же, может быть, используются как-то иначе; человек не знает возможностей и принципов действия собственной психики, которая во многом остаётся для него самого «черным ящиком», относительно происходящего в котором существуют весьма различные версии. Когда человек поспешно моделирует деятельность этого своего «черного ящика», то это в определенном смысле напоминает ситуацию из детской сказки про волшебника Хоттабыча, который по образцу телефонного аппарата делает его подобие из *цельного* куска мрамора.

Многие поведенческие акты человек действительно может передать ЭВМ, потому что и сам может разложить их на микроэтапы (микрооперации), поддающиеся формализации. Однако разложить подобным образом можно поведение (отдельные действия или их последовательность) как ответ человека на те или иные внешние сигналы или стимулы, но не психику как таковую, ибо пока она и для самого человека остается феноменом, хотя и специфически характеризующим его, однако им самим до конца не понятым и недостаточно изученным.

В обычных случаях человек при своем мышлении использует целый ряд мыслительных стереотипов, которые для него явились результатом предшествующих когнитивных усилий. Теперь это своего рода готовые клише, в которых ему *удобно* моделировать действительность, ибо это экономит время и мыслительные усилия. Именно подобными упрощенными стереотипами и может действовать машина, которая не предназначена для индивидуализации информационно-вычислительных процессов. Это актуализация, во-первых, *типового*, во-вторых, *поведения*. Первое не предполагает эвристичности, спонтанности, что мы связываем с процессом творческого мышления (а настоящее мышление — всегда в той или иной мере творческий процесс), второе говорит о том, что это именно поведение, а не собственно мышление, даже пусть упрощенное. Именно это и имеется в виду, когда говорят о необходимости адекватной интерпретации машинного интеллекта, т.е. о применении к нему бихевиористских подходов.

Подобное упрощенное мышление машины можно назвать своего рода *рефлекторным* мышлением, которое достаточно далеко от мышления в собственном смысле слова и тем более от творчества. Человек с помощью машины может воспроизводить операции, действия, даже очень сложные, но не то, что за ними стоит и их обуславливает. ЭВМ накапливает объективную информацию, но в нее не могут быть включены цели и планы, надежды и ожидания. Без этого картина реальности может быть лишь ее упрощенной схемой. Психика человека несводима к его лишь внешнему поведению, и пока человек не расшифрует собственную психику, говорить о машинном творчестве преждевременно. Машина способна взять на себя большинство рутинных операций и обеспечить недоступную для человека скорость действий, но не может выйти за назначенные ей пределы и не может предложить какой-то иной, новый ответ на ту или иную, даже типовую, ситуацию. ЭВМ не осознаёт свои действия и не может вмешаться в них. У нее нет той структуры, которая человеку обеспечивает наличие собственно человеческого сознания: наряду с «я» действующим и сознающим присутствует и «я» самоосознающее. Человек может говорить о своем сознании, когда осознаёт себя сознающим. Иначе говоря, сознание есть живой диалог между «мозгом-деятелем» и «мозгом-наблюдателем» деятельности. И потому интеллект, с одной стороны, и манипулятор стереотипными схемами, с другой, — по сути и в принципе различны. В известной степени использование понятия «интеллект» по отношению к машине выпадает как преимущественно метафора, а не однозначный термин. Только с известными оговорками можно согласиться с тем, что ЭВМ, обладающая способностью быстро и точно совершать некую сово-

купность операций с информацией, имеет то, что *формально* можно определить как интеллект.

Конечно, человек в своем поведении часто осуществляет множество автоматических (неосознаваемых и не окрашенных эмоционально) действий. Мы, как правило, не задумываемся над теми своими действиями (и не переживаем их), которые выполняем по привычке. Хотя осознаём и внешнюю цель действия, и способ ее достижения. Утром, например, мы встаем, включаем плиту, делаем себе кофе. Мы обычно делаем подобное, что называется, на автомате, не задумываясь, в какую руку берем чашку, в какую банку с кофе и т.д., т.е. не контролируем действия в микропоследовательности их составных элементов. Нам действительно нет нужды задумываться над привычными действиями, над тем, что хорошо именно своим автоматизмом — например, дыханием (над этим мы задумываемся лишь в том случае, когда, например, практикуем дыхательные техники с целью «перепрограммирования» своего процесса дыхания для перевода его на нужный режим) или хождением (кроме случаев, когда мы, например, после снятия гипса на сломанной ноге вынуждены как бы заново учиться ходить).

Однако ученый, по профессии своей отличающийся от «обычного человека» тем, что время от времени задумывается над обычными действиями и вещами, над их причинами, смыслом и последовательностью, задается вопросами «что» и «как» мы делаем. Что это лично ему дает? Начинает ли он производить эти действия быстрее? Как правило, нет, ибо утрачивается их автоматизм. Может быть, эффективнее? Вряд ли, ибо часть энергии уходит не на действия, а на «бесцельное» их прослеживание и осмысление. Тогда зачем ему это?

В том-то и дело, что понимание действительных причин этих якобы ненужных размышлений принципиально недоступно машине, поскольку ей, в отличие от человека, нет необходимости составлять *программы* подобных действий, в основе которых лежит антиципирование нужного результата. Автоматизм привычных человеку действий теряет смысл и ценность, когда возникает необходимость передать наши функции *другому* (например, той же машине): тогда-то мы и оказываемся вынуждены задуматься над содержанием и порядком наших действий. Этот *другой* (который может быть из другой экзистенциальной традиции или даже другой цивилизации) не умеет совершить подобные действия интуитивно (на автомате), как мы. Ему-то мы и должны разложить наши действия «по полочкам». В результате сделает ли машина эти действия лучше нас? Вряд ли. Но зато быстрее — наверняка. Она делает те же операции, что и мы, но разложенные на десятки или сотни мелких последовательных этапов; однако делает это в тысячи раз быстрее, чем мы. Затем она нам и нужна. Мы можем то, что она, но мы — медленнее. Она — не лучше

нас, но — быстрее.. Значит, главный критерий здесь — *скорость*. Однако ни в науке, ни в искусстве скорость, как правило, не является ни решающим фактором, ни решающим критерием.

Искусственный интеллект не может ничего того, чего не можем мы, ибо он может только то, чему *мы* его научили (при условии, что предварительно освоили это сами). Но его быстродействие, то, что он делает это гораздо быстрее нас, — есть его выигрыш. Он экономит нам время, за секунду перебирая в своем, заложенном нами, «уме», миллионы возможных вариантов решения. Это, собственно, наша *машина времени* (но не интеллект в собственном смысле слова).

