

В. Л. Васюков

Проблема сознания с точки зрения логического функционализма*

Введение

Функционализм, как возникшее в 60–70-х гг. прошлого столетия направление в философии сознания, все еще не исчерпал своих возможностей, существуя и развиваясь и поныне. Основную его идею кратко можно охарактеризовать как утверждение о том, что «виды ментальных состояний следует считать не видами физического и вообще не какими-либо свойствами, материальными или идеальными, а функциональными состояниями» [Юлина 2004, с. 39].

Существует много разновидностей функционализма, например теоретический (каузальный) функционализм («главное — определить переживание через его каузальную роль, так как это удостоверяет его реальность и действенность и одновременно укореняет его в обыденном языке» [Вильянуэва 2006, с. 111]), машинный или вычислительный функционализм (психика — это вычислительное устройство, а ментальные состояния — вычислительные состояния) и т.п. Притягательность функционализма кроется в том, что сознание рассматривается как чистая функция, безотносительно к породившим эту функцию свойствам, феномен «сознания» может быть функционально реализован в человеке, компьютере или каких-либо иных вещах и устройствах (аргумент множественной реальности). Последнее положение близко по своему духу к антипсихологическому тезису логики: логика не является наукой о мышлении (в отличие от психологии,

* Исследование выполнено при финансовой поддержке РГНФ в рамках научно-исследовательского проекта РГНФ «Философия сознания: перспективы развития и возможные ограничения», проект № 04–03–123а.

эпистемологии, нейрофизиологии и других наук о мышлении), но наукой о *правильном* мышлении, а правильно думать может человек, компьютер, робот и т.д.

Подобная декларативная «нейтральность» функционализма породила с течением времени неизбежные вопросы: можно ли понимать «функцию» в отрыве от свойств ее носителей? Каков механизм «реализации» функции, все ли свойства сознания в состоянии имитировать компьютер?

Как замечает Н.С.Юлина, «возникли подозрения: не является ли замена ментальных свойств функциональными отношениями просто подменой понятий; действительно ли функционализму удалось снять корневую проблему об отношении духовного и телесного, или она просто отодвинута в тень» [Юлина 2004, с. 41].

Пытаясь ответить на эти вопросы, функционалисты выдвинули несколько концепций, способные, по мнению их создателей, дать ответы на эти вопросы в рамках функционализма. Одна из них — это концепция *супервентности*, являющаяся, по мнению Джегвон Кима и Дэвида Чалмерса, ключевой в понимании сознания.

Сам по себе термин «супервентность» появился довольно давно (мы обнаруживаем его еще у Дж.Мура); в философии науки он использовался для обозначения особого типа объяснения законосообразного и вместе с тем нередуктивного отношения между наукой большей общности (например, физикой) и науками меньшей общности (биологией, психологией, социологией и т.д.). Философы сознания и когнитивисты термин «супервентность» стали использовать для объяснения особого типа психофизической связи, когда психическое является следствием физического, сопутствует ему, но не сводится к тем или иным конкретным процессам мозга. При этом в зависимости от ответа на главный вопрос о том, какой характер связи имеется в виду, когда говорят об отношении супервентности (является ли она логической, физической, необходимой, законосообразной и т.д.), используется «сильное» или «слабое» ее истолкование. Второе из них является гораздо более популярным, чем первое.

Так, например, слабое понимание супервентности мы находим в концепции «аномального монизма» Дональда Дэвидсона, считающего, что, поскольку связь физических и ментальных событий невозможно подвести ни под один из известных естественнонаучных законов, эту аномалию следует принять за естественную данность. Психофизическое отношение является законосообразным, но не в смысле «сильной» психофизической супервентности, предполагающей строгий закон, а в смысле допущения аномалий. Сходным образом Ко-

лин МакГинн (McGinn) утверждает, что наличие отношения супервентности можно только декларировать, но сказать с определенностью, является ли это отношение необходимым и законосообразным — невозможно; его попросту следует принять за особенность природы [McGinn 1991].

В целом слабое понимание супервентности удивительно близко к концепции расслоения, используемой в математике. Само понятие расслоения и расслоенного пространства тоже не ново: оно возникло в 1930-х гг. в связи с задачами топологии и геометрии многообразий. Расслоенное пространство проще всего представить себе как совокупность двух пространств, связанных между собой специальным отображением. Одно из двух пространств выбирается в качестве базы расслоенного пространства и в каждую точку базы проектируется экземпляр второго пространства (слой).

Эта математическая конструкция получила широкое распространение к концу XX в. в физике элементарных частиц, в частности в теории калибровочных полей. Для нас, однако, важно то обстоятельство, что сама по себе конструкция расслоения явным образом не налагает каких-либо ограничений на вид пространств базы и слоя. Иначе говоря, природа этих пространств ничем не обусловлена — это могут быть просто множества, векторные пространства, пространства симметрии и т.п.

Если вспомнить теперь, что английское прилагательное *supervenient* в буквальном смысле означает «возникающий как нечто новое в дополнение к чему-либо», то отсутствие ограничения на вид пространства в конструкции расслоения (новизна слоя в отношении базы) позволяет рассматривать расслоение как некую математическую разновидность супервентности. В этом случае роль отношения супервентности играет проектирующее отображение, к виду которого опять же никаких требований не предъявляется.

