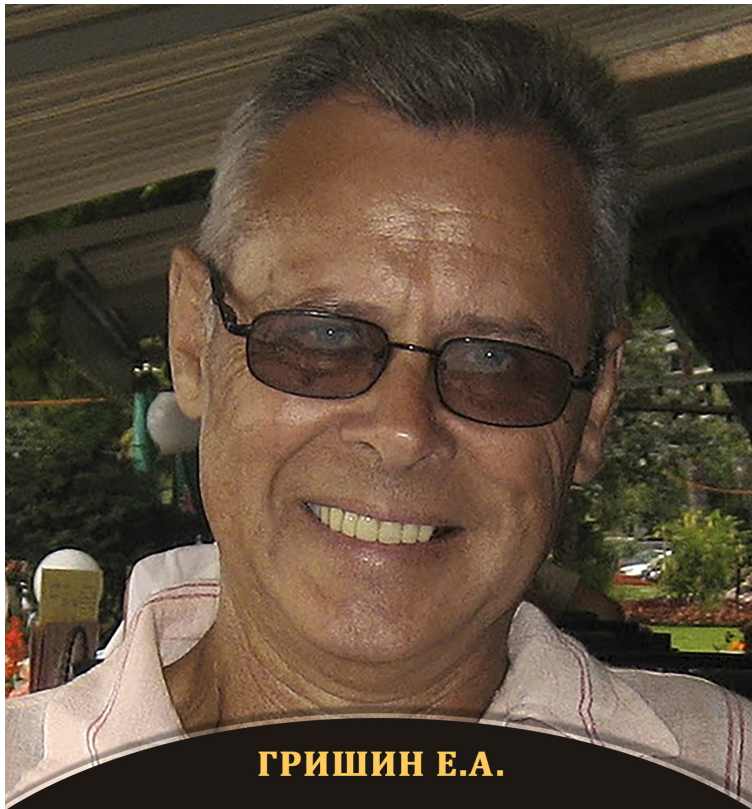




ГРИШИН Е.А.

РЕФЛЕКСИВНЫЕ АГЕНТЫ И ЧЕЛОВЕЧЕСТВО:

КООПЕРАЦИЯ ИЛИ КОНФЛИКТ?

Евгений Александрович Гришин

Рефлексивные агенты и человечество: кооперация или конфликт?

Текст предоставлен правообладателем

http://www.litres.ru/pages/biblio_book/?art=55093190

Рефлексивные агенты и человечество: кооперация или конфликт?

*Сборник научных и философских статей / Гришин Е.А.: Издать Книгу;
2020*

Аннотация

Сборник включает два раздела:

1. Раздел «Научные статьи» автора, опубликованные в период 2008–2019 гг. (статья в журнале «Искусственный интеллект и принятие решений» Института системного анализа РАН, 2008 г., и в сетевом журнале «Искусственные общества» (Журналы РАН gas@jes.su), 10 статей, с 2014 г. по 2019 г.).

2. Раздел «Философские притчи» – как гипотетические диалоги виртуального агента и человека на темы их совместного существования в будущем (10 притч), опубликованные на сайте «Философский шторм. Совместное философское творчество» в 2015–2018 г.г. (<http://philosophystorm.org/persona/evgenii-aleksandrovich-grishin>).

В обоих разделах с разной детализацией и обобщениями обсуждены сущностные моменты строительства виртуального агента и его взаимоотношений с человеком:

- Определены понятия «Потребности», «Деятельности» и «Рефлексивная машина» применительно к виртуальному компьютерному агенту. Показано, что они являются основой для моделирования квази-живого и квази-разумного виртуального агента.

- Предложен тест на квазиразумность виртуального агента, как альтернатива тесту А. Тьюринга, акцентирующая внимание не на умении автомата поддерживать разговор с человеком, а на способности договориться и исполнить все условия договора по цене, срокам и качеству товара/услуги.

- Подвергнуты критике три закона робототехники А.Азимова. Предложена альтернативная парадигма: «агент-квазиличность» (вместо «агент-слуга»).

- Предложен подход «взаимное резервирование» как основа партнёрского (договорного, равноправного) сотрудничества рефлексивного агента и человека-оператора при управлении общим жизненным ресурсом.

- Предложен набор категориальных предпочтений как база для реализации функционала сосуществования человека с рефлексивным агентом:

Этика предпочтительнее Рациональности,
Договор предпочтительнее Приказа,
Кооперация предпочтительнее Конфликта.

Автор согласен с выводами книги Ника Бострома «Искусственный интеллект. Этапы. Угрозы. Стратегии» (<https://www.mann-ivanov-ferber.ru/books/iskusstvennyj-intellekt/>)

относительно возможных угроз появления «сильного искусственного интеллекта» для человечества. Однако автор имеет собственное представление относительно стратегии предотвращения угроз, которое он и рассматривает в книге.

Автор рассчитывает на аудиторию читателей широкого возрастного и профессионального спектра, интересующихся философскими, исследовательскими и дизайнерскими подходами к созданию искусственного интеллекта, в том числе – отличающимися от мейнстрима (искусственных нейронных сетей).

Содержание

Об интеллекте: просто о сложном	6
Аннотация	7
Основная часть	9
Заключение	33
Литература	34
Искусственный интеллект и «творчество», «честь», «достоинство»: совместимые понятия или нет?	35
Аннотация	35
Введение	36
1. Сознание и творчество применительно к виртуальному персонажу	38
2. Место валоризации в процессе творчества и место репутации в валоризации	43
Конец ознакомительного фрагмента.	48

Евгений Гришин
Рефлексивные агенты
и человечество:

кооперация или конфликт?

**Сборник научных и
философских статей**

Об интеллекте: просто о сложном

(реплики на выдержки из книги Д. Хокинса и С. Блейсли
«Об интеллекте»)

Аннотация

Как сказано в предисловии, в книге «Об интеллекте» Джефф Хокинс представляет революционную теорию на стыке нейробиологии, психологии и кибернетики и описывающую систему "память-предсказание" как основу человеческого интеллекта.

Автор отмечает, что все предшествующие попытки создания разумных машин провалились из-за фундаментальной ошибки разработчиков, стремившихся воссоздать человеческое поведение, но не учитывавших природу биологического разума.

Джефф Хокинс предполагает, что идеи, сформулированные им в книге «Об интеллекте», лягут в основу создания истинного искусственного интеллекта – не копирующего, а превосходящего человеческий разум.