Мы можем найти решение — не обладая всем набором нужной информации — используя интуицию и творческий поиск. Автомат не может *найти* там, где ничего нет: он может лишь *перебрать* все имеющиеся (заложенные нами) варианты и выбрать подходящий — в случае, конечно, если такое поведение предусмотрено нами в его программе. Мы можем создать программу и выработать алгоритм действия. Автомат может извлечь это из своей памяти (если это туда заложено) и следовать заданному алгоритму.

Ученый (как это демонстрирует «ученый консультант по науке» Брунцов из старинного кинофильма Г.Александрова «Весна») — «сел, задумался, открыл». Даже самый лучший автомат (во всяком случае пока) не может *открыть*, ибо его «задумался» есть лишь технический процесс переборки возможных вариантов из набора имеющихся в памяти. Интеллект (человек) сам избирает себе объект «задумывания», сам ставит задачу. Автомату объект задается; ему ставится уже сформулированная задача. Его функция — сблечь наше *время* и тем самым как бы увеличить *пространство* охвата действительности нашим мышлением.

Может ли автомат, используя эти свои временные и условно пространственные козыри, — переиграть нас? В рутинной ситуации — безусловно! В ситуации дефицита исходной информации или ее неопределенности, в ситуации, требующей творческого подхода, — нет! «Творчество» автомата ограничено, хотя он и способен на некоторые самостоятельные действия (но они обусловлены определенными программными установками). Машина может дать новое композиционно и технологически, но не смысло-содержательно.

Итак, повторяем, что целостная психика человека не сводима к внешнему предметному поведению. Машина демонстрирует поведение, внешне похожее на разумное, ибо оно является реализацией определенных программ. Однако как человек не может смоделировать собственную психику, так и машина поэтому не может «обладать» психикой или интеллектом. Проблему так называемого машинного интеллекта можно отнести скорее к проблемам бихевиористики, чем когнитивной психологии, ибо моделируется не сам интеллект и тем более не

психика в целом, а только лишь поведение. Машина не осознаёт и не оценивает свои действия, она не может сама корректировать их в случае возникшей по ходу дела необходимости, как это может делать человек. У человека, как уже говорилось выше, его самосознающее «я» отражает «я» сознающее и действующее, в то время как у ЭВМ подобной «двусубъектности» нет. Природа живого интеллекта и машины, которая лишь манипулирует готовыми мыслительными схемами, различна. «Нет никаких оснований, — утверждает психолог С.Э.Поляков, — для антропоморфизма в машиностроении, так как человек — это не биологическая машина, как полагали бихевиористы, а психика — не эпифеномен. Психика — это то, что *ответственно* за интеллект и разум»<sup>3</sup>. Он даже считает, что само применение терминов типа «машинное мышление», «машинный разум» наносит реальный вред науке созданием мифических когнитивных конструкций, ибо на самом деле человек никогда не создавал даже самого элементарного искусственного интеллекта; он лишь конструировал приборы, имитирующие возможные человеческие ответы на определенные внешние воздействия, и подобные приборы только *кажутся* разумными<sup>4</sup>. На современном этапе знания человеку еще не удалось создать искусственную психику.

По той же причине неоправданно говорить и о создании (моделировании) искусственной личности. Человек осмысленно приобретает знания и целенаправленно, сознательно использует их, в то время как «понимание» машины носит оперативно-манипулятивный характер. Искание истины у него не отделимо от эмоционального «фона» — как в плане мотиваций, так и в смысле переживания процесса и результатов поиска. Безэмоциональная искусственная личность — оказалась бы личностью в известном смысле патологической. Кроме того подобная личность была бы и асоциальной, ибо ее «социальность» могла бы носить лишь искусственный, рефлекторно-принудительный характер. Таким образом, у нас фактически нет оснований считать подобную конструкцию — с заменой сложных социопсихических функций имитационными поведенческими актами — аналогом *человеческой* личности. В этом случае можно было бы говорить лишь о формировании нового искусственного «социума» рядом с традиционным человеческим социумом (а может быть, и вместо него), т.е. о формировании некоей новой квази-социальной реальности.

Обратимся теперь к феномену *творчества*, при рассмотрении которого обозначенные различия между человеком и машиной выступают наиболее очевидно. Конечно, можно и творчество по-бихевиористски представить как всего лишь сложно мотивированное и столь

---

<sup>3</sup> Поляков С.Э. Мифы и реальность современной психологии. М., 2004. С. 447.

<sup>4</sup> См.: Там же. С. 448, 449.

же сложно организованное поведение. Однако в этом случае остались бы совершенно непонятны движущие мотивы творческой деятельности, непредсказуемый ход и сама специфика процесса творчества, а также и непредсказуемость его результата, особенно, например, в искусстве. Большинство феноменов и операций, относящихся к творчеству, не поддаются формализации. Иногда, правда, утверждают, что качество непредсказуемости операций, имитирующих творчество, присуще и машинному творчеству. Так, можно ввести в программу некую мутацию, способную на выходе дать результат, который будет неизвестен и самому программисту. Однако здесь следует сделать ряд смысловых уточнений. Во-первых, «непредсказуемый» отнюдь не всегда означает «творческий»; так, могут быть непредсказуемы спонтанные проявления разрушительной энергии (и природы, и человека). Во-вторых, подобный «непредсказуемый» результат есть следствие не творчества, а как бы сжатия времени, на которое способна машина, выдающая этот «непредсказуемый» результат, аналогичный тому, который выдал бы человек, перебирающий разнообразные варианты в течение не секунд, а сотен лет. Иными словами, это опять феномен скорости, но не творчества.

И, наконец, в-третьих, согласно известному «критерию Тьюринга», идентификацию можно формализовать: если, например, машина выполняет функции человека, то она вполне человек. Но даже исходя из подобного функционального подхода, нельзя признать за машиной способности исполнения функций человека: творческие функции, принципиально не формализуемые, не могут быть адекватно и в полной мере осуществлены машиной, ибо, как уже говорилось выше, обеспечением этих функций у человека являются такие свойства его психики, которые не могут быть ни формализованы самим человеком, ни «освоены» машиной, ни продублированы ею.