Таким образом, если попытаться интерпретировать отношение супервентности между физическими и ментальными событиями в рамках конструкции расслоения как отображение пространства физических событий в пространство ментальных событий, то отсутствие каких-либо предварительных требований к этому отображению можно расценивать, следуя слабому пониманию супервентности К.МакГинном, как принятие его в качестве особенностей природы, как постулирование наличия связи между физическим и психическим просто в качестве особенностей природы. Понятно, что подобная аналогия сводит феномен сознания к психофизическому «расслоению», при этом сознание получает интерпретацию в рамках функционалистского подхода.

Заметим, однако, что хотя психофизическое расслоение толерантно к принимаемому виду пространств базы и слоя, однако эта декларируемая толерантность чисто внешняя — после фиксации конкретного вида пространств между свойствами ваших пространств неминуемо возникает внутренняя корреляция, навязываемая принимаемым интерпретирующим его математическим аппаратом. Вы вольны выбирать ваши физические и ментальные структуры и связь между ними (слабая супервентность), однако сделав этот выбор, вы получаете формулировку некоторых структурных ограничений и связей (сильная супервентность), вызванных математическими особенностями вашего расслоения (ваше описание влияет на вид отношения супервентности). Как следствие, чем более абстрактный аппарат вы выбираете, тем слабее влияние самого описания на отношение супервентности, и тем более общим получается ваше понимание психофизической супервентности.

С этой точки зрения становится понятным и утверждение ученых, например Р. Пенроуза, что «научный путь к пониманию феномена сознания несомненно существует, и начинаться этот путь должен с более глубокого познания природы собственно физической реальности» [Пенроуз 2003, с. 11].

Действительно, для понимания сознания новая квантовая теория необходима хотя бы потому, что она неминуемо будет использовать более общий математический аппарат, приводя к более общей психофизической супервентности. Точно так же понятно и стремление Пенроуза к невычислимости сознательного мышления: компьютационная модель сознания чересчур конкретна — она навязывает жесткую внутреннюю и внешнюю супервентность. В этом случае нужно либо вводить более общее понятие и аппарат вычислимости, либо более общее понятие невычислимости, включающее в себя вычислимость как частный случай.

Неклассическое «классическое» сознание, альтернативная машина Тьюринга и паранепротиворечивость

Проиллюстрируем, как степень абстрактности при определении психофизической супервентности отражается на функционалистских формулировках. Рассмотрим с этой целью принимаемый Пенроузом тезис об отношении вычисления и сознательного мышления, который выглядит следующим образом: «Осознание является результатом соответствующей физической активности мозга, однако эту физиче-

скую активность невозможно должным образом смоделировать вычислительными средствами» [Пенроуз 2003, с. 33]. Таким образом, предложенное еще в 1960 г. Хилари Патнэмом объяснение сознания по аналогии с машиной Тьюринга, по всей видимости, ошибочно: виды ментальных состояний не следует считать нейтральными функциональными состояниями, аналогичными логическим состояниям машины Тьюринга.

Не занимаясь анализом аргументов Пенроуза в пользу своей точки зрения, обратим внимание на то, что этот тезис в такой формулировке неточен, ибо скрыто предполагает важное ограничение. Заключается оно в том, что в нем речь идет о *рекурсивных* вычислениях, в частности о *рекурсивной* машине Тьюринга. Но дело в том, что плюрализм современной логики, ведущий к пролиферации систем неклассической логики, заставляет нас учитывать и возможность существования, например, *альтернативной* машины Тьюринга.

Так, Я.Хинтикка в работе «Альтернативная концепция вычислимости», написанной им совместно с А.Мутаненом [Hintikka 1998], основываясь на своей системе так называемой (дружественной к независимости) IF-логики, формулирует нестандартную концепцию вычислимости. Описать ее можно кратко следующим образом.

Рассмотрим машину Тьюринга с двумя лентами, рабочей и результирующей. Машина может записывать и стирать символы на каждой ленте. С точки зрения понятия ошибкоустойчивости машины Тьюринга это ограничение не существенно. Допустим, что на результирующей ленте записаны уравнения вида $f(a) = b$, где a и b — натуральные числа. Эти уравнения определяют вычислимую функцию. Рассматриваемая машина Тьюринга вычисляет функцию тогда и только тогда, когда все (и только они) подобные истинные уравнения записываются на ленте в процессе работы. Это означает, что каждое истинное уравнение $f(a) = b$ появится на результирующей ленте после нормального завершения некоторого конечного числа шагов. Нет необходимости говорить, что на результирующей ленте для любого данного a должно в конце концов появиться лишь одно уравнение вида $f(a) = b$. Если на результирующей ленте не появится для данного a ни одного подобного уравнения, или оно будет постоянно меняться, то $f(x)$ не определена для $x = a$.

Хотя на первый взгляд кажется, что это определение в точности совпадает с определением рекурсивной вычислимости, но в действительности это не так. Чтобы сузить это определение до рекурсивного, требуется, чтобы: а) машина никогда ничего не стирала на результирующей ленте; б) можно было бы эффективно решить, когда полу-

чено последнее уравнение. Вычислимость без пунктов а) и б) Хинтика называет *tae*-вычислимостью (trial-and-error-computability – вычислимостью с пробами и ошибками).