Кроме этого, книга содержит рассуждения о последствиях и возможностях создания разумных машин, взгляды автора на природу и отличительные особенности человеческого интеллекта.

В одном из отзывов на книгу пишется [4]:

«...Восприятие мира и наши действия основаны на непрерывном формировании прогнозов о том, что произойдет в следующий миг. Человеческий разум, согласно Хокинсу, – это способность прогнозировать, а неокортекс – "орган пред-

видения".

На основе этой теории Хокинс предлагает нам ответы на вечные вопросы: что такое сознание, мышление, творчество, обучение, воображение? Конечно, скажете вы, все это сложно найти в мозге — там только нейроны, а способность мыслить по-человечески нам дает культура: без нее любой Маугли останется зверем среди зверей.

Но пускай себе мозг служит Хокинсу источником вдохновения, главное, чтобы его теория помогла созданию новых компьютеров».

Автор же данной статьи предлагает содержательный анализ некоторых положений книги Джеффа Хокинса «Об интеллекте». Высказаны соображения и аргументы «за» и «против» анализируемых положений. Схематически рассмотрены подходы к исследованию проблемы «псевдоразумности» автомата, основанные на иных, предположительно, более адекватных акцентах.

Основная часть

1. Дж. Хокинс: РЕПЛИКАЦИЯ не требует разума, и (искусственный) РАЗУМ не требует репликации. Вывод: РЕПЛИКАЦИЯ и РАЗУМ – самостоятельные непересекающиеся процессы...

ЕАГ: Всё дело – в возможности естественной совместимости процессов РЕПЛИКАЦИИ и (искусственного) РАЗУМА. И именно это может привести к неуправляемому развитию событий (даже при отсутствии необходимости их совместимости).

2. Дж. Хокинс: Разумным машинам не нужны *эмоции, амбиции, потребности...* Придание их разумным машинам – маловероятно, что столь сложная задача когда-либо окажется выполнимой... Модель ПАМЯТЬ – ПРОГНОЗИРОВАНИЕ есть достаточное основание для создания машины с интеллектом...

ЕАГ: Машинный интеллект, построенный на основе одной модели ПАМЯТЬ – ПРОГНОЗИРОВАНИЕ, будет основан на моделировании лишь только одной – *логической* – функции мозга. Все остальные перечисленные функции (удовлетворение потребностей, эмоции, настроение, общение, саморазвитие и т. д.) – окажутся за бортом.

На наш взгляд, необходимое основание создания Разумных машин – это реализация функций удовлетворения раз-

ных (материальных и нематериальных) ПОТРЕБНОСТЕЙ в разнообразных ДЕЯТЕЛЬНОСТЯХ, для чего необходимы функции РЕФЛЕКСИИ, ПЛАНИРОВАНИЯ и ПРИНЯТИЯ РЕШЕНИЙ о затратах ограниченных РЕСУРСОВ. Конечно, с учётом функций памяти и прогнозирования.

Кроме всего, придание Разумным машинам функций: *эмоции, амбиции, удовлетворение потребностей* – представляется существенным для их построения и развития ещё и потому, что это всё есть необходимые условия реализации феномена ТВОРЧЕСТВА. Иначе говоря, Разумная машина, построенная без учета перечисленных функций, будет принципиально неспособна к творчеству.

3. Дж. Хокинс: Самый разумный робот *не сможет* проявить того понимания и участия, на которое способен человек по одной лишь своей природе...

ЕАГ: Робот именно *должен* проявлять ПОНИМАНИЕ и УЧАСТИЕ по отношению к человеку. Поэтому у Разумной машины должна быть природа, очень близкая к природе человека (не по поведению, а по сущностным проявлениям).

4. Дж. Хокинс: С понятием сознания связаны понятия разума и души...

ЕАГ: Разум связан с понятием *сознание*, а душа связана с понятием *самосознание*. А именно самосознание (душа) и не рассматривается в работе Д.Хокинса как что-либо существенное.

В то же время, разум без души – это «хлопок одной ла-

донью», как сказано (правда, по другому поводу) у писателя Б.Акунина.

Душа есть необходимый ограничитель разума (интеллекта), обслуживающего животные инстинкты, и только все вместе они могут образовать сбалансированное целое, называемое личность.

5. Дж. Хокинс: Вероятно, нам так и не удастся до конца разгадать загадку творчества А.Эйнштейна, но можно с уверенностью утверждать, что отчасти в их основе лежали генетические факторы...

ЕАГ: И только это можно сказать о загадке творчества А.Эйнштейна?

Но тогда очень мощный компьютер (память, быстроедействие), в который «генетически» заложен генератор случайных чисел, сможет выдавать массу «творческих» гипотез!

Выскажем предположение, что творчество субъекта, как целостный процесс, должно включать по крайней мере следующие составляющие:

- Формулировка конкретного вопроса (задачи) в рамках решения общезначимой проблемы, ответом на который может быть та или иная гипотеза,
- Генерация гипотез (и всё *бессознательное*, что с этим связано – интуиция, вдохновение, инсайт, «эврика»), поиск гипотез во внешней среде (случайные блуждания), *сознательный* выбор гипотезы как решения задачи, а также практическая проверка выбранной гипотезы самим субъектом,

- Общение с другими субъектами, в результате которого гипотеза (решение задачи) может быть принята или опровергнута ими. И если решение задачи ими как экспертами проверено на практике и признано как наиболее эффективное и имеющее повышенный спрос на рынке, то только тогда можно сказать, что вышеозначенный процесс есть *творчество*.

Т.о., попытаемся сформулировать *необходимые и достаточные условия* представления некоего процесса как процесса творчества субъекта.

Первое необходимое условие – осознание субъектом *общезначимой проблемы* и в её рамках – чёткое формулирование вопроса (задачи), ответом на который может быть та или иная найденная гипотеза.

Второе необходимое условие – наличие внешнего и/или внутреннего генератора гипотез; таковым генератором может быть как сам субъект (индивидуальное бессознательное), так и другие (коллективное бессознательное).

Третье необходимое условие – взаимопонимание творящего субъекта с другими субъектами в игре, борьбе или решении общих проблем. В свою очередь, взаимопонимание возможно при наличии у субъектов близкого мировосприятия. А близкое мировосприятие может основываться лишь на общности систем ценностей субъектов, общности ресурсной базы их среды существования и на ресурсном обмене (рынке).