В самом деле, творчество есть, пожалуй, самый сложный, самый значительный, самый необъяснимый и самый интересный по своим проявлениям и результатам феномен во Вселенной. Если вспомнить утверждение Платона, что творчество — это то, что вызывает переход из небытия в бытие, то под этим в равной степени можно подразумевать и некий уникальный процесс (в том числе и процесс самого возникновения мира), и самого Бога. Формально творчество определяют как внесение в мир нового, небывшего, каковым может быть не только предмет или произведение, но и новая интерпретация, новое прочтение, новое видение и т.п. Однако в творчестве присутствует и некий метафизический план, в котором оно связано с самим становлением человека, его утверждением как рода и предназначением в Космосе.

В природе наряду с силами энтропийного разрушения работают, как известно, и силы самоорганизации, совокупная деятельность которых приводит к становлению сложных систем. Однако живому с

его недолгими сроками жизни не приходится уповать на самоорганизацию, действие которой проявимо лишь в масштабе чрезвычайно больших исторических периодов. Чтобы выжить, живое развивает формы деятельности, способные к упорядочению окружающего. Разумное и сознательное живое вырабатывает такие формы поведения, которые помогают адаптироваться в мире и сохраниться. Творчество культуры есть вершина разумного поведения живых организмов, объединенных в сообщества, способные к высшим формам адаптации. С известной долей упрощения можно сказать, что творчество становится для человека как бы *заменой* той самоорганизации, которая возможна в природе, но на социальном уровне развития выступает в форме культуры. Действительно, с одной стороны, только культура способна обеспечить человеку качественно новый и эффективный способ сохранения и трансляции накопленного опыта, но, с другой стороны, кто видел хотя бы один культурный продукт, возникший сам собой в процессе своей самоорганизации?

Стремление к развитию, совершенствованию, самореализации, выступающее частью самой природы человека, определяет творчество не только как способ его деятельности, но как общий *модус* существования и поведения в мире человека, который осознал эту свою «особость» в мире. Таким образом, творчество, будучи способом утверждения человеческого рода, выступает как своего рода высший инстинкт, оформляющийся у него в целях сохранения, совершенствования и развития. Именно в этом смысле и можно рассматривать творчество как высший уровень сложно организованного поведения. Однако мотивации его и сам творческий процесс не всегда поддаются формализации, ибо включают в себя множество «внесемантических» элементов, содержание которых отнюдь не всегда может быть адекватно представлено в вербализованной форме.

Конечно, можно сказать, что и художник своей творческой работой вызывает к бытию то, что в линиях, звуках, красках, формах уже как-то существует в мире, проявляя всё это из скрытого состояния в явленное, вызывая это из в определенном смысле небытия. Однако, во-первых, роль художника при этом отнюдь не пассивно-исполнительская: он должен увидеть *возможность* такого проявления, должен знать, *как* это делается, и обладать соответствующим умением, он, наконец, должен проявить *волю к реализации* этой цели. К этому сложному процессу на разных его этапах оказываются подключены такие способности человека, как воображение, антиципация, спонтанность, интуиция, вдохновение. Поэтому, во-вторых, художник всегда нечто «добавляет» к используемому материалу, который он не просто организует, но *претворяет* в соответствии с творческим замыслом, задачами, качествами и возможностями собственного духовно-личностного материала; недаром утверждают, что произведе-



ние искусства больше факта действительности на величину души художника.

В силу всего этого сложного состава творческой деятельности и еще более сложного комплекса целе-мотивационных его факторов вопрос о возможности полноценного машинного творчества выглядит весьма проблематично, если не мистифицировать возможности машины и не подменять объяснение процесса «мифом по ассоциации». Если в науке, где информация самоценна, машина может взять на себя какие-то определенные функции, то в сфере искусства, где столь большую роль играют процессы эмоционально-ассоциативного плана, апеллирующие к специфически человеческой природе, ее творческие возможности сомнительны.

Когда говорят о творчестве, то прежде всего упоминают, что осуществление творчества свободно. Оно свободно и в том смысле, что, как известно, только человек производит «даже будучи свободен от физической потребности, и в истинном смысле только тогда и производит, когда свободен от нее»<sup>5</sup>. Оно свободно и в том отношении, что к творчеству нельзя принудить. Его процесс нельзя включить, например, нажатием кнопки — ни по желанию, ни по заказу, ни из чувства долга. И в творчестве нельзя притвориться: в сотворенном им творце может предьявить только то, чем он располагает или чем является. В творчестве он адекватен себе и своей природе, адекватен своей сущности, своей целостности, ибо может осуществлять свободное перемещение в пространстве открытых его духу смыслов. Однако можно ли о машине сказать, что она творит *сама*, по собственному почину, по своему *свободному* волеизъявлению? Творчество (творческое поведение) иногда может быть вызвано той или иной необходимостью, но *обеспечено* оно может быть только внутренним импульсом; когда его нет, никакие внешние обстоятельства не смогут «вызвать» ни процесса творчества, ни творческого состояния. Творческий человек — это человек с развитым творческим импульсом, обусловленным развитой творческой потребностью. Представляется, что все эти слова не применимы к машине как предполагаемому субъекту творчества.

Кроме того, как уже говорилось, в творчестве создается нечто небывшее, не имеющее прецедентов существования. По мнению И.Канта, творчество есть создание нового, для которого еще не существует какого-либо правила создания<sup>6</sup>. В то же время, если известно, как осуществляется то или иное действие, или как нечто делается (т.е. это делалось уже ранее, возможно, неоднократно, и существует определенная технология, алгоритм этого действия), то это уже не столько творчество, сколько *технология*. Собственно творчество, по выраже-

<sup>5</sup> Маркс К., Энгельс Ф. Из ранних произведений. М., 1956. С. 566.

<sup>6</sup> См.: Кант И. Соч. в 6 т. Т. 5. М., 1966. С. 323–324.

нию Г.Марселя, в сути своей метатехнично; сферой метатехники может быть лишь человеческое, программирующее, но не машинное, запрограммированное, сознание<sup>7</sup>. В данном пункте, правда, существуют определенные расхождения относительно того, что считать собственно творчеством и обязательно ли в творчестве создание именно нового. Полагают, например, что творчеством может быть любая деятельность, к которой человек способен подойти творчески, т.е. творчески себя в ней ощутить, по-новому прочувствовать старое, вообще что-то пережить впервые и т.п. Здесь, на наш взгляд, проблема новизны в творчестве переносится из сферы его онтологии в область психологического и когнитивного. Когда человек нечто открывает для себя в том, что может быть давно открыто и известно, то это вопрос его личных ощущений и переживаний, т.е. вопрос индивидуальной психологии, а не собственно онтологии творчества. При этом не только происходит методологическое смешение онтологического, психологического и когнитивного уровней анализа, но и анализ собственно феномена творчества часто подменяется феноменологическими иллюстрациями.