Место, которое занимает *tae*-вычислимость в систематической логической теории, можно продемонстрировать с помощью следующей аналогии:

$$\frac{\textit{tae}\text{-вычислимость}}{\text{рекурсивность}} = \frac{\text{классическая логика}}{\text{конструктивистская логика}}$$

Это проливает свет на интересную, но в основном неисследованную связь между логической теорией и теорией вычислимости. Среди других философских последствий принятия альтернативной концепции *tae*-вычислимости можно отметить отмену известного тезиса А.Чёрча (1936), гласящего, что механически определимые (конечно вычислимые) функции (в дотеоретическом смысле) являются в точности рекурсивными функциями. Поскольку понятие *tae*-вычислимости включает в себя рекурсивность, а не наоборот, то, по-видимому, следует взамен перейти к какой-то «*tae*-разновидности» тезиса Чёрча.

Парадоксальным образом аналогия Хинтики предполагает, что рассматриваемая им альтернативная машина Тьюринга – *классическая* машина Тьюринга, то есть, если уж считать виды ментальных состояний нейтральными функциональными состояниями, аналогичными логическим состояниям машины Тьюринга, то речь должна идти о логических состояниях *альтернативной* машины Тьюринга, ибо классическая логика по умолчанию и есть логика нашего мышления. Но тогда тезис Пенроуза о невычислимости должен быть скорректирован, ибо в нем ничего не говорится о возможности или невозможности моделирования физической активности «классическими» вычислительными средствами.

Далее, Пенроуз в [Пенроуз 2003а] приводит доказательство теоремы Гёделя, а в [Пенроуз 2003] приводит обзор возможных формальных возражений против его тезиса о невычислимости, основанных на критике его доказательства. В большинстве из них речь идет о непротиворечивых формальных системах, что и позволяет защитить от нападок теорему Гёделя, на которой в значительной степени основывается понятие Пенроуза о невозможности вычислительной модели сознания. Однако среди систем современной неклассической логики существуют так называемые системы паранепротиворечивой логики. Эти системы не непротиворечивы, но лишь паранепротиворечивы, в них парадоксальность не играет решающей роли, поскольку не работает принцип сведения к абсурду (из абсурда не следует все

что угодно). Как указывает Г.Прист [Priest 2002], по этой причине в паранепротиворечивой логике (например, в так называемой системе LP) гёделевские теоремы не доказуемы. Но тогда требуют модификации и аргументы Пенроуза, ибо во всех его рассмотрениях фигурируют непротиворечивые формальные системы.

Обратная комбинированная трехзначная логика нейрона и логическая супервентность

Попробуем теперь повысить уровень абстрактности психофизической супервентности, прибегнув к логическим построениям. Воспользуемся с этой целью тем обстоятельством, что в последнее время конструкции расслоения проникли и в современную логику.

В работе [Юрьев 2001] описана трехзначная логика, разработанная с целью моделирования работы нервной клетки — нейрона, которая позволяет описать нейрон как простейший элемент сложной системы (например, головного мозга). В качестве значений истинности здесь фигурируют значения постсинаптического потенциала реального биологического нейрона, закодированные следующим образом: значению 1 (истина) соответствует возбуждающий постсинаптический потенциал, значению $1/2$ (неопределенность) соответствует потенциал покоя, а значению 0 (ложь) — тормозной постсинаптический потенциал.

Табличные определения основных логических операций этой логики в стандартном виде выглядят следующим образом:

p	$\sim p$
1	0
$1/2$	$1/2$
0	1

$\vee$	1	$1/2$	0
1	1	1	$1/2$
$1/2$	1	$1/2$	0
0	$1/2$	0	0

$\&$	1	$1/2$	0
1	1	$1/2$	$1/2$
$1/2$	$1/2$	$1/2$	$1/2$
0	$1/2$	$1/2$	0

Как нетрудно заметить, свойства приводимых операций в высшей степени необычны, поэтому трудно говорить здесь о какой-либо логической теории, т.е. трудно ее как-то классифицировать. А.С.Карпенко было показано, что посредством данного множества операций не выразимы минимаксные операции трехзначной логики Клини, т.е. операции $x \ll y = \min(x, y)$; $x \gg y = \max(x, y)$ [Карпенко 2001, с. 42]. Фактически это означает и невозможность получения соответствующих операций классической двузначной логики в рамках подобной системы. Данное обстоятельство означает для нас то, что мы имеем в этом

случае дело с «логикой» физических состояний нейрона, а не ментальных состояний сознания, поскольку именно классическая логика по умолчанию принимается как логика нашего правильного мышления (и, следовательно, сознания).

Чтобы получить классические двузначные операции, пополним язык за счет новой одноместной логической операции:

p	θp
1	1
$1/2$	0
0	0

Используя эти табличные определения, можно установить, что введенная операция обладает, например, следующими свойствами:

$$(1) \theta(p \& q) = \theta(\theta p \& \theta q)$$

$$(2) \theta(\sim p) = \theta \sim (p \vee \sim \theta(p \& \sim \theta p))$$

$$(3) \theta \theta p = \theta p$$

Поскольку θ -оператор, по сути дела, превращает трехзначные формулы в двузначные, то нетрудно проверить, что таблицы для $\theta(p \& q)$ и $\theta(\sim p)$ совпадают с таблицами для классической конъюнкции и отрицания.