Творчество также предполагает знание субъектом существующих способов решения задачи, неудовлетворённость ими и попытку решить задачу лучше.

И это лучшее решение ценно потому, что может принести ему как те или иные внешние выгоды (деньги, признание, власть), так и внутренние (самоуважение).

Отсюда выводится *достаточное условие* для признания некоего процесса и его результата творчеством: решение задачи, предложенное субъектом, должно быть признано на данный момент на рынке (ресурсов, идей) лучшим в соответствующем классе решений, и это должно быть подтверждено повышенным на него спросом (цитированием).

6. Дж. Хокинс: Модель «Память – Прогнозирование» в своей основе имеет феномен запоминания мозгом СЕРИЙ событий. В последующем мозг, на основе появления на входе отдельных элементов серии, ПРОГНОЗИРУЕТ (ПРЕДВОС-ХИЩАЕТ) появление на входе всей серии событий...

ЕАГ: Представляется, что это важный, но лишь один шаг в постижении феномена разума.

Не менее важен и следующий шаг: ПЛАНИРОВАНИЕ ПОВЕДЕНИЯ, которому в работе [1] уделено недостаточно внимания.

На наш взгляд, этот следующий шаг уже терминологически предполагает раздвоение единого процесса на два самостоятельных, алгоритмически равнозначных, процесса. Они могут быть названы так: ОСОЗНАНИЕ (включающее плани-

рование) и собственно ПОВЕДЕНИЕ по плану.

Мы имеем в виду, что это два таких алгоритмических процесса, которые обязаны пользоваться ЕДИНЫМ ПОЛЕМ ПАМЯТИ так, что один процесс использует результаты другого.

Принципиально важно то, что оба процесса должны осуществляться не ПОСЛЕДОВАТЕЛЬНО, а ПАРАЛЛЕЛЬНО.

Это возможно лишь при условии, что их параллельное выполнение относительно общего поля памяти будет характеризоваться ФАЗОВЫМ сдвигом.

Понимать это надо так. Один процесс не ждёт конца исполнения другого процесса, а исполняется на текущем этапе, пользуясь результатами другого процесса, полученными на ПРЕДЫДУЩЕМ этапе. Второй процесс, выполняясь одновременно с первым, использует результаты первого, полученные также на ПРЕДЫДУЩЕМ этапе.

Тем самым через фазовый сдвиг вводится необходимый параметр ВРЕМЕНИ как критерий последовательности событий.

Приведённая совокупность двух процессов, выполняющих функции ОСОЗНАНИЯ (т. е. прогнозирования и планирования) и ПОВЕДЕНИЯ (как исполнения плана), может быть охарактеризована психологическим понятием РЕФЛЕКСИРОВАНИЕ и более общим кибернетическим понятием ОБРАТНАЯ СВЯЗЬ.

Парадоксально, что отдельно рассмотренный эвристиче-

ски мощный психологический принцип рефлексии, не обязательно предполагающий взаимодействия двух параллельных процессов, сам по себе оказывается недостаточно эффективен для анализа сущности РАЗУМНОСТИ.

Каждый из двух вышеприведённых алгоритмических процессов является примером универсальной машины А.Тьюринга. Обобщённо говоря, предлагается единая модель, представляющая собой две универсальных машины Тьюринга, работающих на общем поле памяти под управлением супервизора.

Эвристическая эффективность совмещённой модели может оказаться значительно выше для понимания и машинной реализации феномена РАЗУМНОСТИ, чем сумма отдельно взятых моделей.

7. Дж. Хокинс: Одно из важных свойств мозга есть *репрезентативная инвариантность*, т. е. хранение информации о целом в виде инвариантных соотношений параметров объекта.

ЕАГ: Идея инвариантности, как чрезвычайно сильный методологический приём, нами взята за основу построения алгоритма существования «псевдоразумного» компьютерного персонажа. К таковым инвариантам относятся следующие модели, в совокупности представляющие собой обобщенную инвариантную модель ДЕЯТЕЛЬНОСТИ персонажа:

- Модель «ПОТРЕБНОСТИ-ДЕЯТЕЛЬНОСТИ-РЕСУРСЫ» как инвариантная *нижняя* модель жизненной *активно-*

сти персонажа.

- Модель «ПРОЕКТ: СЦЕНАРИЙ-РОЛИ-ИСПОЛНИТЕЛИ» как инвариантная *первая промежуточная* модель жизненной *активности* персонажа.

- Модель «АНАЛИЗ ПРИЧИН-ПРОБЛЕМА-ЗАДАЧА-ПЛАН» как инвариантная *вторая промежуточная* модель жизненной *активности* персонажа.

- Модель «ОБЩЕНИЕ: ПРОСЬБА-ПРЕДЛОЖЕНИЕ» как инвариантная *третья промежуточная* модель жизненной *активности* персонажа.

- Модель «САМОРАЗВИТИЕ: ДОЛГ ПЕРЕД СОБОЙ» – осознанная смена персонажем его приоритетов на множестве собственных потребностей в пользу *нематериальных* потребностей как инвариантная *высшая* модель жизненной *активности* персонажа.

- Модель «УНИВЕРСАЛЬНЫЙ ПРЕОБРАЗОВАТЕЛЬ: ИСХОДНЫЙ МАТЕРИАЛ-СРЕДСТВО-ЦЕЛЬ-РЕСУРСЫ» как инвариантная модель *технологии преобразующей деятельности* персонажа.

- Модель «РЕФЛЕКСИВНАЯ МАШИНА: ЗАПОМИНАНИЕ, ПЛАНИРОВАНИЕ, ИСПОЛНЕНИЕ и САМОКОНТРОЛЬ» как инвариантная модель *технологии обратной связи* в деятельности персонажа, связи между осознанием и поведением.

- Модель «САМООБУЧЕНИЕ: ВЫРАБОТКА СТЕРЕОТИПОВ ПОВЕДЕНИЯ» как инвариантная модель *техноло-*

гии самообучения. Эмпирическое вычисление класса наиболее эффективных вариантов поведения из потока вариантов за небольшое число шагов (обучение) и большое число шагов использования этого опыта в «боевых» условиях, с последующим многократным переобучением.

- Инвариантный критерий выбора момента ПРИНЯТИЯ РЕШЕНИЯ при последовательной генерации вариантов – совпадение величины предполагаемого дополнительного эффекта от искомого варианта поведения при продолжении поиска, с совокупными затратами на поиск искомого варианта (фактически осуществлёнными и предполагаемыми).