В процессе познания также, безусловно, может присутствовать не только творческая составляющая, но и возможность открытия субъективно нового, однако при этом отнюдь не всегда достигается создание нового знания, что, конечно, не умаляет ценности самого индивидуального познавательного усилия. Конечно, компьютер может создать относительно новое — комбинацию, сочетание тех или иных элементов из имеющихся в его памяти. Однако создание истинно нового, для которого нет соответствующего алгоритма, — это всё-таки особенность и преимущество живого мозга, и оно, как правило, означает известное подключение *спонтанности*, наличие которой не предполагается у машины, даже самой умной. Пожалуй, единственная спонтанность, которая может быть доступна разумной машине, это включение «реакции» самосохранения при угрозе ее существованию; другого *содержания* у машинной спонтанности быть не может. В целом, представляется, что в вопросе об уме машины происходит невольная мистификация достаточно формального предмета, объясняемая энтузиазмом; ум машины — это всё-таки всего лишь чётко составленная и технически безукоризненно реализуемая программа, основанная на заложенной в памяти способности воспринимать и фиксировать тончайшие сигналы, распознавать их и «ассоциировать» с теми или иными действиями, заложенными в «памяти».

В отличие от машины, которая, как уже отмечалось, может наиболее продуктивно работать с однозначно, четко определенными поня-

---

<sup>7</sup> Даже здесь, в различии активного и пассивного залога проявляется это кри- териальное требование к творчеству, которое возможно только в активном залоге.

тиями, человеку противопоказан излишне жесткий порядок в организации мыслительного процесса, ибо своеобразная психическая инерция, которая формируется при этом, препятствует образованию новых неожиданных сочетаний, понятий, приемов, смыслов. Строгая упорядоченность элементов информации мешает образованию новых порядков, подавляет воображение. Неопределенность, непредсказуемость, т.е. известная доля энтропийности является необходимой принадлежностью всякого творческого процесса, когда в состоянии свободного поиска, включающего игру воображения, человек открывает новое, поскольку его ищущий дух всегда недоволен существующим, наличным, он всегда хочет лучшего, большего, он всегда в поиске совершенства (которого не находит) и устремлен к нему. Как отмечал Н.Бердяев, «человек есть существо, недовольное самим собой и способное себя перерастать»<sup>8</sup>.

Материалом для творчества у человека является обычно совокупность его широко и разносторонне понимаемого опыта. Но, с другой стороны, человек, если нужно, умеет и отказаться от него, уйти от навязанного этим опытом мыслительно-поведенческого стереотипа, ставшего непригодным в новых условиях, — и принять нешаблонное решение в пользу неизвестного, неизведанного. Трудно предположить, что у машины может существовать, с одной стороны, нечто похожее на подобный жизненный опыт, с другой стороны — своего рода диалектическая способность выбора между «использовать» или «пренебречь»? Трудно также предположить наличие у машины достаточно сложно организованной системы многогранной, самостоятельной и эмоционально окрашенной системы *мотиваций* для подобных действий.

Что же касается содержания самого человеческого опыта, то оно складывается на основе впечатлений мира — таких и так отобранных, — как это соответствует запросам и интересам познавательной деятельности данного человека. Особенности *личности* при этом имеют особенно важное значение в художественном творчестве, где личностные компоненты являются неременной составляющей создаваемого творческого продукта. Художественная реальность формируется из притока живых ощущений, впечатлений творческой работой духа. Машина пока не может предоставить аналога подобного единства духа и плоти. Известны случаи создания так называемых электронных Маугли (в США) как симбиоза виртуальной личности и ее машинного носителя, однако у подобной личности происходит развитие лишь манипулятивно-операторского компонента «ума» при отсутствии духовно-творческих составляющих. Подобная личность вполне может быть признана машиноравной, о достижении чего меч-

---

<sup>8</sup> Бердяев Н.А. О назначении человека. М., 1993. С. 55.

тал герой антиутопии Е.Замятина «Мы»: такой машиноравный человек представляется весьма оптимальным в плане своего прагматического функционирования: он не будет отрывать от работы драгоценное время на бесполезные мечты, не будет страдать от любви, ухудшая тем самым свои трудовые кондиции.

Основой способности человека к творческому поведению является способность его сознания к *антиципации*, лежащей в самой природе отражательной деятельности человеческой психики. Антиципация — это способность действовать с упреждением в отношении ожидаемых или возможных событий. В этом смысле антиципация есть чисто человеческое качество. Так, например, еще английский философ Томас Гоббс писал, что человека, в отличие от животного, делает голодным уже грядущий голод, который и побуждает его заботиться о возможности его утоления, т.е. охотиться, работать и т.п.<sup>9</sup>

Данные науки подтверждают, что антиципация как способность предвидеть будущее и овладевать им есть цель высокоорганизованного мозга, выражением которой и является высшая форма антиципации — творчество. Именно способность к антиципации обеспечивает возможность формулирования цели (т.е. будущего результата) до начала действий по ее достижению, планирования (построения программы действий на будущее, чтобы обеспечить нужный результат) и прочих способов «работы» с будущим. Без этой способности нельзя ни создать художественного произведения, ни организовать любую культурную деятельность, устремленную к получению будущего результата. Человек должен мысленно увидеть результат, чтобы начать осознанно действовать для его достижения. В этом смысле всё быстрое действие ЭВМ разворачивается тем не менее всегда только *post factum*; человек же принципиально и конститутивно способен к реакции опережения.

В свою очередь, антиципирующее «забегание» работы мозга, как и вся сложно организованная и осуществляемая деятельность творчества в целом, невозможны без воображения — этой способности человека мысленно создавать, воплощать в идеальной сфере свои представления, образы, мечты, идеалы. Без активности воображения невозможно ни создание, ни существование и развертывание идей, замыслов, символов, на которых строится всякая мыслительная деятельность — и в науке, и в искусстве.