Возникающий вопрос связан со статусом θ -оператора. Учитывая, что, значению 1 соответствует возбуждающий постсинаптический потенциал, $1/2$ — потенциалу покоя, а 0 — тормозному постсинаптическому потенциалу, то, принимая во внимание, что θ -оператор имеет дело только с потенциалами возбуждения и торможения, можно было бы рассматривать его как своеобразный модальный оператор «необходимости», когда потенциал обязательно есть либо его обязательно нет, но промежуточного состояния покоя для θ -формул не существует. С другой стороны, наличие θ -оператора подразумевает существование некоего механизма, когда некоторые из нейронов переводят сигналы от других нейронов только в «двузначные», образующие области с классической, а не трехзначной логики (например, области коры мозга в отличие от областей подкорки).

Существует и еще одна возможность интерпретации θ -оператора. В так называемых комбинированных логиках предложений и событий, разработанных В.А.Смирновым, вводится θ -оператор, переводящий события (образующие алгебру событий) в формулы, так что, например, θa можно истолковать как высказывание « a является актуальным событием» (см. [Vasyukov 1999, p. 130]). Поскольку в на-

шем случае мы имеем дело не с логическим исчислением, а с некоторой алгебраической семантикой, то, учитывая направление интерпретации, можно было бы скорее говорить об «обратных» комбинированных логиках, когда высказывания переводятся в события и рассматривается алгебра событий. Если учесть, что таблицы для $\theta(p \& q)$ и $\theta(\sim p)$ совпадают с таблицами для классической конъюнкции и отрицания, то можно было бы с учетом (3) рассматривать (1) и (3) как

$$(4) \theta(p \cap q) = \theta[\theta p \& \theta q]$$

$$(5) \theta(\neg p) = \theta[\sim([\theta p] \vee \sim \theta[\theta p \& \sim \theta p])]$$

где $\cap$ — конъюнкция, $\neg$ — дополнение в булевой алгебре, а $[x]$ означает элемент булевой алгебры, соответствующий обратному переводу. Однако в этом случае изменяется смысл нашего θ -оператора, поскольку мы теперь должны под p , q понимать элементы подобной булевой алгебры. Можно предположить, что это булева алгебра представляет собой алгебру ментальных состояний, которая согласно концепции психофизиологического дуализма связана с физическими процессами в нейроне. Последнее станет еще более очевидным, если предположить, что наша булева алгебра является алгеброй множеств. В этом случае $\cap$ и $\neg$ будут представлять собой теоретико-множественное пересечение и дополнение, а θp будет представлять собой связь между сигналом (потенциалом) нейрона и множеством ментальных состояний, соответствующих данному потенциалу при наблюдении.

Подобная интерпретация подразумевает конструкцию алгебраического расслоения $\langle A, B, f, g \rangle$, где $A = \langle A, \&, \vee, \sim \rangle$ является алгеброй формул трехзначной логики, определяемой нашими таблицами, $B = \langle B, \cap, \neg \rangle$ представляет собой булеву алгебру ментальных состояний, $f: B \rightarrow A$, $g: A \rightarrow B$ являются функциями вложения. Для f и g должны выполняться следующие условия:

$$f(a \cap b) = f(g(f(a) \& f(b)))$$

$$f(\neg a) = f(g(\neg(g(f(a)) \vee \neg f(g(f(a) \& \neg f(a)))))$$

где $a, b \in A$.

Полученное расслоение можно было бы рассматривать как описание психофизической супервентности, основанное на логических конструкциях в рамках функционализма, т.е. как некоторую версию «логического» функционализма. При этом, однако, само отношение супервентности носит алгебраический характер и представлено конструкцией функции f , да и, собственно говоря, в роли базы и слоя выступают не логические системы, а алгебры (алгебра формул трехзначной логики и булева алгебра ментальных состояний). Получить чисто логическую (синтаксическую) формулировку не удастся ввиду того, что с самого начала трудно было как-то квалифицировать саму трехзначную логику нейрона.

Заметим, что тот факт, что наша булева алгебра действительно должна быть алгеброй множеств, подтверждается следующим обстоятельством. Против введения супервентности в объяснительные схемы были высказаны в свое время возражения, сводящиеся к тому, что нет оснований вводить понятие «супервентность» и делать из него «последнее убежище современного физикализма», поскольку для него нет никакой эмпирической очевидности: два человека могут иметь одни и те же внутренние физические состояния, но еще не факт, что их ментальные состояния будут одинаковыми (Д.Меллор и Т.Крейн, Дж.Серль). С точки зрения рассмотренной выше конструкции психофизического расслоения этот аргумент бьет мимо цели: если у двух человек одни и те же нейроны находятся в одинаковом физическом состоянии (например, их постсинаптический потенциал имеет одно и то же значение), то это действительно не означает, что они находятся в одном и том же ментальном состоянии, однако это несколько не влияет на отношение супервентности. Согласно конструкции расслоения, если в роли слоя выступает булева алгебра множеств ментальных состояний, то физическому состоянию нейрона у двух людей могут соответствовать разные ментальные состояния из множества эквивалентных ментальных состояний: от них требуется лишь только то, чтобы они были элементами этого множества, они неразличимы в этом отношении.

Рассмотрим этот вопрос более подробно.