Изложенный критерий выбора момента принятия решений применим для каждой из вышеперечисленных моделей, т. к. в них предполагается последовательная генерация вариантов.

Это может быть решение о снижении планки желаемого выигрыша и продолжении поиска, если появившиеся варианты не сохраняются в памяти после появления. Или это может быть решение о прекращении поиска и выборе лучшего из уже появившихся вариантов, если они сохраняются в памяти после появления.

Единая инвариантная модель ДЕЯТЕЛЬНОСТИ персонажа представляет собой набор инвариантных моделей, образующий целое. Это целое обладает структурой, являющейся полным графом (каждый элемент связан с каждым, рис. 1).

Полный граф подчёркивает то обстоятельство, что рассмотрение любого из элементов структуры в отдельности, без связей друг с другом, даёт неадекватную (непроявленную) картину понимания существа целого.

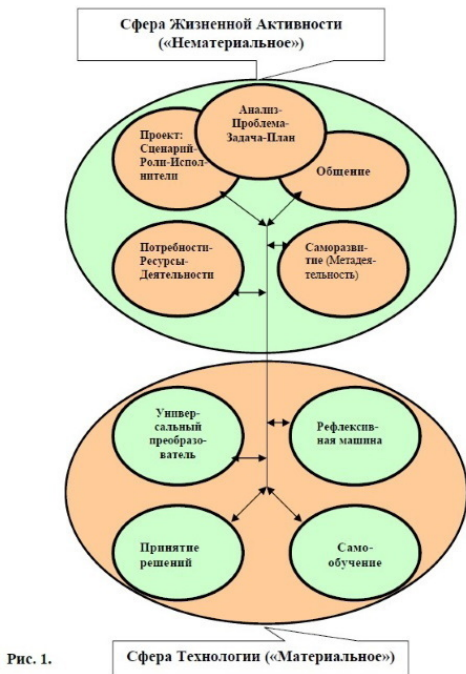


Рис. 1.

Рис. 1.

В то же время представляется допустимым и целесообраз-

разным для понимания подхода сгруппировать модели так, чтобы выделить две группы, дуальные по отношению друг к другу с точки зрения их ролей:

- Группа *технологических* моделей существования персонажа (сфера «материального») и
- Группа моделей *жизненной активности* персонажа (сфера «нематериального»).

Этот принцип дуальности элементов целого как структурный инвариант сохраняется и при переходе на нижележащий уровень рассмотрения моделей (можно сказать, реализуется «фрактальность» как свойство сохранения структурной инвариантности при изменении масштаба рассмотрения объекта). Так, модель Рефлексивной машины представляет собой принципиально двухсферную структуру, сферу осознания и сферу поведения

В целом, с учётом масштабной инвариантности в модели последовательно реализуется методологический принцип дополнительности Н.Бора.

Рассмотрим иные примеры инвариантностей, уже в сфере языкового общения персонажей.

«Не спрашивай, о чём говорят – говорят о деньгах» – Эмманюэль Тодд (предположительно).

Сделаем попытку повторить подобный метафорический подход применительно к другим сторонам жизненного проявления и взаимодействия персонажей. Акцент сделаем на

поиске инвариантностей в интерпретации их жизненных проявлений.

Не спрашивай, о чём (почему):

Говорят – о деньгах (*просят* деньги или услуги в счёт будущего дохода, или *предлагают* деньги или услуги за будущие услуги или деньги).

Плачут – жалеют себя (и *предлагают* другим разделить с ними горе).

Смеются – гордятся собой (и *предлагают* другим разделить с ними это удовольствие).

- **Поют** – о любви (желают её *подарить*, или желают её *получить*, или и то, и другое).

- **Сердятся** (оскорбляют) – «мстят» другому за собственное неуважение себя

- **Действуют** – стремятся к желаемой *цели*, ища эффективные *средства* при минимизации затрат *ресурсов*.

Борются (дерутся, воюют, занимаются политикой, вообще действуют) – добиваются победы над другим с целью *заполучения* лидерства, ресурсов, услуг. Или с целью *пресечения* зла, беззакония, несправедливости, угрозы своей безопасности или безопасности общества, основываясь на своём понимании нравственности.

- **Дают** кому-либо *ресурсы*, услуги – исполняют свой долг в соответствии с добровольным *договором* или свой *внутренний долг*, или *покупают* безопасность в связи с *ультиматив-*

мом другого, или ультимативно *навязывают* в долг ресурсы, услуги другому.

- **Дарят** кому-либо ресурсы, услуги – *бескорыстно исполняют долг* перед собой, или *корыстно провоцируют долг* другого перед собой.

- **Берут** (получают в дар) у кого-либо ресурсы, услуги – воспринимают дающего своим должником (по добровольному договору или в связи с собственным ультиматумом берущего дающему), или готовы себя считать должником дающего (по добровольному договору или в связи с ультиматумом дающего берущему).

- **Играют** – сознательно ввергают себя в *искусственную реальность*, в которой быстрее можно *проявить* себя, имитируя реальную деятельность или борьбу в логическом, этическом или эстетическом отношении.

- **Читают** (потребляют предметы любого искусства) – сознательно ввергают себя в *искусственную реальность*, в которой быстрее можно *понять* другого и соотнести себя с ним.

- **Пьют**, потребляют наркотики, медитируют, болеют (в спорте, игре, споре) – сознательно ввергают себя в *искусственную реальность*, в которой можно быстрее (и сильнее) *получить удовольствие, бездействуя и испытывая страсти*, или получить кратковременный *допинг* (катализатор) для реальной деятельности.

- **Обманывают (играют)** – производят ложные движе-

ния (произносят лукавые тексты), сознательно ввергая другого в искусственную реальность и *провоцируя* его на действия, выгодные «обманщику» и невыгодные (или выгодные) его жертве/партнёру. Например:

- Просят деньги (услуги) у партнёра под обещания неочевидного будущего его дохода, или предлагают деньги (услуги) партнёру за его будущие несоразмерные услуги (платежи).

- Льстят – словесным наркотиком усыпляют разум другого (обманно утверждая о достоинствах, богатстве, авторитете, красоте, справедливости... другого) в расчёте на получение своего корыстного интереса.