Таким образом, для аутентичного осуществления творчества необходимо как умение создавать некоторое субъективное идеальное пространство, где происходит работа с разного типа психическими конструкциями, имеющими сложный механизм образования и опередмечивания, так и осознание наличия определенного метафизиче-

---

<sup>9</sup> См.: Гоббс Т. Избр. произведения. М., 1964. Т. 1. С. 235.

ского пространства, где «содержатся» или формируются идеи, ценности, смыслы. Нужно, чтобы человек умел входить в эти условные пространства и взаимодействовать с ними. Нужно, чтобы он обладал и умел пользоваться своим воображением, интуицией, способностью к спонтанному проявлению всех граней своего творческого потенциала. Многие специалисты признают, что продублировать подобную психическую деятельность в машине пока однозначно невозможно.

Интеграционная деятельность мозга, включающая обработку специализированной (рациональной, вербально-понятийной и эмоциональной, чувственно-конкретной, логически выстроенной и образно представленной, ориентировочной пространственно-временной и локально идентификационной) информации, поступающей из функционально асимметричных правого и левого полушария, позволяет человеку сочетать эти разные способы восприятия действительности; при этом он получает комплексное представление о мире, которое не распадается на две разные, «конкурирующие» между собой модели, но дает именно единую и адекватную картину мира, возникающую целостно и одновременно. Сотрудничество обоих полушарий обуславливает возможность протекания творческой деятельности, а характер индивидуально представленного различия между ними, сама величина этого различия определяют потенциал, «масштаб» и качество творческой одаренности личности, оригинальность и продуктивность ее творчества, преимущественную направленность ее творческой деятельности.

В целом можно сказать, что главным содержанием творческого проявления человека является то, что в нем осуществляется специфически протекающая и выражаемая *поисковая* деятельность способного к *саморазвитию* мозга, побуждаемого к этому человеческими эмоциями, которые известный ученый П.В.Симонов считал ее двигателями, а другой известный ученый, П.К.Анохин, «пеленгами поведения». Эмоции — от простейших до сложнейших их форм — неизменно окрашивают человеческое восприятие мира и выступают в сфере мотивации деятельности. И хотя мозг человека можно в определенном смысле уподобить точной магнитофонной ленте, на которой с детства записывается всё увиденное, услышанное и пережитое, однако, по свидетельству В.П.Эфроимсона, вспоминаются и подключаются (актуализируются) лишь «те сенсорные элементы, на которые пациент обращал внимание, выделяя их из массы других впечатлений»<sup>10</sup>. Иными словами, эмоциональный фон «записи в память» того или иного вида опыта имеет для человека безусловное значение, тогда как относительно машины об эмоциях говорить не приходится.

---

<sup>10</sup> Эфроимсон В.П. Генетика этики и эстетики. М., 2004. С. 87.

Автомат проигрывает человеку и в случае необходимости включения интуиции (в ситуации нечетко определенных понятий, в ситуации дефицита исходных данных и т.п.). Он проигрывает там, где надо не просчитать, а именно понять и открыть. Ибо вся база данных автомата человека просчитывается тем, что известно человеку-программисту. База данных человека как такового — всё практически неисчерпаемое, безграничное информационное пространство Земли (Космоса). Его «объем» невозможно, подобно компьютеру, «перебрать», чтобы найти нужное. Здесь необходим мгновенный интуитивный прорыв, необходимо то, что в разных практиках называют вдохновением, откровением, озарением, просветлением...

Когда (и если) автомат «научится» подключаться к информационному полю Земли, тогда человек и автомат сначала окажутся *равны*, потом человек *проиграет*. Но сможет ли человек заложить в автомат подобную программу подключения, если она пока и для него самого не поддается точной формулировке? Сможет ли он научить машину *вдохновению*? Человек сам далеко не всегда может достичь этого состояния, и издавна пытался использовать различные средства и особые психотехники для выхода на этот особый канал связи с Высшими уровнями бытия (Богом, сверхсознанием). В религиозных системах это право выхода неизменно связывали с соответствием человека определенным нравственным требованиям, ибо расширение сознания открывало человеку и новую истину, и новые возможности. В восточных системах для достижения этого особого состояния сознания разработали специальную комплексную систему психотехник со статическими и динамическими способами медитации, технику трансцендентальной медитации, интегральную йогу и т.п. В современных методиках, используемых, например, в американской трансперсональной психологии, с этой целью используют не только специальные психотренинги (опирающиеся в том числе на современные психо-лингвистические разработки), но и некоторые наркотические (психоделические) средства (Т.Лири, С.Гроф, Ч.Тарт и др.). Можно ли представить себе медитирующую ЭВМ или робота-наркомана? Кстати, методы трансперсональной психологии открывают человеку и тайны его бессознательного, где также, по свидетельству, например, Ф. Достоевского или Ю. Мисимы, скрыты глубинные истоки как творческих образов, так и самих творческих потребностей. Доводилось ли даже самому прилежному юзеру общаться с ЭВМ, обремененной так называемыми комплексами?

А что могут означать, например, нравственные требования применительно к «творчеству» машины? Если для *машинного* творчества данные понятия не имеют смысла, означает ли это, что следует вообще отказаться от определенных религиозных и вообще каких-то традиционных культурных представлений, связанных с творчеством, ради при-

знания таковым творчеством машинного? И вообще столь серьезное отношение к пребыванию в состоянии творчества говорит о значимости этого особого состояния сознания и для человека, и для оценки содержания его творения. Не происходит ли уплощения и схематизации самого понимания сознания, когда мы говорим о машинном творчестве? Что вообще может означать для машины такое понятие, например, как «измененное состояние сознания» (*измененное* — по сравнению с чем? измененное как? по каким параметрам? наконец, кем? *состояние* — в какой форме оно фиксируется? *сознания* — что такое само сознание для ЭВМ?). Либо творчество не есть особое состояние сознания (коль оно свойственно и «творящему» компьютеру), либо (если оно таковым всё-таки является) оно компьютеру недоступно, как, в этом случае, и само собственно творчество. Человек может говорить о своем сознании, когда осознаёт себя сознающим. Иначе говоря, сознание есть живой диалог между мозгом-деятелем и мозгом-наблюдателем деятельности. Простейший пример этой своего рода двусубъектности сознания — хрестоматийное восклицание А.С.Пушкина: «Ай да Пушкин...». Можем ли мы обнаружить у компьютера наличие чего-либо подобного такой двусубъектной структуре?