Физические и ментальные Чу-пространства

Возн Пратт в работе «Рациональная механика и естественная математика» [Pratt 1995] обращает внимание на следующее обстоятельство, способное, по его мнению, служить основой для функционального рассмотрения проблемы психофизического дуализма. Если мы будем различать тела и сознания, кодируя их с помощью переменных типа A , B и X , Y соответственно, то следует различать, как они преобразуются друг в друга. Тела связаны между собой помощью *функций*, в то время как сознания связываются с помощью *антифункций*. Антифункция $g: X \rightarrow Y$ представляет собой бинарное отношение между X и Y , обратным к которому является функция $g: Y \rightarrow X$ (заметим, что функция тоже является отношением, главное различие, которое нас интересует заключается лишь в том, что функция не допускает сопоставления «одно-многое», разрешая лишь сопоставление «многое-одно», т.е. разным точкам в области определения функции может соответствовать одна и та же точка в области ее значения, но не наоборот).

Функции отождествляют и сопрягают, в то время как антифункции копируют и удаляют. Отождествление функцией означает, что некоторые элементы отображаются в один и тот же элемент (случай, когда функция $f: A \rightarrow B$ определяет тождественность a и b из A и B соответственно, как результат выполнения условия $f(a) = f(b)$), сопряжение появляется в том случае, когда $f: A \rightarrow B$ переводит сначала A в $f(A)$, а затем мы сопоставляем A дополнение $f(A)$ до B . Антифункция $g^{\perp}: X \rightarrow Y$ копирует в том случае, когда для обратной функции $g: Y \rightarrow X$ выполняется условие $g(y) = g(y')$. В этом случае мы говорим, что $g^{\perp}$ копирует $g(y)$ в y и y' одновременно. Антифункция $g^{\perp}: X \rightarrow Y$ удаляет, если функция g не является сюръективной (если g задана не на всем Y), т.е. антифункция $g^{\perp}$ удаляет в точности дополнение $g(Y)$ до X . Таким образом, отождествление и сопряжение характерны для области физического, в то время как копирование и удаление являются характерными признаками ментального.

Пратт предлагает использовать математическую конструкцию так называемых *Чу-пространств* (или *пространств Чу*) для описания феномена психофизического дуализма, так как она, по его мнению, позволяет учесть в одной структуре вышеописанные свойства физических тел (нейронов) и сознаний (ментальных состояний). Чу-пространство представляет собой структуру $A = (A, X, \models)$ над множеством K , состоящую из множества A точек (тело), антимножества X состояний (сознание) и $A \times X$ -матрицы $\models$, образуемой элементами K (т.е. функции $\models: A \times X \rightarrow K$). Множество состояний названо антимножеством, так как его элементы связывают не функции, но антифункции, а матрица $\models$ призвана описать психофизическое взаимодействие.

На первый взгляд подобное Чу-пространство можно легко адаптировать для описания трехзначной логики нейрона. Возьмем, например, $A = \{a, b, c\}$ (т.е. рассмотрим систему трех нейронов), $K = 3$ (ввиду трехзначности рассматриваемой логики нейрона) и $X = \{x, y, z, k\}$ (т.е. принимаем, что существуют четыре состояния системы нейронов). Рассмотрим следующее Чу-пространство, представляющее нашу систему нейронов:

	x	Y	z	k
a	1	$1/2$	0	0
b	$1/2$	$1/2$	$1/2$	$1/2$
c	0	$1/2$	1	1

Нетрудно заметить, что в нашем Чу-пространстве $a = \sim c$ и $b = a \& b$ (это легко проверить, используя таблицы трехзначной логики нейрона, рассматривая справедливость этих равенств для каждого состояния).

Первое уточнение, которое сразу же приходится сделать, заключается в особенностях употребления термина «состояние». Ибо когда речь шла о состояниях нейрона при рассмотрении истинностных таблиц трехзначной логики, то под состоянием понималось просто истинностное значение. В рассматриваемом же Чу-пространстве под состояниями понимается уже совокупность истинностных значений всех трех нейронов (столбец матрицы). В этом случае понятно, что, несмотря на то, что два нейрона могут находиться в одном и том же «физическом» состоянии (иметь одно и то же значение постсинаптического потенциала), они одновременно могут находиться в разных «ментальных» состояниях. Так, в нашем Чу-пространстве нейрон a имеет значение потенциала 1 и нейрон c имеет значение потенциала 1, однако нейрон a имеет это значение в состоянии x , а нейрон c имеет это значение в состояниях z и k . Тем самым уже даже в такой «примитивной» функциональной модели сознания видно, что утверждение Меллора-Крейна-Серля (о том, что нет никакой эмпирической очевидности того, что два человека могут иметь одни и те же внутренние физические состояния, но еще не факт, что их ментальные состояния будут одинаковыми) справедливо (в нашем Чу-пространстве нет даже теоретической очевидности), но, тем не менее, мы в состоянии учитывать все возможные ментальные состояния, коррелирующие с данным внутренним физическим состоянием.

Однако хуже выглядит другое обстоятельство особенностей употребления термина «состояние». Являются ли состояния нашего Чу-пространства «ментальными» состояниями? Вспомним, что наш θ -оператор был введен для описания связи между сигналом (потенциалом) нейрона и множеством ментальных состояний, соответствующих данному потенциалу при наблюдении. Необходимость его введения мотивировалась результатом о невыразимости классических двузначных операций на основании принятых таблиц, т.е. внутри нашей логики нейрона не содержится классической логики. Это означает, что и наши состояния не подчиняются классической логике (на модели это можно проверить, переходя к дуальному Чу-пространству $A^\circ = (X, A, \lrcorner)$, т.е. к обратной $X \times A$ -матрице, и проверяя на ней действие табличных операций).