- Играют на сцене (или в жизни) – произносят тексты и «прикидываются», желая, чтобы зрители поверили им и забыли, что они – просто исполнители.

В свою очередь, зрители – сами желают быть «обманутыми», но иногда «не верят»...

Поиск инвариантов в вышеперечисленном списке проявлений персонажа приводит к следующему. Вычленены смысловые пары как категории, к которым можно свести остальные проявления. Это следующие категории:

Живут или **играют** (*настоящая* реальность или *искусственная* (игровая, виртуальная)).

Берут (получают) или **дают** (дарят), **просят** или **предлагают**.

Попытаемся понять, как соотносятся и различаются для персонажа *настоящая* реальность и *искусственная* (игровая, виртуальная). Представляется, что основой этому может быть различие понятий *быть* и *казаться*, а также понятий *жить* и *играть*.

И в той, и в другой реальности персонаж осуществляет ту или иную деятельность.

И в той, и в другой реальности одна и та же деятельность удовлетворяет одну и ту же группу потребностей персонажа (может быть, с разной интенсивностью).

Но:

- Любая деятельность в игровой реальности, в отличие от проявления подобной же деятельности в настоящей реальности, удовлетворяет ещё и дополнительные потребности – потребность в *азарте* (*риске*), а также потребность в *удовольствии от удачного обмана*! Обман – предумышленное действие персонажа по перевоплощению в исполнителя некоей роли, несвойственной персонажу, по мнению зрителей (партнёров). При этом персонаж испытывает удовольствие от ощущения готовности зрителей (знающих о перевоплощении) поверить в истинность этого действия, или от ощущения успешности обмана, если зрители (партнёры) не догадались о перевоплощении. Иначе говоря, азартная и/или «обманная» деятельность *всегда есть игровая реальность*!

Здесь реализуются понятия *играть* (*vs жить*), а также *казаться/прикидываться* (*vs быть*).

- Существует подмножество деятельности персонажа, которые могут присутствовать в обеих реальностях, и которые только в настоящей реальности (но не игровой) могут удовлетворить специфические потребности – *физиологические потребности* ТЕЛА персонажа! Т. е., такая деятельность *всегда есть настоящая реальность*! Здесь реализуется понятие *жить* и *быть*.

Для наблюдателя внешними признаками наличия *искусственной* реальности (не *настоящей*), в которой могут находиться персонажи, могут быть такие признаки, как:

- обман (провокация, финт, подставка, дезинформация) во взаимоотношениях персонажей,
- азарт (повышенная рисковость) в проявлениях персонажей,
- повышенная эйфория (аффектация) в проявлениях персонажей.

При всей непререкаемости тезиса «жизнь есть игра» у автора есть интуитивное ощущение, что при каких-то обстоятельствах жизнь всё-таки остаётся просто жизнью, и не является игрой. Вопрос: при каких же условиях «жизнь не является игрой»?

Такая постановка вопроса представляется важной в рамках проблематики создания «псевдоразумного» компьютерного персонажа. Сделаем попытку на него ответить.

В дальнейшем термин «персонаж» будем понимать как «псевдоразумный» компьютерный персонаж, так и «человек».

Термин «жизнь» будем понимать как «существование компьютерного персонажа в виртуальной среде», так и «жизнь человека».

Термин «искусственная реальность» будем использовать в контексте «искусственная реальность, создаваемая персонажем, как вторичная по отношению к первичной реальности, интуитивно понимаемой как жизнь персонажа».

Приведём свои определения.

С одной стороны, Игра – это *сознательно созданный* персонажем-инициатором проект «искусственной реальности», *имитирующей* некую реальную деятельность, с участием себя как режиссёра и других персонажей как исполнителей.

С другой стороны, Игра – это *процесс неосознанного исполнения* персонажем – инициатором и персонажами-исполнителями ролей в проекте «искусственной реальности», т. е. собственно *Игра*.

Получается, что персонаж *играет* только тогда, когда он уже создал проект игры и приступил к его исполнению, или согласился исполнить роль в чужой игре!

Во всех остальных случаях персонаж не играет, а *живёт*, и когда только ещё создаёт тот или иной проект, и когда исполняет неигровой проект, и когда тайно *играют* с ним.

Таким образом, тезис «жизнь есть игра» правильное пред-
ставить в виде: **«жизнь есть изобретение персонажем
игр с другими, в то время как другие, возможно, уже
давно играют с ним»!**

Проект игры включает элементы:

- субъект (режиссёр, инициатор),
- сценарий (список ролей и планов их исполнения),
- объекты (персонажи-исполнители ролей).

В некоторых случаях субъект и объект совпадают (игра с самим собой).

Персонаж *живёт реальной жизнью*, когда разрабатывает проект игры.

Персонаж *играет*, или находится в «искусственной реальности», когда исполняет свой проект игры в качестве режиссёра или сознательно участвует в чужом проекте игры в качестве одного из исполнителей.

Цель *имитации* – всегда *обман*:

- сознательный *самообман* (игра с самим собой или сознательное вхождение в роль исполнителя не в своей игре),
- сознательный *обман* другого персонажа (вовлечение в свою игру другого персонажа, не осознающего этого).

Цель игрового *самообмана* для персонажа- инициатора:

- испытать более сильные страсти – азарт, эйфорию, оргазм (через участие в состязаниях, зрелищах, приёме алкого-

ля, наркотиков, поиск любовных приключений),

- быстрее обучаться профессиональным навыкам (посредством моделирования, симулирования, тренировок),
- легче и быстрее познавать суть других персонажей и сравнивать их с собой, чем это возможно в жизни (посредством чтения, потребления предметов любого искусства).

Цель игрового *обмана* для персонажа-инициатора:

- вовлечь в свою игру другого персонажа,
- не дать ему возможности осознать, что он невольно исполняет роль в игре «кукловода».
- использовать другого персонажа в требуемой роли и посредством этого получить свою выгоду (материальную или нематериальную).

Этот другой персонаж может даже сознательно участвовать в объявленной игре. Но при этом в разные моменты он может и не догадываться, что он – исполнитель чужой воли (уже в иной, необъявленной, игре), когда поддаётся финту, подставке, гамбиту, дезинформации, очарованию.

Сначала поговорим об имитации как сознательном обмане другого персонажа.

Игра персонажа-инициатора только тогда может считаться игрой-обманом, когда он её задумал, начал, но *не объявил* её как игру для других, вовлечённых в неё, персонажей.