Таким образом, способность к творческому поиску, обусловленная как определенными возможностями мозга, так и возможностью достижения особого состояния сознания (что в большинстве религиозных систем связано с необходимым следованием человека определенным нравственным регулятивам), определяют эту «специфику» человека, пока всё-таки отличающую его от машины. С известными оговорками (в отношении определения интеллекта) можно согласиться с тем, что ЭВМ обладает способностью быстро и точно совершать операции с информацией, что формально можно определить как интеллект. Но в определение интеллекта не входит безоговорочно способность к творчеству, которое, в свою очередь, отнюдь не сводимо к операциям и манипуляциям с информацией. Даже очень разумные технологии, будучи продуктом творчества, сами как таковые не являются творчеством. Иными словами, если еще можно представить себе функционального робота-солдата, то вообразить себе робота-поэта всё-таки невозможно.

Человек, в отличие от машины, которая способна лишь к более или менее связанной компиляции данных, способен к конструированию такой сложной системы, как новое *целостное* мироощущение (что, как правило, сопровождается переходом от одной эпохи к другой, например, от Средневековья к Возрождению), когда творческий разум, а не просто препарирующий и оперирующий данными рассудок, восходит к видению нового единства в понимании мира, существующего в совокупности разных способов его понимания и представления, в *синтезе* научных, философских, религиозных знаний. Целостность ми-

роощущения может быть достигнута только целостным человеком, способным соединить разные грани и разные уровни понимания действительности, создать непротиворечивую картину из значительно расходящихся индивидуальных видений и оценок. Только живой, способный к ассоциациям и синтезу человеческий мозг, обладающий даром интуиции, воображения и прозрения, может выстроить комплексный образ мира, во многом скрытый от человека из-за недостатка его знаний, но конструируемый, воссоздаваемый на основе наличествующей у человека способности охватывать своим сознанием то, что пока еще не существует (потенциальные возможности, тенденции, перспективы).

Человеческое творчество существует как одновременно и высшая духовная потребность, и как интенсивная психоментальная деятельность, при этом, как уже говорилось выше, в нем неизменно и мощно присутствуют энергии самоутверждения, потребность в самореализации. О какой потребности в самореализации, в самоутверждении можно говорить применительно к машине? В этом отношении творчество может быть деятельностью только человека со всеми сопутствующими этому и позитивными, и негативными проявлениями (например, патологическими утверждениями) в психологическом и социальном планах.

В целом можно сказать, что особенностью творчества как *человеческого* феномена является наличие совокупности следующих качеств или характеристик:

- наличие *потребности* в творчестве, причем эта потребность осознаётся как личная потребность, независимо от уровня реализации;
- *свободное* следование своей потребности и свобода ее реализации, отсутствие внешнего принуждения;
- *мотивированность* творческой деятельности, включая высшие духовные мотивации и мотивы самореализации и самоутверждения;
- *эмоциональность* как окраска мотивации, как условие и как фон творчества;
- *целенаправленность*, ибо творец в принципе знает, что он намерен создать;
- создание *ценности* как результата творческой деятельности, реализующей человеческую потребность.

В самом деле, человеческое творчество создает именно ценности, а не просто вещи, ибо за бытием творческого продукта разворачивается целый глубинный пласт смыслов, обеспечиваемый душой художника, его позицией, уровнем мотивации, спецификой индивидуального видения мира, в котором произошло преломление всей совокупности культурного опыта. Хотя в наше прагматичное время всё труднее говорить о подлинных ценностях, которые всё чаще подменяются техническими иллюзиями или имитациями, позволяющими камуфлировать отсутствие этого культурного пласта значений и смыслов.



И хотя в функционировании ЭВМ не удастся выделить ни личной заинтересованности, ни специальной мотивации ее деятельности, ни целенаправленности, ни даже инстинкта самосохранения, тем не менее в настоящее время всё чаще исследователи говорят о том, что наряду с биосферой или антропосферой формируется подобного же типа автономная реальность, именуемая техносферой. В техносфере, утверждают они, возможно даже осуществление *самостоятельной эволюции*, способной породить умные машины, которые могут вытеснить человека, что, скорее всего, и последует.

Однако при этом выпускается из виду, что техносфера всё-таки, в отличие от биосферы, например, есть *искусственное*, а не естественное образование, и, будучи искусственным образованием, она подлежит управлению не космическими законами, но законами создавшей его человеческой деятельности (даже если ее результаты грозят иногда выйти из-под контроля самого человека). Действительно, Бог (или, в другой парадигме, Космос) создал земную природу и человека; но ни Бог, ни Космос не создавали ни троллейбусов, ни компьютеров — их создал уже сам человек. Согласно законам космической структуризации, созданное управляется создателем: Космос — управляет на уровне биосферы, человек — на уровне техники. Космос не может непосредственно, минуя человека, управлять эволюцией техники, как он это «делает» относительно той же биосферы, создавая соответствующие (или несоответствующие) условия для ее жизнедеятельности, и не содержит законов или принципов для организации функционирования техносферы. И потому перенос некоторых понятий, характеризующих жизнедеятельность биосферы, на уровень функционирования техники неправомерен. Иными словами, ни Бог, ни Космос не снисходят до формулировки законов управления «эволюцией» троллейбусов и процессоров. Точно так же понятие творчества как космического принципа не относимо к машине как продукту технического развития человека, и ей доступно лишь ограниченное число форм *условно* «творческого» участия в человеческой деятельности.

Человек как космопланетарный феномен есть *естественное* порождение Космоса, и его естественность проявляется, в частности, в том, что он смог возникнуть и существовать только при схождении достаточно узкого набора из числа характеризующих жизнь Космоса фундаментальных физических констант. Само же человеческое творчество, если иметь в виду творчество в собственном смысле этого слова, есть хотя и особый, но также являющийся и остающийся естественным человеческий феномен — человеческий способ выражения космического принципа сотворения и развертывания бытия.

Сама способность человека к творчеству, будучи истинно человеческой способностью, означает новый уровень приспособления к действительности, новый уровень и смысл самих отношений с ней. Творче-

ское поведение в широком смысле выступает как умение быстро и правильно реагировать в непредсказуемых ситуациях. В этом отношении именно творчество является в подлинном смысле слова *адекватным* поведением в мире, который постоянно находится в движении, непрестанно изменяется. Такие свойства живой психики, как способность к антиципации, как воображение, интуиция, спонтанность, способные явить человеку всю неисчерпаемость его сознания, устремленного к сверхсознанию, а своими основаниями скрытого в бессознательном, определить богатство творческого переживания, обеспечить сложность непосредственного творческого процесса, — все эти свойства живого вряд ли когда станут доступны машине.