Следовательно, состояния рассматриваемого Чу-пространства не являются настоящими ментальными состояниями, но лишь физическими состояниями нейрона. Чтобы связать эти состояния с ментальными состояниями, нам следует ввести в рассмотрение Чу-пространства с двузначной матрицей, т.е. с множеством K , состоящим из двух элементов. В этом случае можно определить классиче-

ские операции на значениях потенциалов нейрона и получить действительно «ментальные» состояния, подчиняющиеся классической двужначной логике.

Чтобы определить теперь связь между ментальными и физическими состояниями, введем понятие *Чу-преобразования* $(f, g): (A_1, X, \models) \rightarrow (A', X', \models')$, состоящего из функции $f: A \rightarrow A'$ и антифункции $g: X \rightarrow X'$ т.е. отношения, обратного к функции $g: X' \rightarrow X$, удовлетворяющей условию непрерывности $g(x') \models a \mapsto x' \models' f(a)$ для всех $a \in A$ и $x' \in X'$. Для наших целей требуется Чу-преобразование из Чу-пространства A_1 с двухэлементным K_1 в Чу-пространство A_2 с трехэлементным K_2 , при том, что в A_1 определены булевы операции, а в A_2 — операции трехзначной логики нейрона. Условие непрерывности у нас означает, что элементы 0,1 двухэлементного множества $K_1 = \{0, 1\}$ отображаются в элементы 0,1 трехэлементного множества $K_2 = \{0, \frac{1}{2}, 1\}$. Функция f преобразует точки (нейроны), отождествляя некоторые из них и сопрягая другие, но не удаляя и не удваивая их. Антифункция g преобразует ментальные состояния в физические состояния, удаляя некоторые из состояний, ограничивая степени свободы пространства точек и копируя другие так, чтобы не нарушить степеней свободы новосопряженных точек (нейронов). Преобразования, таким образом, должны сохранять структуру и не несут ответственность за то, что происходит в сопряженных нейронах.

Такая формулировка определяет связь между нейронами и ментальными состояниями двояким образом: через связь нейронов с двумя значениями потенциала с нейронами с тремя значениями потенциала и через связь ментальных и физических состояний. При этом условие непрерывности гарантирует, что значение потенциала двужначного нейрона в ментальном состоянии совпадает со значением потенциала трехзначного нейрона в физическом состоянии, отвечающем этому ментальному состоянию.

В целом, нетрудно прийти к выводу, что сама по себе конструкция *Чу-преобразования* $(f, g): (A, X, \models) \rightarrow (A', X', \models')$ дает нам не что иное, как отношение психофизической супервентности. Таким образом, связь между психическими и ментальными состояниями описывается с помощью комбинации идей подхода Пратта к проблеме психофизического дуализма и обратной комбинированной трехзначной логики нейрона как расслоение с базой $(A', X', \models')$, слоем $(A, X, \models)$ и отображением (f, g) . Если бы трехзначная логика нейрона допускала бы синтаксическую формулировку, то конструкция подобного Чу-расслоения описывала бы, по сути дела, семантику комбинированной трехзначной логики нейрона.

Квантовая природа сознания и логический функционализм

Вернемся к утверждению, что чем более абстрактный аппарат вы выбираете, тем слабее влияние самого описания на отношение супервентности и тем более общим получается ваше понимание психофизической супервентности. В этом отношении для понимания феномена сознания было бы полезным, например, использование квантовой теории хотя бы потому, что она основана на более общем математическом аппарате, приводя к более общей психофизической супервентности.

Существуют, конечно, и другие причины для использования квантовой теории с целью объяснения феномена сознания. Так, Р. Пенроуз, развивая один из вариантов квантовой версии физикализма, высказывает гипотезу о том, что неалгоритмический характер процессов сознания, возможно, связан с объективным коллапсом волновых функций макроскопических переменных (см. [Пенроуз 2005]). Главный ключ к пониманию сознания он видит в создании революционной теории, основанной на новом — невычислительном — типе квантовой механики, которая примирит квантовую теорию с теорией относительности, объяснит гравитацию, физику нейрофизиологических процессов, и на этой основе объяснит и феномен сознания.

Однако попытки объяснения феномена сознания предпринимались до Пенроуза и на основе старого — вычислительного — типа квантовой механики (Х.Стапп, Л.Рикьярди и Х.Умезава, Дж.Витиелло, М.Джибу и К.Ясуэ). Способ, которым при этом часто формулировался вопрос о квантовой природе субстрата сознания, подсказывает аналогию с проблемой скрытых параметров в квантовой теории, когда все отклонения от обычной классической (булевой) теории описания реальности полагаются обусловленными неполнотой любого нашего знания, а выход видят во введении некоторых новых, ранее скрытых параметров-переменных, позволяющих отвергнуть специфическое описание на языке квантовой теории. Более того, подобная аналогия может быть обоснована путем отсылки к некоторым научным теориям сознания, использующим скрытые переменные при попытке объяснения физических оснований феномена сознания. Например, согласно Э.Х.Уокеру [Walker 1970, p. 175], если мы допускаем существование «скрытых переменных», чьи значения (в настоящем и будущем) позволяют просто детерминировать все будущие состояния психической системы, вызванной функционированием мозга, то они отождествляются с самим сознанием.