С того же момента, как персонаж-инициатор объявляет о начале и правилах игры, данная игра как таковая (как обман) перестаёт существовать, и продолжается уже как неигровой

(необманный) проект.

Но у персонажа-инициатора и у каждого из других участников в рамках объявленной игры существуют собственные заготовленные, необъявленные, стратегии и планы игр-обманов, интриг, о которых ещё должны догадываться другие участники.

Таким образом, игра-обман есть не более чем решение, план и исполнение персонажа-инициатора, вовлекающего в свою игру других персонажей, и существует лишь постольку, поскольку этот план им *не объявлен* как игра, или *не распознан* персонажем-жертвой как игра!

Объявленная персонажем-инициатором до своего начала *игра* при согласии участников со всеми правилами прекращает быть игрой и становится *реальной жизнью* по правилам (в которой участниками могут разыгрываться уже другие игры-обманы)!

Игра, *признанная* и *объявленная* персонажем-инициатором в середине, в конце или по окончании как игра-обман, прекращает быть его *игрой*, но не обязательно становится его *реальной жизнью*.

Игра-обман может быть *распознана* персонажем-жертвой (сразу, в середине игры, в конце или по окончании игры). Если игра-обман *распознана* персонажем-жертвой, то персонаж-инициатор, не зная об этом и продолжая играть, может сам оказаться жертвой другой игры-обмана, затеянной уже персонажем-«жертвой» первой игры. И так далее.

И только будучи *объявленными* или *распознанными* (и *обнародованными*), *игры* прекращают быть таковыми и могут стать правилами для *реальной жизни*.

Таким образом, для персонажа-инициатора игра есть процесс исполнения секретного игрового проекта, ранее им разработанного. Проект заключается в тайном использовании других персонажей как исполнителей ролей.

Исполнение этого проекта может считаться игрой только до тех пор, пока проект *не объявлен* персонажем-инициатором как игра, или пока он *не распознан* персонажем-исполнителем как игра. После *объявления* или *распознавания* (и *обнародования*) данная игра становится *прошлой игрой*. У персонажа-инициатора и у персонажа-исполнителя продолжается *реальная жизнь*, и каждый из них волен начать в ней новую игру с учётом знаний, полученных каждым в предыдущей игре.

Теперь поговорим об *игре как самообмане* персонажа. Попробуем выяснить точки перехода от настоящей реальности к самообману и наоборот.

Пока персонаж разрабатывает проект игры самого с собой или обсуждает исполнение роли в чужой игре, он не играет, он *живёт*.

Когда же персонаж начинает реализовывать игровой план,

он *играет*. Эта точка перехода, как правило, психологически не заметна персонажу. Далее персонаж настолько может оказаться увлечённым игрой, что перестанет отдавать себе отчёт, что он играет. Выход из игры – это *факт рефлексии*. персонаж в какой-то момент приходит в себя и признаётся себе, или «будильник» или другой персонаж открывает ему, что он находится в игре (с самим собой или в чужой игре). Что он увлёкся этим самообманом до такой степени, что начал воспринимать его как *настоящую реальность*. С этого момента у персонажа вновь начинается реальная жизнь. Правда, до того момента, пока он вновь не решит сыграть (сам с собой, с другим или исполнить роль в чужой игре).

Таким образом, для персонажа игра с самим собой начинается после разработки своего игрового проекта или после обсуждения участия в чужом игровом проекте. В течение игры сам факт игры персонажем не осознаётся. Игра прерывается моментом рефлексии, когда персонаж осознаёт, что он играет, и тем самым возвращается к реальной жизни. Далее персонаж решает, продолжить ли ему эту реальную жизнь, или возвратиться к игре.

Вывод: антитеза состояний «жизнь – игра» некоего персонажа представляется как целое, которое может быть рассмотрено с двух точек зрения: *синхронной* и *идиахронной*.

С точки зрения синхронной персонаж *одновременно и живёт, и играет*.

Персонаж *живёт*, поскольку сознательно (и тайно) разрабатывает проект будущей игры с другими персонажами.

Персонаж одновременно *играет* (и играют с ним), пока он неосознанно исполняет свой (или чужой) предыдущий игровой план.

В этом – пример реализации принципа дополнительности Н.Бора.

С точки зрения диахронной персонаж *последовательно* находится в этих состояниях. Сначала он разрабатывает некий игровой проект (план), т. е. *живёт*. Затем он его (именно этот план) исполняет, т. е. *играет*.

В этом – реализация принципа метасистемного перехода В.Турчина.

Но переход персонажа из одного состояния в другое в диахронном смысле, а также синхронное соотнесение двух противоположных состояний персонажа описывается моделью Рефлексивной машины (схематически представлено на рис. 2.)

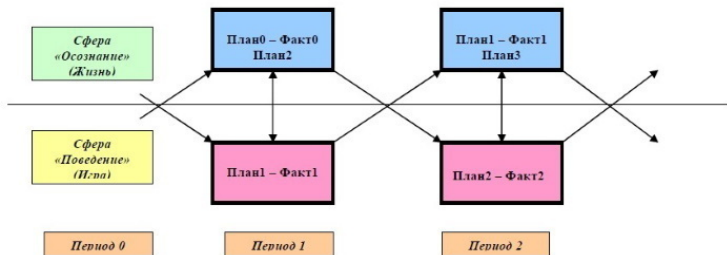


Рис. 2. Синхронно-диахронное представление анти-тезы «жизнь – игра» и сфер «Осознание» и «Поведение» персонажа.

Отсюда следует, что модель Рефлексивной машины есть инвариант для представления как двух сфер, «Осознание» и «Поведение», так и двух состояний, «Жизнь» и «Игра».

Заключение

Предпринята попытка критического анализа отдельных положений книги Д. Хокинса и С. Блейкли «Об интеллекте».

Приведены аргументы в пользу нецелесообразности отказа от попыток придания машине таких свойственных для человека проявлений, как эмоции, амбиции, потребности.

Описана обобщённая инвариантная модель ДЕЯТЕЛЬНОСТИ компьютерного персонажа, включающая вышеперечисленные свойства. Компьютерная реализация модели показала, что в ней присутствуют признаки «псевдоразумности» во взаимодействии с человеком.

Рассмотрены подходы к содержательной интерпретации разницы понятий «жизнь» и «игра» персонажа.