Поэтому остаётся констатировать, что творчество в собственном смысле по-прежнему может существовать лишь как человеческий феномен, как общий продукт и чрезвычайно сложно устроенного живого мозга, и одновременно живой плоти с ее вкусом к живым ощущениям многоцветного бытия. Именно эта двойная природа человека, обусловившая и устремления духа, и переживания плоти, — источник напряжения, из которого возникает искра творчества.

И потому принципом адекватного смыслового и методологического различения в оценках естественного и искусственного интеллекта остаётся правило, которое было сформулировано еще А.И.Бергом и которое можно перефразировать приблизительно так: человеку — человеку, процессору — процессорово. Представляется, что подобный человеческий «шовинизм», во-первых, *объясним*, ибо наука в целом, на которую в свое время, как известно, возлагались большие надежды, в конечном счете лишь бесконечно отдалила человека от природы (в том числе и его собственной); во-вторых, подобный шовинизм в известной мере и *оправдан* (необходимостью защиты живого, всё более теснимого технико-цивилизационным) — с точки зрения широко понимаемой экологии: природно-биологической, психо-духовной, культурной. Ведь главная проблема в том, чтобы к тому времени, когда человек, может быть, наконец научит машину творческому вдохновению, он сам бы не утратил этой способности.

## ОБ АВТОРАХ

**АЛЕКСЕЕВ Андрей Юрьевич** – кандидат философских наук, учёный секретарь Научного совета по методологии искусственного интеллекта РАН

**АНТИПЕНКО Леонид Григорьевич** – кандидат философских наук, старший научный сотрудник Института философии РАН

**БАХТИЗИН Альберт Рауфович** – кандидат экономических наук, старший науч. сотрудник Центрального экономико-математического института РАН

**ВАСЮКОВ Владимир Леонидович** – доктор философских наук, ведущий научный сотрудник Института философии РАН

**ГЛАЗУНОВ Виктор Аркадьевич** – доктор философских наук, доктор технических наук, профессор, главный научный сотрудник Института машиноведения РАН им. А.А. Благонравова

**ДУБРОВСКИЙ Давид Израилевич** – доктор философских наук, профессор, ведущий научный сотрудник Института философии РАН, заместитель председателя Научного Совета по методологии искусственного интеллекта РАН

**ЗОТОВ Анатолий Федорович** – доктор философских наук, профессор философского факультета МГУ им. М.В.Ломоносова

**ИВАНИЦКИЙ Алексей Михайлович** – доктор медицинских наук, член-корреспондент РАН, главный научный сотрудник Института высшей нервной деятельности и нейрофизиологии РАН

**ИГНАТЬЕВ Михаил Борисович** – доктор технических наук, профессор, заведующий кафедрой вычислительных систем и сетей Санкт-Петербургского государственного университета аэро-космического приборостроения

**КАТРЕЧКО Сергей Леонидович** – кандидат философских наук, доцент МГУ им. М.В. Ломоносова

**КОЛИН Константин Константинович** – доктор технических наук, профессор, главный научный сотрудник Института проблем информатики РАН

- 
- КОЧЕРГИН Альберт Николаевич** – доктор философских наук, профессор МГУ им. М.В. Ломоносова
- КУЗИЧЕВ Александр Сергеевич** – кандидат физико-математических наук, старший научный сотрудник МГУ им. М.В. Ломоносова
- ЛЕКТОРСКИЙ Владислав Александрович** – доктор философских наук, член-корреспондент РАН, Главный редактор журнала "Вопросы философии", заведующий отделом Института философии РАН, заместитель председателя Научного Совета по методологии искусственного интеллекта РАН
- МАКАРОВ Валерий Леонидович** – академик, директор Центрального экономико-математического института РАН, Академик-секретарь Отделения общественных наук РАН, член Совета при Президенте Российской Федерации по науке, технологии и образованию, председатель Научного Совета по методологии искусственного интеллекта РАН
- МЕРКУЛОВ Игорь Петрович** – доктор философских наук, профессор, заведующий сектором эволюционной эпистемологии, заместитель директора Института философии РАН
- МИРОНОВ Владимир Васильевич** – доктор философских наук, профессор, проректор МГУ им. М.В.Ломоносова, декан философского факультета МГУ им. М.В.Ломоносова
- МЦХВЕТАДЗЕ Арчил Владимирович** – доктор медицинских наук, профессор, ведущий сотрудник Института психологии АН Грузии
- НАРИНЬЯНИ Александр Семёнович** – кандидат физико-математических наук, научный руководитель Российского НИИ искусственного интеллекта
- НИКИТИНА Елена Александровна** – кандидат философских наук, доцент МИРЭА
- ОГУРЦОВ Александр Павлович** – доктор философских наук, профессор, заведующий отделом Института философии РАН

---

**ПАВЛОВ Сергей Афанасьевич** – кандидат философских наук, старший научный сотрудник института философии РАН

**ПАНОВ Олег Владимирович** – Научный сотрудник объединения «Магнум-АРС»

**РЕДЬКО Владимир Георгиевич** – доктор физико-математических наук, заместитель директора Института оптико-нейронных технологий РАН

**РОЗИН Вадим Маркович** – доктор философских наук, профессор, заведующий сектором философии техники Института философии РАН

**САМОХВАЛОВА Вера Ильинична** – доктор философских наук, профессор, ведущий научный сотрудник Института философии РАН

**СКВОРЦОВ Алексей Алексеевич** – кандидат философских наук, доцент МГУ им. М.В. Ломоносова

**СМИРНОВА Наталия Михайловна** – доктор философских наук, профессор, ведущий научный сотрудник Института философии РАН

**ХОЛОДНАЯ Марина Александровна** – доктор психологических наук, профессор, заведующая лабораторией психологии способностей Института психологии РАН

**ЧЕРНИГОВСКАЯ Татьяна Владимировна** – профессор, доктор биологических наук, Заведующая отделом общего языкознания и лаборатории когнитивных исследований ИФИ СПбГУ, г. Санкт-Петербург

**ЧИСТОХОДОВА Людмила Ивановна** – доктор психологических наук, кандидат экономических наук, директор Института экономики, управления и права