В книге «Формальная феноменология» [Васюков 1999] автором была разработана концепция формальной феноменологии и разработаны логические системы, для семантики которых характерны предметные области с интенциональными объектами. Все эти системы представляли собой некоторое «объектное» расширение системы онтологии Ст.Лесьневского. При этом были разработаны и системы формальной квантовой феноменологии, описывающие с позиции логического функционализма, взаимодействие квантовых структур с интенциональными объектами сознания.

В онтологии Лесьневского свойства объектов описывается булевой алгеброй их имен, а с точки зрения семантики этой системы совокупности объектов, имена которых фиксируются в синтаксисе, образуют объектные состояния сознания. Если определить на этих состояниях отношения достижимости и определить интенциональность объектов как существование их лишь в некоторых (или, наоборот, во всех) состояниях, то возникает возможность дополнить объектное содержание состояний за счет таких «феноменологических» объектов, что приводит к понятию интенциональных состояний сознания.

Если принять допущение, что существование «скрытых переменных», ответственных за описание квантовых феноменов (объектов) сознания, может проявляться в наличии отдельной специальной небулевой алгебры имен, то интенциональное содержание сознания начинает зависеть от взаимодействия этих двух алгебр имен. Во всяком случае, наличие ортомодулярности (или ортодополнительности) во взаимоотношении ментальных объектов может быть истолковано как косвенное свидетельство их квантовой природы.

В работе [Васюков 1999] строятся системы формальной феноменологии, содержащие одновременно и булеву алгебру имен и ортомодулярную решетку имен, что позволяет говорить о системах *квантовой феноменологии* в том смысле, в каком говорят о системах квантовой логики, возникающей в рамках той или иной логической системы. Более того, подобно тому, как квантовая логика допускает модальную интерпретацию (перевод квантовых формул в модальные), системы квантовой феноменологии допускают интенциональную интерпретацию (рассматривая интенциональность как особый вид «объектной» модальности). Это можно было бы расценивать с точки зрения логического функционализма как наличие следов (отражений) квантовых феноменов в интенционально направленном сознании и самом сознании соответственно. Или же наоборот, рассматривать это как возможность элиминации «квантовых оснований» сознания в пользу неразличимости физического субстрата и чистого сознания.

Однако, несмотря на все эти возможности появления квантовой структуры в различных системах формальной феноменологии, наша позиция не должна быть избавлена от известной осторожности: мы не могли бы сказать, что подобная картина является свидетельством квантового основания сознания, потому что имеются некоторые примеры макроскопических моделей квантовых систем (см., например, [Finkelstein Finkelstein 1983]). Это означает, что мы не в состоянии различать, когда квантовая алгебра имен описывает квантовые микроскопические объекты, а когда речь идет о макроскопических не-квантовых объектах.

Тем не менее все предложенные феноменологические системы, хотя они и не являются окончательным решением, все же дают нам некоторое формальное описание (может быть, чересчур абстрактное) сплетения телесного субстрата сознания с самим интенционально направленным сознанием. В этой связи заметим, что некоторые конструкции, рассмотренные в [Васюков 1999], позволяют сослаться на гипотезу интерференции образов в [Гриб 1989, с. 156], где переход от дистрибутивной решетки нейронов и связей между ними к недистрибутивной решетке приводит к перестройке сознания и замещению классической картины квантовой. Как следствие, может иметь место интерференция образов, что в свою очередь может вызвать, например, спонтанное появление человека и коня (кентавра) одновременно (во сне) и т.д. В нашем случае спонтанная перестройка интенционально направленного сознания может быть интерпретирована как сплетение субстрата и сознания, а интерференция образов может быть также обоснована специфической структурой семантики систем формальной феноменологии, где образование интенционального объекта в некотором интенциональном состоянии осуществляется из двух различных объектов, находящихся в двух различных состояниях, но эти состояния одновременно достижимы из рассматриваемого интенционального состояния.

Как бы там ни было, с точки зрения логического функционализма все эти возможности можно рассматривать как результат реализации психофизического расслоения, когда в качестве базы подобного расслоения рассматривается квантовая алгебра имен (в основном ортомодулярная решетка), ответственная за описание квантовой природы феномена сознания, а в качестве слоя принимается булева алгебра имен онтологии Ст.Лесьневского. Конечно, отношение супервентности здесь варьируется в зависимости от особенностей формулировки взаимодействия квантовой и булевой алгебр имен в соответствующих системах формальной феноменологии, однако само его наличие во всех случаях несомненно.

Заключение

Итак, действительно ли, что чем более абстрактный аппарат вы выбираете, тем слабее влияние самого описания на отношение супервентности и тем более общим получается ваше понимание психофизической супервентности? В данной статье с целью проверки этого тезиса была предпринята попытка адаптации аппарата современной неклассической логики к проблемам философии сознания. Как следствие, подход к описанию психофизической супервентности, основывающийся на логических конструкциях в рамках функционализма, был квалифицирован как некоторая версия «логического» функционализма.