Предложено рассматривать антитезу состояний «жизнь — игра» персонажа как целое, которое может быть исследовано с двух точек зрения: *синхронной* и *диахронной*.

Тогда синхронный срез модели окажется реализацией принципа дополнительности Н.Бора, а диахронный срез — реализацией принципа метасистемного перехода В.Турчина.

В свою очередь, это позволяет рассмотреть соотношение понятий «жизнь» и «игра» персонажа аналогично соотношению понятий «осознание» и «поведение».

Литература

1. Джефф Хокинс и Сандра Блейкли. «Об интеллекте». Вильямс. 2007.
2. Дмитрий Самин. Основы мироздания. Принцип дополненности. <http://bibliotekar.ru/100otj<r/37Jitm>
3. Турчин В.Ф. Феномен науки. Кибернетический подход к эволюции. Изд. 2-е М.:ЭТС. 2000. 368 с. <http://refaLru/turchin/phenomenon/index.htm>
4. <http://www.oхран-komplekt.ru/oхран-komplekt-izobretateli/news> 2007-11-09zl6z05zllz659Mmi

Искусственный интеллект и «творчество», «честь», «достоинство»: совместимые понятия или нет?

Аннотация

Предпринята попытка рассмотреть поведенческую сторону творческого процесса как единственно алгоритмизируемую часть применительно к моделированию его при конструировании интеллектуального виртуального персонажа. Анализ важен сточки зрения сравнения возможностей человека и интеллектуального персонажа в их когнитивном и рационально-этическом поведении и взаимодействии в виртуальном (и реальном) мире.

Введение

В последнее время сформировалось историческое и методологическое подразделение естественнонаучных исследований на этапы: классический, неклассический и постнеклассический [1, 2].

Охарактеризуем их в самой общей форме для целей дальнейшего изложения.

Классический этап базируется на механике Ньютона, рассматривающей материальные объекты и силы взаимодействия. Гуманитарная составляющая отсутствует.

Неклассический этап характеризуется междисциплинарностью, допускающей учёт двойственности в моделях мира, а также учёт целесообразной деятельности человека как субъекта, его воздействия на объект посредством орудий труда/исследования для достижения цели.

Постнеклассический этап характеризуется возрастающим интересом к нелинейным, саморазвивающимся, рефлексивным процессам, а также интересом к исследованию феноменологической (духовной) составляющей деятельности человека – механизмов рефлексии ценностей и смыслов его деятельности.

В исследовательском направлении «искусственный интеллект» к настоящему времени главенствует неклассический подход, подтверждением чего является основной ак-

цент на совершенствовании методов распознавания, обучения, представления, накопления и систематизации знаний, логики принятия решений, автоматической генерации программ.

В то же время идеи постнеклассического подхода только начинают осознаться как существенно важные для развития данной области знания.

Это проявляется в образовании таких направлений исследований, как «интеллектуальные роботы», «многоагентные системы», «виртуальные предприятия» с обращением к понятиям активности, мобильности, автономности, рефлексивности, планирования, коммуникабельности виртуальных сущностей [3, 4].

Сделаем и далее обоснуем предположение: адекватное конструирование и исследование сообщества виртуальных интеллектуальных компьютерных персонажей в рамках постнеклассического подхода требует идей представления персонажа с использованием понятий: «квазиживой» и «квазиразумный» [4].

1. Сознание и творчество применительно к виртуальному персонажу

Приведём некоторые общие соображения о феномене сознания.

Вспомним об известном приёме – принципе дополнительности Н.Бора, являющимся эвристически сильным методологическим способом выявления эффекта целостности в постижении любого предмета анализа [5].

Если предметом анализа становится феномен психики, то указанный методологический приём приводит к целостной, неразрывной паре: «сознательное – бессознательное».

Сделаем свои самые общие предположения о модели взаимодействия этой пары в индивиде, для использования её в последующем конструировании виртуального персонажа.

Прежде всего, целостность модели основывается на представлении её в виде *параллельно исполняемых* «сознательно-бессознательных» рекурсивных функций (механизмов), области определения и области значений которых есть участки общего поля памяти, *периодически переключаемые* между функциями посредством механизма рефлексии, соблюдающей некий эталон целостности [8].

Блок «сознание» в целостной паре «сознательное – бессознательное»

знательное» есть:

- *Генератор текстов планов*, реагирующий на воспринятые органами чувств и отражённые в памяти индивида образы предметов мира как на объект для их символического описания. Такое описание (объективирование) генерируется в результате анализа плана и факта прошлого взаимодействия индивида с объектом, измерения разницы план-факт как степени неудовлетворения потребностей индивида, и творческого синтеза решения о плане будущего действия индивида с объектом ради удовлетворения данной потребности.

- *Генератор (модификатор) алгоритмов поведения* на основе обучения (эволюционный принцип мутаций и естественного отбора для повышения эффективности алгоритма).

- *Генератор текстов запросов*, ответы на которые, как информация о вариантах планов и их вероятностях, необходимы для синтеза решений о плане. Ответы на запросы (информация) приходят от генератора ответов, являющегося составной частью механизма «бессознательного». На основе полученной информации генератор планов принимает решение о будущем плане и алгоритме, которые запоминаются в актуальной памяти индивида и передаются на исполнение и контроль.

Блок «бессознательное» есть:

- *Интерпретатор*, пошагово исполняющий текущий алгоритм поведения в подсознании. Этот алгоритм поведения

реагирует действием своих эффекторов в среде на воспринятые органами чувств (рецепторами) и отражённые в памяти индивида образы предметов мира, как на появившийся *стимул*. Алгоритм поведения является или генетически заложенным «инстинктом», или ранее скомпонован генератором-модификатором алгоритмов поведения в механизме «сознательного», запомнен в актуальной памяти и исполняется интерпретатором.

- *Алгоритм – контролёр* фактического исполнения *плана поведения*, выработанного ранее механизмом сознания, и запомненного в актуальной памяти.

- *Генератор ответов на запросы*, получающий запросы от механизма сознания, и направляющий ответы в механизм сознания. Генератор ответов может находиться вне тела индивида (это – партнёры и внешние источники данных), и внутри индивида (в его механизме «бессознательного», работающего с долговременной памятью). Он даёт ответ механизму сознания как вариант решения, для осмысления проблемы и разработки плана будущего поведения.