**ЮЛИНА Нина Степановна** – доктор философских наук, ведущий научный сотрудник Института философии РАН

---

## СОДЕРЖАНИЕ

От редакторов .....	3
<b>I. Методологические и эпистемологические вопросы .....</b>	<b>5</b>
<i>Макаров В.Л.</i> Получение нового знания методом компьютерного моделирования .....	5
<i>Лекторский В.А.</i> Философия, искусственный интеллект и когнитивная наука .....	12
<i>Зотов А.Ф.</i> Роль феноменологии в организации междисциплинарных исследований в области искусственного интеллекта .....	22
<i>Огурцов А.П.</i> Возможности и трудности в моделировании интеллекта.....	32
<i>Нариньяни А.С.</i> Между Знанием и Незнанием – Наивная Топография 2 .....	49
<b>II. Сознание, мозг, искусственный интеллект .....</b>	<b>75</b>
<i>Дубровский Д.И.</i> Сознание, мозг, искусственный интеллект .....	75
<i>Иваницкий А.М.</i> Физиологические основы сознания и проблема искусственного интеллекта .....	90
<i>Михветадзе А.В.</i> Естественный и искусственный интеллект. Некоторые аспекты сходства и различия .....	100
<i>Панов О.В.</i> Функциональная структура бессознательного и возможность формирования новых принципов искусственного интеллекта .....	107
<i>Антипенко Л.Г.</i> Квантовый компьютер и квантовый мозг .....	111
<i>Черниговская Т.В.</i> Зеркальный мозг, концепты и язык: цена антропогенеза .....	127
<b>III. Естественный и искусственный интеллект .....</b>	<b>149</b>
<i>Холодная М.А.</i> Структура и функции естественного интеллекта в контексте проблемы искусственного интеллекта.....	149
<i>Смирнова Н.М.</i> Трансцендентальная интересубъективность, проблема «чужих сознаний», искусственный интеллект .....	163
<i>Юлина Н.С.</i> Улучшение качества мышления: «сократический диалог» и интеракция с компьютером .....	180

<i>Розин В.М.</i> Как в настоящее время можно осмыслить концепцию искусственного интеллекта?.....	194
<i>Кочергин А.Н.</i> Искусственный интеллект, психика, творчество .....	209
<i>Алексеев А.Ю.</i> Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга?.....	223
<b>IV. Проблемы моделирования .....</b>	<b>243</b>
<i>Редько В.Г.</i> Задача моделирования когнитивной эволюции .....	243
<i>Глазунов В.А., Чистоходова Л.И.</i> Моделирование творческого процесса.....	259
<i>Игнатъев М.Б.</i> Мир как модель внутри сверхмашины и виртуальные миры .....	264
<i>Катречко С.Л.</i> Трансцендентальная (кантовская) модель сознания как новая парадигма «искусственного разума» .....	276
<b>V. Логические и математические вопросы .....</b>	<b>289</b>
<i>Меркулов И.П.</i> Эпистемологический анализ когнитивной природы математических и логических формализмов.....	289
<i>Васюков В.Л.</i> Формальная онтология и искусственный интеллект.....	305
<i>Павлов С.А.</i> О логике символьных выражений для искусственного интеллекта .....	325
<i>Кузичев А.С.</i> Программа Колмогорова, интеллектуальные системы и теоремы Гёделя о неполноте .....	330
<i>Бахтизин А.Р.</i> Искусственный интеллект и его приложения в экономико-математических моделях .....	347
<b>VI. Социокультурные аспекты .....</b>	<b>379</b>
<i>Колин К.К.</i> Становление информационного общества в России и национальная безопасность.....	379
<i>Миронов В.В.</i> Информационное пространство: диалог культур.....	392
<i>Никитина Е.А.</i> Антропологический поворот в искусственном интеллекте.....	399
<i>Скворцов А.А.</i> Мироззренческие и теоретические основы этических исследований виртуальности.....	411
<i>Самохвалова В.И.</i> «Человеческое, слишком человеческое», или процессор в экстазе.....	423
Об авторах .....	442

# ИЗДАТЕЛЬСТВО ИИНТЕЛЛ

специализируется  
в публикации  
литературы по  
когнитивным и  
компьютерным  
наукам,

по теоретическим и философско-методологическим вопросам  
искусственного интеллекта.

В ИЗДАТЕЛЬСТВЕ «ИИНТЕЛЛ» ВЫШЛИ В СВЕТ  
КНИГИ

1. **Новое в искусственном интеллекте. Методологические и теоретические вопросы.** Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2005. – 280 с.
2. **Искусственный интеллект: междисциплинарный подход.** Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2006. – 448 с.
3. **Методологические и теоретические аспекты искусственного интеллекта.** Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексева - М.: МИЭМ-ИИНТЕЛЛ, 2006. – 192 с.
4. **Тест Тьюринга, зомби, роботы.** Пер. с англ. Под ред. А.Ю. Алексева - М.: МИЭМ-ИИНТЕЛЛ, 2006.– 112 с.
5. **Искусственный интеллект: философия, методология, инновации.** Материалы Первой Всероссийской конференции студентов, аспирантов и молодых учёных. Под ред. Д.И. Дубровского и В.И. Самохваловой - М.: МИРЭА-ИИНТЕЛЛ, 2006. – 320 с.

По вопросам издания и приобретения книг:

Адрес: <http://aintell.ru>

Электронная почта: [book@aintell.ru](mailto:book@aintell.ru)



Научное издание

**Искусственный интеллект:  
междисциплинарный подход**

Под редакцией доктора философских наук, профессора  
*Д.И. Дубровского* и доктора философских наук и члена-  
корреспондента РАН *В.А. Лекторского*

Оригинал–макет: М.А. Сёмочкин

Художник: А.А. Лобков

Технический редактор: Т.А. Кураева

Корректор: Д.А. Алексеев

Подписано в печать с оригинал-макета 14.03.2006 г.  
Формат 60х90/16. Печать офсетная. Гарнитура Таймс.  
Усл.печ.л. 28. Уч.-изд.л. 32. Тираж 1000 экз.  
Заказ № \_\_\_\_\_

Отпечатано в с готового оригинал-макета в  
ФГУП «Производственно-издательский комбинат ВИНТИ»,  
140010, г. Люберцы Московской обл.,  
Октябрьский проспект, 403. Тел. 554-21-86