Что же можно сказать, подводя итоги исследования, о достоинствах и недостатках логического функционализма? Скорее всего, против логического функционализма можно выдвинуть все те же возражения, которые выдвигались против функционализма вообще. Джон Серль, например, в свое время критиковал Д.Деннета за то, что сознание у него сведено к функционированию когнитивных информационных процессов (в человеке или компьютере) и лишено самого главного свойства — осознанности себя как живого сознающего существа. На первый взгляд логический функционализм явно грешит тем же самым недостатком — ведь логика всегда безлична и безразлична к тому, кто мыслит.

Однако если просто говорить об *осознанности* как о способности самосознания, как о самосознании, то положение здесь не столь уж безнадежно. Во-первых, самосознание формально можно симулировать с помощью конструкции вложенного (двойного последовательного) расслоения. В этом случае в качестве слоя нужно выбирать уже расслоенное пространство, причем это расслоение второго уровня осуществляется за счет выбора одинаковой базы и слоя. В рассмотренном случае трехзначной логики нейрона расслоение второго уровня должно производиться за счет рефлексивной конструкции *Чу-преобразования* типа $(f, g): (A, X, \models) \rightarrow (A, X, \models)$, где база и слой совпадают. Антифункция g в этом случае преобразует ментальные состояния в ментальные состояния, удаляя некоторые из состояний, ограничивая степени свободы пространства точек и копируя другие так, чтобы не нарушить степеней свободы новосопряженных точек (двузначных нейронов).

С синтаксической точки зрения конструкция вложенного расслоения сводится к допустимости двойного применения оператора $[-]$. Если в рамках комбинированной трехзначной логики нейрона $[x]$ означает элемент булевой алгебры, соответствующий обратному пе-

реводу, то $[[x]]$ теперь будет означать элемент булевой алгебры самосознания, а уравнение $\theta\theta p = \theta p$ можно понимать теперь как утверждение о классическом двузначном характере самосознания.

Вторую часть упрека Серля — отсутствие осознанности себя как *живого сознающего существа* — можно было бы точно так же парировать посредством двойного расслоения, однако теперь принимая в качестве базы наше расслоенное пространство, а в качестве слоя — Чу-пространство двузначных нейронов. В этом случае антифункция финального Чу-преобразования будет преобразовывать антифункции преобразования физических состояний в ментальные состояния самосознания, т.е. подразумевая, что ментальные состояния самосознания детерминированы актами соответствия физических состояний ментальным.

Ясно в этом случае лишь одно: логический функционализм, в том виде, как он здесь представлен, приводит к излишне громоздким формальным построениям, что вряд ли можно поставить ему в заслугу. С другой стороны, ясно и направление дальнейших исследований: они должны быть направлены на устранение этой громоздкости.

Литература

- [Васюков 1999] *Васюков В.Л.* Формальная феноменология. М., 1999.
- [Вильянуэва 2006] *Вильянуэва Э.* Что такое психологические свойства. М., 2006.
- [Гриб 1989] *Гриб А.А.* Макроскопические реализации квантовой логики // Современные исследования по квантовой логике /Под ред. В.С.Меськова, Б.Н.Пятницына, Москва: МГУ, 1989. С. 150—156.
- [Карпенко 2001] *Карпенко А.С.* Не истинностно-функциональная логика Клини с антибулевой операцией. I // Труды научно-исследовательского семинара по логике. Вып. XV. М., 2001. С. 41—45.
- [Пенроуз 2003] *Пенроуз Р.* Тени разума. В поисках науки о сознании. Ч. 1. М., 2003.
- [Пенроуз 2003а] *Пенроуз Р.* Новый ум короля. О компьютерах, мышлении и новых законах физики. М., 2003.
- [Пенроуз 2005] *Пенроуз Р.* Тени разума. В поисках науки о сознании. Ч. 2. М., 2005.
- [Юдина 2004] *Юдина Н.С.* Головоломки проблемы сознания. М., 2004.
- [Юрьев 2001] *Юрьев Д.Н.* Новая трехзначная логика // Труды научно-исследовательского семинара по логике. Вып. XV. М., 2001. С. 120—125.
- [Finkelstein Finkelstein 1983] *Finkelstein D., Finkelstein S.R.* Computational Complementarity // Int. J. Theor. Phys. 22 (1983). P. 753—779.
- [Hintikka 1998] *Hintikka J., Mutanen A.* An Alternative Concept of Computability // Hintikka J. Selected Papers. Vol. 3. Language, Truth and Logic in Mathematics, Kluwer Academic Publishers, Dordrecht, Boston—London, 1998. P. 174—188.
- [McGinn 1991] *McGinn C.* The Problem of Consciousness. Cambridge, 1991.
- [Pratt 1995] *Pratt V.R.* Rational Mechanics and Natural Mathematics // Proc. TAPSOFT'95, LNCS 915, 108—122, Aarhus, Denmark. May 1995.

[Priest 2002] *Priest G.* Paraconsistent Logic // Handbook of Philosophical Logic (Second Edition), Vol. 6 /D.Gabbay and F.Guenthner (eds.), Kluwer Academic Publishers, Dordrecht, 2002. P. 287–393.

[Walker 1970] *Walker E.H.* The Nature of Consciousness // Mathematical Biosciences. 7 (1970). P. 131–178.

[Vasyukov 1999] *Vasyukov V.L.* Combined da Costa Logics (world according to N.C.A.da Costa) // Logique et Analyse. 165–166(1999). P. 127–138.