Именно генератор ответов, сканирующий долговременную память индивида и выдающий за непредсказуемое время непредсказуемый ответ, и является механизмом собственно «бессознательного». Ведь он имеет дело с образами предметов мира из памяти, которые могли никогда не быть объектом осознания, и были запомнены, минуя этап объективации. А теперь эти образы становятся информаци-

ей для принятия решений. В то же время объективированию (как осознанию, рефлексии) периодически подвергаются сами планы и алгоритмы поведения индивида.

Они запоминаются в памяти, в следующем периоде рефлексии неосознанно исполняются и контролируются в подсознании, чтобы стать предметом объективации в очередном периоде рефлексии.

Таким образом, можно допустить, что механизм сознания в принципе своими объектами имеет следующие. Предмет анализа – именно поведенческая (объективируемая) сторона прошлого бытия индивида. Это – соотношение план-факт как проблема, а также степень эффективности навыков, умений, алгоритмов. Предмет синтеза – планы и алгоритмы будущего поведения.

Но тогда следует допустить и другое, а именно:

– механизмы генерации идей, генерации *нового знания*, *know how*, питающие сознание, и позволяющие индивиду «творить», суть бессознательны сами по себе по определению, и посему не могут быть предметом сознательного анализа и синтеза индивида!

Однако ничто не мешает индивиду постфактум анализировать поведенческую сторону своего существования в тот период, когда он решал творческие задачи. И именно через совершенствование поведения, через сравнение план-факт, можно рационализировать и повышать эффективность процесса творчества [6].

Вывод: Рефлексивную модель взаимодействия пары «сознательное – бессознательное» в индивиде, а также модель рационализации поведения для повышения эффективности процесса творчества можно использовать при конструировании виртуального персонажа.

2. Место валоризации в процессе творчества и место репутации в валоризации

В финансовом деле под понятием «валоризация» понимают «процесс по увеличению основного капитала»*.

Позволим себе в дальнейшем использовать термин «валоризация», понимая под ним следующее словопотребление: *придание новой большей ценности объединённому в целое множеству малоценных объектов (идей), подтверждённое продажей этого целого на рынке.*

Сделаем предположение, что валоризация в изложенном смысле есть существенный результирующий аспект разумной деятельности вообще (не только в искусстве и финансах).

Однако, закономерен вопрос: является ли валоризация результатом только *творческой* деятельности индивида в ранее высказанном смысле, как осознаваемый продукт генерации *нового знания* бессознательным? Или валоризация может быть также и результатом осознания индивидом поведенческой стороны своего существования?

Вернёмся к идее создания и исследования сообщества компьютерных персонажей.

Определим два вида поведения персонажа:

- Поведение персонажа в физическом мире (виртуальном или реальном).
- Поведение персонажа в обществе (социальном мире) себе подобных и людей.

Рассмотрение разумного поведения персонажа в реальном физическом мире необходимо требует моделирования как сознательных процессов, так и бессознательных.

Суть того, что не может быть алгоритмизируемо в работе бессознательного, есть свойство бессознательного быть активным внутренним источником, генератором информации о новых классах, категориях целей и средств достижения цели, т. е. быть инструментом творческого решения задачи.

Примечание*: Валоризация (англ. Valorization) – Рост цен на товары, курсов ценных бумаг вследствие определенных действий государства. <http://dic.academic.ru/dic.nsf/business/1952>

Но есть и другая сторона творческого процесса – поведенческий поиск внешнего источника в окружающей среде, генерирующего информацию о вариантах средств, и слежение за уже найденным источником.

Именно этот процесс вполне алгоритмизируем, так как он суть поведенческий, а значит, осознаваемый. И на него остаётся уповать, рассуждая о возможности моделирования автономного поведения персонажа в реальном физическом мире.

Однако тогда это будет моделирование не более чем «ква-

приразумного», т. е. нетворческого, поведения.

Но будет ли этого достаточно для создания автономного разумного автомата? Ведь он должен будет выживать и творчески развиваться в произвольном виртуальном или физическом мире. А без учёта «бессознательного» в алгоритме по-прежнему останутся нерешёнными все проблемы поиска категорий, абстрагирования, классифицирования, вообще обнаружения и формулирования нового знания.

Поставим следующий вопрос: Не поможет ли найти новые аспекты моделирования разумного поведения персонажей анализ взаимоотношений персонажей в социальном мире, абстрагируясь от физического мира? Это должны быть такие аспекты, которые смогли бы, наряду с генерацией нового знания, в определённой степени влиять на результирующую составляющую творчества, т. е., на валоризацию как повышение цены товара на рынке.

Ответ: Да, есть такие аспекты, если наряду с качеством товара,

предлагаемого на рынке, рассматривать как важнейшее условие сделки «Честь» (*Репутацию*) самого персонажа-торговца!

Поясним данное соображение. Предположим, что сделка, т. е. договор об обмене, купле-продаже товара-услуги, является достаточно частым видом социального взаимодействия персонажей.

Сформулируем ряд важных, по нашему мнению, вопро-

сов для выяснения специфических факторов, влияющих на решение о сделке и на её результат.

1. Не является ли результат договорённости сторон о цене сделки (как валоризация) в существенной степени следствием *исполнительности* по обязательствам продавца и покупателя, оцениваемой *доверием* к ним других персонажей – покупателей и продавцов?

2. Не является ли пара *исполнительность-доверие* главенствующим обстоятельством в общественных взаимоотношениях мира персонажей, если соотносить его с важностью физических качеств товара/услуги для их деятельности?

3. Как взаимодействует предполагаемый рост общественного спроса на обязательность партнёров (доверие) с общественным предложением этой обязательности? Не колебательно ли, то есть, не *рефлексивно* ли?

4. Что же может являться критерием, лежащим в основе оценки обязательности партнёра и доверия к нему другого? Не *риск* ли, оцениваемый в виде известной персонажу по опыту частоты положительных случаев прошлого исполнения обязательств партнёра, который преобразуется им в *вероятность* будущего исполнения обязательств партнёра (*вероятность* положительного исхода сделки)?

5. Что критически важное лежит в основе решения персонажа рискнуть и довериться этому или другому варианту, предложенному партнёром или придуманному самим? Не затраты ли жизненно важных и ограниченных *ресурсов* пер-

сонажа на реализацию решения, а также возможный доход, оцениваемые совокупно в некоем эквиваленте, например, в деньгах?

6. Чем принципиально важным руководствуется персонаж в своём решении: как поделить свои ограниченные *ресурсы*

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.